

# Enabling Intercontinental e-Infrastructures – a Case for Africa

Ognjen PRNJAT<sup>1</sup>, Bruce BECKER<sup>2</sup>, Roberto BARBERA<sup>3</sup>, Christos KANELLOPOULOS<sup>1</sup>,  
Kostas KOUMANTAROS<sup>1</sup>, Rafael MAYO-GARCÍA<sup>4</sup>, Federico RUGGIERI<sup>3</sup>

<sup>1</sup>Greek Research and Technology Network, 56 Mesogion Av., Athens, 11527, Greece

Emails: {oprnjat, skanct, kkoum}@admin.grnet.gr

<sup>2</sup>Meraka Institute, Meiring Naudé Road, Pretoria, 0001, South Africa

Email: [bbecker@csir.co.za](mailto:bbecker@csir.co.za)

<sup>3</sup>Italian National Institute of Nuclear Physics

Division of Catania, Via S. Sofia 64, Catania, 95123, Italy

Division of Roma Tre, Via della Vasca Navale 84, Rome, 00146, Italy

Emails: [roberto.barbera@ct.infn.it](mailto:roberto.barbera@ct.infn.it); [federico.ruggieri@roma3.infn.it](mailto:federico.ruggieri@roma3.infn.it)

<sup>4</sup>Centro de Investigaciones Energéticas Medioambientales y Tecnológicas,  
Av. Complutense 40, Madrid, 28040, Spain, Email: [rafael.mayo@ciemat.es](mailto:rafael.mayo@ciemat.es)

**Abstract:** CHAIN-REDS, an EU co-funded project, focuses on promoting and supporting technological and scientific collaboration across different e-Infrastructures established and operated in various continents. The project implemented a Regional Operations Centre (ROC) model for enabling Grid computing interoperation across continents, and an operational ROC has been set up for Africa. Moreover, the project operates a global Cloud federation test-bed, where also nascent African Cloud sites contribute. Finally, the project is supporting a real-life use-case from Africa that of APHRC, using its data e-Infrastructure services.

**Keywords:** e-Infrastructures, Grid interoperations, Cloud federations, data management, persistent identifiers.

## 1. Introduction

Electronic Infrastructures or e-Infrastructures - including networking, computing (Grid, Cloud, High-performance computing), and data infrastructures - are the key enabler for Research and Education activities worldwide. Building on the outputs and the momentum of the European Commission's CHAIN project and expanding on the background of knowledge and a long experience gained through a large number of initiatives addressing collaboration between European and other regional e-Infrastructures over the past decade, the CHAIN-REDS project [1] is the de facto flagship EU project in the field of global e-Infrastructure collaboration. CHAIN-REDS is focused on promoting and supporting technological and scientific collaboration across different e-Infrastructures established and operated in various continents. It aims to facilitate e-Infrastructures uptake and interoperability, and intercontinental e-Infrastructure use by established and emerging Virtual Research Communities (VRCs) but also by single researchers.

This paper focuses on the CHAIN-REDS approach to computing aspects of e-Infrastructures, including Grid and Cloud infrastructures. In sections 2.1 and 2.2 it presents the model for the Regional Operations Centre, and its applicability in Africa in section 2.3. In section 3, it describes the global cloud test-bed where also a cloud installation from Africa participates. The paper then discusses a practical use case which takes advantage of the actual project e-Infrastructure data management services in section 4, and conclusions

are given in section 5. The paper is of interest for the e-Infrastructure operators which can join their computing clusters in Grid or Cloud mode to the global e-Infrastructures, thus effectively sharing resources and increasing the resource access pool available for their users. It is also of interest for the users from the African Research and Education community, who can exploit the capacity and services of e-Infrastructures on offer.

## 2. Grid Interoperations

### 2.1 General Collaboration Model

CHAIN-REDS is enabling 5 Regional Operations Centres (ROCs) around the world to both technically and procedurally interoperate with the European Grid Infrastructure (EGI) [2]. These centres are set up and maintained by the regions themselves independently of the CHAIN-REDS project, while the project itself facilitates their interoperation with their European counterpart and supports the use and adoption of European technologies and approaches. The ROCs are responsible for Grid computing sites in their regions, coordinating daily operations and interfacing with EGI.

