



GISELA

INFRASTRUCTURE STATUS REPORT

EU DELIVERABLE: D4.1

Document Full name	GISELA-D4.1-v1.5
Date	01/12/2010
Activity	WP4 / NGI - LGI Infrastructure Services
Lead Partner	UFRJ
Document status	APPROVED
Classification Attribute	PU (Public)
Document link	http://documents.gisela-grid.eu

Abstract: This document presents the initial situation of the GISELA infrastructure, to be expanded and perfected during the project.



Copyright notice

Copyright © Members of the **GISELA** Consortium, 2010

GISELA (“Grid Initiatives for e-Science virtual communities in Europe and Latin America”) is a project co-funded by the European Commission as an Integrated Infrastructure Initiative within the 7th Framework Programme. **GISELA** began on 1st September 2010 and will run for 2 years.

For more information on GISELA, its partners and contributors please see www.gisela-grid.eu.

You are permitted to copy and distribute, for non-profit purposes, verbatim copies of this document containing this copyright notice. This includes the right to copy this document in whole or in part, but without modification, into other documents if you attach the following reference to the copied elements: “Copyright © Members of the **GISELA** Consortium, 2010. See www.gisela-grid.eu for details”.

Using this document, in a way and/or for purposes not foreseen in the paragraph above, requires the prior written permission of the copyright holders.

The information contained in this document represents the views of the copyright holders as of the date such views were published.

THE INFORMATION CONTAINED IN THIS DOCUMENT IS PROVIDED BY THE COPYRIGHT HOLDERS “AS IS” AND ANY EXPRESSED OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE MEMBERS OF THE **GISELA** COLLABORATION, INCLUDING THE COPYRIGHT HOLDERS, OR THE EUROPEAN COMMISSION BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THE INFORMATION CONTAINED IN THIS DOCUMENT, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Delivery Slip

	Name	Partner/Activity	Date	Signature
From	WP4	UFRJ / NGI - LGI Infrastructure Services	16/11/2010	R. Diacovo
Reviewed by	Technical Board			
Approved by	Management Board		01/12/2010	B. Marechal Ph. Gavillet S. Jalife Villalón L. A. Trejo Rodriguez R. Barbera R. Ramos Pollán

Document Log

Issue	Date	Comment	Author
0-1	16/11/2010	First draft	R. Diacovo
0-2	18/11/2010	Accounting information added	R. Diacovo
0-3	22/11/2010	F. Brasileiro' s comments incorporated	R. Diacovo, F. Brasileiro
0-4	28/11/2010	D. Carvalho ' s review incorporated	R. Diacovo, D. Carvalho
0-5	01/12/2010	Ph. Gavillet' s review incorporated	R. Diacovo, Ph. Gavillet
1-5	01/12/2010	Final review and approval	B. Marechal

Document Change Record

Issue	Item	Reason for Change

TABLE OF CONTENTS

1. INTRODUCTION	5
1.1. PURPOSE OF THE DOCUMENT	5
1.2. DOCUMENT ORGANISATION	5
1.3. APPLICATION AREA	5
1.4. DOCUMENT AMENDMENT PROCEDURE	5
1.5. TERMINOLOGY	5
1.6. GLOSSARY.....	6
2. EXECUTIVE SUMMARY	7
3. COMPUTATIONAL RESOURCES AND OPERATIONS	8
3.1. RESOURCE CENTRES	8
3.2. CORE SERVICES	10
3.3. OPERATIONS.....	11
4. INFRASTRUCTURE USAGE	12
4.1. REPORTING POLICY.....	12
4.2. STATISTICS.....	12
5. CONCLUSION	14

TABLE OF TABLES

Table 1 – Resource Centres situation in M03	8
Table 2 - Resources Situation Summary in M03.....	10
Table 3 – Core Services situation in M03	10

TABLE OF FIGURES

Figure 1 – Accumulated normalised CPU Hours – gLite (date x VO).....	13
Figure 2 – Accumulated normalised CPU Hours – gLite (date x RC).....	13

1. INTRODUCTION

1.1. PURPOSE OF THE DOCUMENT

This deliverable presents the status of the GISELA infrastructure at the end of the first quarter of the Project. The project milestone MS2, due on M06, marks the date at which the infrastructure shall become operational. This will be eased by the fact that GISELA profits from part of the EELA-2 legacy infrastructure, which will be adapted to GISELA needs and augmented.