On the policy level, 5 Memoranda of Understanding (MoUs) have been signed with direct support and facilitation provided by the CHAIN-REDS project, defining the framework for long-term persistent collaboration between European Grid Infrastructure and the other regional Grids. 5 MoUs have been signed between EGI and Africa&Arabia ROC, China, India, South-East Asia and Latin America. The figure below (by EGI), presents the interoperation model. The MoUs with these regions – including Africa&Arabia (A&A) – have the form of as Integrated Resource Infrastructure Providers’ MoUs, since they fully adopt the EGI policies and technical approaches. The only exception is India, which has a MoU as a Peer Resource Infrastructure Provider (using different middleware and technical solutions). The schematic is shown in the Figure 1 below.

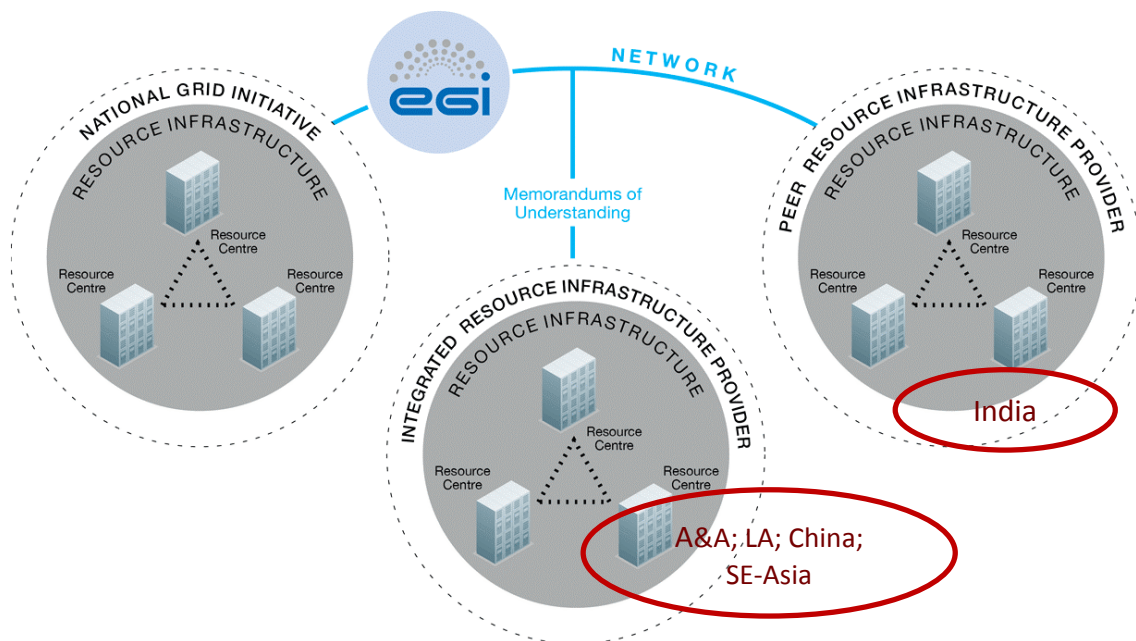


Fig.1. EGI Relationship with Other World Regions

### 2.2 ROC Guidelines

The project has produced a set of guidelines how to build and support Regional Operations Centres for Grid computing, responsible for the coordination and management of the operations of the infrastructures. The functionality encompasses authentication and

authorization, monitoring, user and operational support, management of Service Level Agreements, helpdesks, etc – all for the final benefit of Virtual Research Communities.

In order for a ROC to become an Integrated Resource Infrastructure Provider it has to fulfil the requirements of the Resource Provider OLA (operational level agreement). [3]

The first requirement for the Resource Infrastructure Provider is to adhere to the Grid Security and Operational Policies and Procedures.

The ROC has to provide a Helpdesk service, either by using the EGI Helpdesk [4] directly, or by using a helpdesk service that can interoperate with the EGI Helpdesk. Implementers can find recommendations for interfacing with GGUS at the GGUS web page [5]. In either case a new Support Unit must be created on the EGI Helpdesk in order to have the ability to assign and identify tickets specific to the ROC. It is the responsibility of the ROC to monitor progress of incident and problem tickets and to ensure that Resource Centres work on tickets opened against them and respond in a timely manner.

The ROC has to operate first- and second-level support. This consists in assisting in the resolution of advanced and specialised operational problems, which cannot be solved by Resource Centre administrators. Resource Infrastructure Provider has to escalate and follow up problems with higher-level operational or development teams.

The site and administrators of the Resource Infrastructure must be registered in the GOCDDB [6], the EGI service and contact registry. The data-scoping feature of the GOCDDB enables the support of multiple registries for different infrastructures holding EGI and non-EGI data. For example an integrated or a peer RP can register its information in the GOCDDB in separate scope than EGI, and create a new separate registry.

The Integrated Resource Infrastructure can choose its own grid middleware with only the requirement being to provide discoverable services using the GLUE 1.3/2.0 schema(s). More information about the implementation of GLUE with EGI can be found in the “EGI Profile for the use of the GLUE 2.0 information schema” [7]. Examples of integrated middleware stacks are ARC, dCache, Desktop Grid, gLite, Globus, QCG, UNICORE.

### *2.3 Africa and Arabia ROC*

Specifically regarding the sub-Saharan Africa the project has facilitated the creation of a single umbrella ROC (together with Arab countries) that coordinates all the activities and that ensures transfer of knowledge and expertise across the wider region. There are 8 Sites running in production-level within the Africa&Arabia ROC. Alongside the South African Grid [8] resources of 8 sites, the Grid infrastructure includes the sites from Egypt, Morocco, Algeria, Senegal, Kenya and Tanzania. The ROC's public website is at [9], and the technical details of AAROC release are at [10] while the ROC blog is at [11]. The status card of the ROC is given in the Figure 2 below.

ROC Africa&Arabia		
Contacts	✓	1. Bruce Becker
Status	✓	8 Sites. The ROC is operational. Registered in GOCDDB
Helpdesk	✓	<a href="https://ggus.eu">https://ggus.eu</a>
Accounting	✓	Accounting records are now published. The ROC runs APEL
Monitoring	✓	Monitoring information is published. The ROC runs SAM-NAGIOS
Website	✓	<a href="http://roc.africa-grid.org">http://roc.africa-grid.org</a>

Fig.2. Africa and Arabia ROC status card

A snapshot of the operational tickets for the ROC is given in the Figure 3 below.

The screenshot shows a ticket management interface with search filters and a list of tickets. The search filters include Ticket ID, Support Unit (AfricaArabia), Priority (all), Type of problem (all), Ticket category (all), MoJ Area, and Status (open states). The list of tickets is as follows:

Ticket-ID	Type	VO	Site	Priority	Resp. Unit	Status	Last Update	Subject
108015	ops	ZA-MERAKA	less urgent	AfricaArabia	assigned	2014-09-01	[Rod Dashboard] Issues detected at ZA-ME...	
107894	ops	MA-01-CNRST	less urgent	AfricaArabia	assigned	2014-08-26	[Rod Dashboard] Issues detected at MA-01...	
107893	ops	MA-01-CNRST	less urgent	AfricaArabia	assigned	2014-08-26	[Rod Dashboard] Issues detected at MA-01...	
107877	ops	ZA-UJ	less urgent	AfricaArabia	in progress	2014-09-01	[Rod Dashboard] Issues detected at ZA-UJ...	
107876	ops	DZ-01-ARN	less urgent	AfricaArabia	assigned	2014-08-26	[Rod Dashboard] Issues detected at DZ-01...	
107875	ops	MA-01-CNRST	less urgent	AfricaArabia	assigned	2014-08-26	[Rod Dashboard] Issues detected at MA-01...	
107874	ops	EG-ZC-T3	less urgent	AfricaArabia	assigned	2014-08-26	[Rod Dashboard] Issues detected at EG-ZC...	
107873	ops	ZA-CHPC	less urgent	AfricaArabia	assigned	2014-08-26	[Rod Dashboard] Issues detected at ZA-CH...	
107872	ops	ZA-MERAKA	less urgent	AfricaArabia	assigned	2014-09-01	[Rod Dashboard] Issues detected at ZA-ME...	
107858	ops	MA-01-CNRST	less urgent	AfricaArabia	assigned	2014-08-28	[Rod Dashboard] Issues detected at MA-01...	
107609	alice	ZA-CHPC	less urgent	AfricaArabia	assigned	2014-08-19	ce is down	
107011	none	ZA-CHPC	very urgent	AfricaArabia	assigned	2014-08-25	CVMFS Client upgrade to version 2.1.19	
106159	none		less urgent	AfricaArabia	in progress	2014-08-19	Sites not publishing accounting data	