For a comprehensive view of the Project and of the GISELA Consortium, the Description of Work¹ (DoW) and the Consortium Agreement (CoA)² should be consulted.

1.2. DOCUMENT ORGANISATION

Section 2 presents the Executive Summary of this document. Section 3 reports on the current situation of the pledged computing resources, comprising both the Resource Centres and the Core Grid Services. It also includes a report on the operation efforts that have been undertaken since the Project started (01/09/2010). Then an accounting report follows in Section 4 that provides an idea about the infrastructure usage from April 2010 (just after the termination of EELA-2) to November 2010. Finally, conclusions are presented in Section 5.

1.3. APPLICATION AREA

The target audience for this document is:

- The members of the Project;
- The European Commission (Project Officer, Reviewers, ...);
- The External Advisory Committee (EAC);
- The general public.

1.4. DOCUMENT AMENDMENT PROCEDURE

Any Project Partner can give comments on, or request amendments to this document. Requests have to be addressed to the Project Coordinator, via the Project Office (hlp-gisela@hlpdeveloppement.fr).

1.5. TERMINOLOGY

In this document, the key words “must”, “must not”, “required”, “shall”, “shall not”, “should”, “should not”, “recommended”, “may” and “optional” are to be interpreted as described in RFC 2119³.

¹ Description of Work (DoW) available upon request to the EELA-2 Project Office (hlp-gisela@hlpdeveloppement.fr)

² Consortium Agreement (CoA) available upon request to the EELA-2 Project Office (hlp-gisela@hlpdeveloppement.fr)

³ <http://www.ietf.org/rfc/rfc2119.txt>

1.6. GLOSSARY

AMGA	ARDA1 Metadata Grid Application
BDII	Berkeley Database Information Index
DoW	Description of Work
EAC	External Advisory Committee
EELA	E-Infrastructure shared between Europe and Latin America
EELA-2	E-science grid facility for Europe and Latin America
EGI	European Grid Initiative
ENMR	Life sciences project (www.e-nmr.eu)
gLite	Lightweight middleware for Grid Computing
GOC	Grid Operation Centre (Regional Operations, GISELA context)
GSC	Grid Support Centre (Central Operations, GISELA context)
IBERGRID	Iberian Peninsula Grid Infrastructure (www.ibergrid.eu)
IGALC	Iniciativa de Grid de America Latina - Caribe
IGI	Italian Grid Infrastructure (www.italiangrid.org)
LA	Latin America
LAC	Latin America and the Caribbean
NGI	National Grid Initiative (EGI Context)
NGI_FR	French National Grid Initiative
OLA	Operational Level Agreement
OurGrid	Opportunistic Grid Middleware (www.ourgrid.org)
RC	Resource Centre
ROC	Regional Operating Centre (EGI context)
ROC_IGALC	The IGALC ROC (www.igalc.org)
ROC_LA	The LA ROC (www.roc-la.org)
VO	Virtual Organisation
VOMS	VO Management Service
VRC	Virtual Research Community
WP3	User Communities Support Work Package
WP4	NGI / LGI Infrastructure Services Work Package

2. EXECUTIVE SUMMARY

The year 2010 saw the end of EELA-2 on March 31st and the launch of GISELA, on September 1st. In 2011, the GISELA infrastructure shall become operational. The period encompassing these three events is very interesting from the point of view of long-term sustainability of e-Infrastructures in Latin America. With National Grid Initiatives in a rudimentary state – if compared to the European ones – how grid entities such as IGALC would fare in maintaining the grid machine working was not completely clear.

With that in mind, this document presents the infrastructure status (size and contributors), as well as what has been done in terms of operation efforts. From the gathered information, it is now possible to conclude that a significant amount of resources has been kept operating during the interval between EELA-2 and GISELA. In addition, usage numbers show that these resources are not only operational, but are being used by scientific communities.

From the operations point of view, the actions taken so far have proven to be adequate, with a reduced team being able to operate the infrastructure up to the time of this writing.

This deliverable also defines an important policy: how to account for committed resources usage on a shared infrastructure. Thanks to this policy, it will be possible to compare the progress of utilisation of the infrastructure during the course of the Project.

3. COMPUTATIONAL RESOURCES AND OPERATIONS

This section presents the current stage of the GISELA infrastructure. Section 3.1 depicts the status of the process of integration of Resource Centres. The Core Services situation is then presented in Section 3.2 and the Operation Efforts in Section 3.3.