Fig.3. Africa and Arabia ROC operational tickets

### 3. Global Cloud Testbed

Cloud computing is an emerging technology aiming to provide increased flexibility as compared to Grid computing.

CHAIN-REDS has built an inter-continental Cloud testbed (Figure 4 below) to demonstrate standard-based cloud interoperability and interoperation and to provide a platform where to run domain-specific and general-purpose applications. CHAIN-REDS Cloud testbed [12] has been organised as a “virtual cloud” made up by the resources belonging to 10 sites from 6 countries, of which, regarding Africa, 1 is a seed Cloud computing site in South Africa and 1 is an SME located in Egypt. 5 out of the 10 sites also belong to the EGI Federated Cloud [13] (including CIEMAT) and 3 different and well-

known cloud stacks are supported, namely Okeanos [14], OpenNebula [15] and OpenStack [16].

Guidelines about how to configure sites in order to be part of the CHAIN-REDS cloud testbed, which is interoperable with the EGI Federated Cloud, are provided in Annexes I, II and III of Deliverable D3.2 [17]. CHAIN-REDS keep these technical guidelines publicly accessible on the project website.

If a region/organisation plans to create a public cloud for research and education and they want to share its computing and storage resources with other clouds available in other parts of the world, for the benefit of the VRCs they support, it is suggested to choose a cloud middleware already supporting the CDMI (Cloud Data Management Interface) and OCCI (Open Cloud Computing Interface) standards. If a region/organisation already owns and operates a public cloud for research and education and chosen middleware is not compliant with CDMI and OCCI standards, it is suggested to create the needed services and endpoints to support those standards.

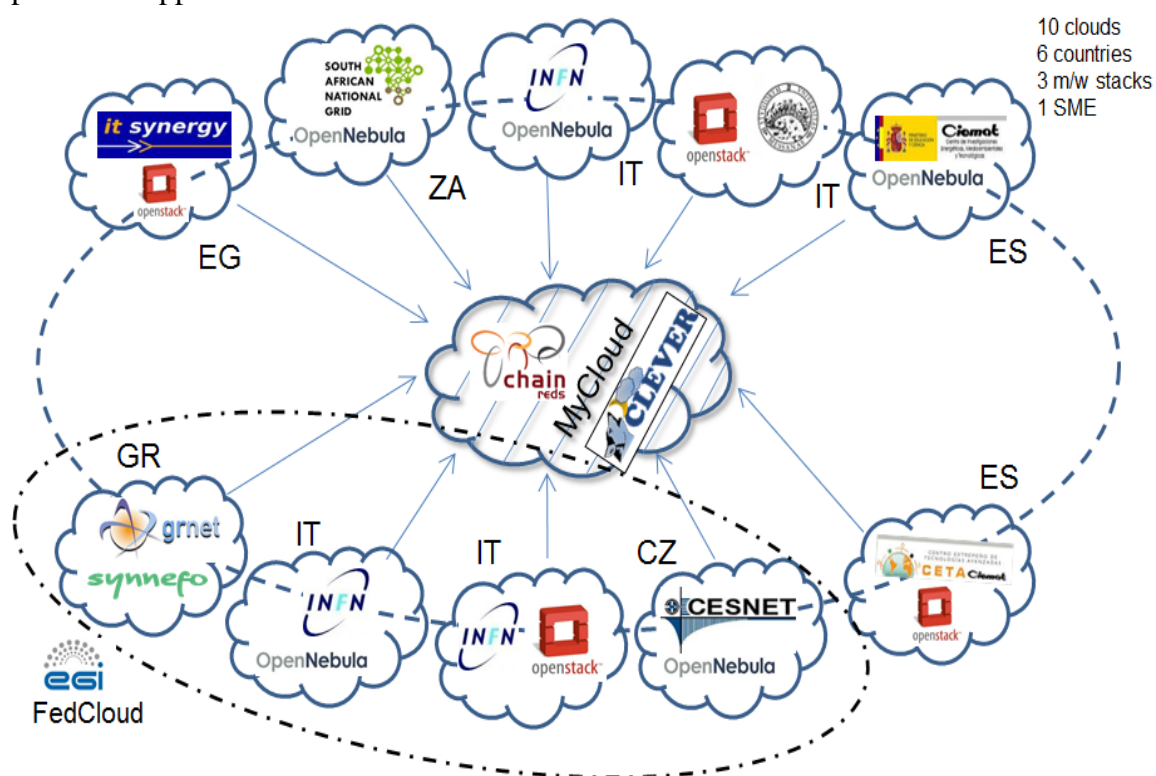


Fig.4. CHAIN-REDS Cloud Testbed

#### 4. Data Management and the APHRC Use-Case

CHAIN-REDS project has also significant developments in managing the data lifecycle. The Knowledge Base (KB) [18] provides information about the deployment of e-Infrastructures per country, as well as links to Data Repositories (DRs), Open Access Document Repositories (OADRs), and Open Access Education Repositories (OAERs) around the world. The Semantic Search Engine [19], allows users to search for any specific term and find the datasets related to it in all the repositories already linked to the KB, but also find potential relationship in terms of authors, subjects and publishers with other repositories. Finally the CHAIN-REDS Persistent Identifiers (PIDs) service [20][21], provides unique, persistent data identifiers to data sets.

In this context, of special interest in the African region is the AHRC case. The African Population and Health Research Centre (APHRC) undertakes research in a wide range of

topics related to societal health and well-being. The APHRC runs around 60 projects, publishes around 60 papers, and trains more than 150 fellows per year. To do so, it counts on 18 donors and 51 partners. The importance of the APHRC's work is contained both in the policy and research documents that it produces, as well as in the raw data sets that are collected through meticulous data collection. These data sets are the bedrock of the APHRC's status of leading institute in the area of population dynamics and social science in the region. The curation, discoverability, dissemination and proper citation of these datasets is important both in academic terms to the researchers involved in collecting them, as well as return-on-investment terms to the institute donors.

APHRC use case involves data management including assignment of Persistent IDentifiers (PIDs) to the APHRC datasets. This is of utmost importance due to these datasets are widely used by almost every country in Africa in order to improve societal health and well-being. APHRC and CHAIN-REDS have first identified which repositories must be catered for and then defined a roadmap for making data labelling: the PIDs are being assigned to an entire data set at the top-level. The software being used to document the data is Nesstar [22].

APHRC has used the CHAIN-REDS access to the EPIC PID issuer, to issue PIDs to the APHRC data sets. This was done by creating a Python script which interacts with the EPIC REST API [21]. The code first scans the APHRC catalogue at runtime and issues PIDs to data sets to which the PIDs have not yet been issued. The code has been published and handed over to APHRC, along with the credentials necessary to execute it, so that future data sets can be associated with PIDs.

The assignment of PIDs to the data sets constitutes a milestone both for the APHRC as well as for the CHAIN-REDS project, since this now provides the ability to the APHRC to track the impact of its work, via data citation networks. Without access to the CHAIN-REDS PID mint, the APHRC would not have had this capability. While this is indeed a success, it could have been arrived at by other means, independently of CHAIN-REDS, albeit with less autonomy and more cost on the part of the APHRC – for example by handing the datasets to third-party repositories which support authoritative PID issuing. The approach proposed by the collaboration between CHAIN-REDS and the APHRC provided a more acceptable means though, since it was done in full co-development with the APHRC, taking into account the needs of the institute and the user communities, as well as taking cognisance of the non-trivial aspects of data accessibility, discoverability and curation implicit in this work. To demonstrate added value in the current collaboration, several follow-on actions are envisioned.

Apart from this case, a number of others from different disciplines are supported by the CHAIN-REDS project, including applications from Latin America, Arab region, China and India.

## 5. Conclusions

CHAIN-REDS provides a platform for e-Infrastructure interoperations of Europe and other continents. A number of distinct achievements were presented.

- First is the model for global Grid interoperation - based on the concept of Regional Operations Centres (ROC) - and its implementation for the African continent. The implementation of the ROC in the African continent is full-blown, with a number of Grid sites operating by providing computing resources, as well as a set of ROC services. The model is fully compatible with Grid operations model in Europe thus enabling cross-continental collaborations. The relevant recommendation is that e-Infrastructure operators in Africa can join their computing clusters under the umbrella of the Africa

ROC - effectively sharing resources with other Grid clusters in Africa and globally and thus increasing the resource access pool available for their users.

- Second, the CHAIN-REDS global Cloud test-bed was presented, which aims to federate global Cloud installations, also integrating 2 sites from Africa. Similar to the Grid, it is recommended that the computing sites in Africa devoted to Research and Education contribute their resources to CHAIN-REDS federated global cloud, paving the way for resource sharing and global collaborations.
- Third, a real-life use-case relevant for Africa, that of APHRC, was presented, which uses specific project developments in data management. Specifically, the CHAIN-REDS Persistent IDentifiers (PIDs) service provides unique, persistent data identifiers to data sets of APHRC, and is available for the other user communities in need of uniquely labelling and identifying their data sets.

The Africa ROC will continue to operate beyond CHAIN-REDS project lifetime, being open to accepting Grid sites from the continent under its federated operational umbrella. The CHAIN-REDS cloud federation will similarly be operational, and the steps are being taken to transfer the operation of the PID service from Europe to an operator within Africa. The barriers of entry into Grid and Cloud federations are lowered for the e-Infrastructure operators which can follow the established guidelines and policies and easily join the global community. Economic benefits of these kinds of computing federations are many-fold, where the common Grid and Cloud operations and resource sharing provide economies of scale, increased hardware and services resource pools, and better structuring in terms of user communities.

## References

- [1] Coordination and Harmonisation of Advanced e-Infrastructures for Research and Education Data Sharing, <http://www.chain-project.eu/>
- [2] European Grid Infrastructure, [www.egi.eu](http://www.egi.eu)
- [3] <https://documents.egi.eu/public/ShowDocument?docid=463>
- [4] <https://ggus.eu>
- [5] [https://ggus.eu/pages/ggus-docs/interfaces/docu\\_ggus\\_interfaces.php](https://ggus.eu/pages/ggus-docs/interfaces/docu_ggus_interfaces.php)
- [6] <https://goc.egi.eu>
- [7] <https://documents.egi.eu/public/ShowDocument?docid=1324>
- [8] <http://www.sagrid.ac.za/>
- [9] <http://roc.africa-grid.org>
- [10] <http://aaroc.github.io/releases/AAROC-in-production/>
- [11] <http://aaroc.github.io/>
- [12] <https://www.chain-project.eu/cloud-testbed>
- [13] <https://www.egi.eu/infrastructure/cloud/>
- [14] [www.oceanos.grnet.gr](http://www.oceanos.grnet.gr)
- [15] <http://opennebula.org/>
- [16] <http://www.openstack.org/>
- [17] CHAIN-REDS deliverable D3.2, “Interoperability guidelines and design” [http://documents.ct.infn.it/record/563/files/CHAIN-REDS-D3.2\\_v0.6.pdf](http://documents.ct.infn.it/record/563/files/CHAIN-REDS-D3.2_v0.6.pdf)
- [18] CHAIN-REDS Knowledge Base, <http://www.chain-project.eu/knowledge-base>
- [19] CHAIN-REDS Semantic Search Engine, <http://www.chain-project.eu/linked-data>
- [20] PID, <http://www.pidconsortium.eu/>
- [21] <http://epic.grnet.gr/guides/api/>
- [22] KLIOS, <http://klios.ct.infn.it>