3.1. RESOURCE CENTRES

The Resource Centres are responsible for providing CPU power and storage to the infrastructure.

Table 1 depicts the situation of the infrastructure at M03. With the exception of Brazil and Mexico, the countries participating in WP4 have only one partner institution. Thus, having the first column referring to *Country* instead of the partner makes it easier to understand the resources distribution. The second column (*Pledged Resources*) is a mere reproduction of the commitments from the DoW.

The *Resource Centres* and *Resource Availability* columns derive from a survey conducted by WP4 in order to identify the actual distribution of the pledged resources within the participating countries, which was not known beforehand. It is important for scheduling the integration: the effort to integrate several small RCs that sum up n CPU cores is greater than the effort to integrate a single RC that provides the same n cores. Some partners, however, have not yet provided the information concerning the Resource Centres in their countries and some RCs have not yet provided the resources availability information within their institutions. This unresponsiveness has already been escalated to the GISELA Technical Board. On the other hand, WP4 profits from good relationships with previous Project (EELA and EELA-2) member institutions, as the portfolio counts on some RCs that have agreed to provide resources to some GISELA-supported VOs, even not being official GISELA partners. For RCs operated by third parties, the information was retrieved directly from the grid information systems.

Moving up to the *Status* column: *N/I* stands for “Not Integrated”, and *I* stands for “Integrated”. A given RC is considered integrated at this stage if, and only if, at least one of the following conditions is satisfied:

1. It is already integrated into EGI, no matter the grid initiative operating it;
2. It is properly configured in the OurGrid community.

The first definition is a temporary one, made necessary due to the fact that the operation of the “catch-all” VO (prod.vo.eu-eela.eu) has not yet begun. Once done, it is likely that the support for this VO will be used as a delimiter between “Integrated” and “Not Integrated” RCs.

Finally, the *Notes* column and its contents should be self-explanatory, and reflect the differences between the DoW and the reality in M03.

Table 1 – Resource Centres situation in M03

<i>Country</i>	<i>Pledged Resources</i>	<i>Resource Centres</i>	<i>Resource Availability (M03)</i>	<i>Status (M03)</i>	<i>Notes</i>
Argentina	130 CPUs	HIBA	8 CPUs	N/I	2 extra RCs
		UBA	8 CPUs	N/I	
		UNLAM	Not informed	N/I	
		UNLP	8 CPUs	I	
		UNRC	16 CPUs, 0.5 TB	N/I	
		UNS	16 CPUs, 0.6 TB	N/I	

Country	Pledged Resources	Resource Centres	Resource (M03)	Availability	Status (M03)	Notes
Brazil UFRJ	762 CPUs 21 TB	UNSL	Not informed		N/I	
		CEFET-RJ	22 CPUs, 0.5 TB		I	DoW ammendment necessary (2 RCs instead of 4)
		UFRJ-IF	212 CPUs, 11 TB		I	
Brazil UFCG	450 CPUs 4 TB	EELA-UFCG	8 CPUs, 1 TB		N/I	
		LCC2	74 CPUs		I	
		GMF	6 CPUs		I	
		DCA	36 CPUs		I	
		AESA	36 CPUs		I	
		LCC1	78 CPUs		I	
		LSD	76 CPUs		I	
Chile	0	UTFSM ⁴	54 CPUs		I ⁵	1 extra RC
Colombia	200 CPUs 2 TB	UNIANDES	134 CPUs		I ⁵	4 RCs to be added
		UIS	20 CPUs, 1 TB		N/I	
Cuba	50 CPUs	CUBAENERGIA	60 CPUs		N/I	
Ecuador	100 CPUs 1 TB	Not informed	-		-	5 RCs to be added
France	40 CPUs	Not informed	-		-	1 RC to be added
Italy	100 CPUs 30 TB	INFN-CATANIA	100 CPUs, 14 TB		I ⁶	
Mexico	198 CPUs 2 TB	CICESE-GRID	16 CPUs		N/I	3 RCs to be added
		ICN-UNAM	138 CPUs		I ⁵	
		ITESM	16 CPUs		N/I	
		Peyote-Grid	16 CPUs		N/I	
		UAEM-GRID	16 CPUs		N/I	
Panama	100 CPUs 1 TB	CIDETYS	Not informed		N/I	1 RC to be added
Peru	110 CPUs 8 TB	Not informed	-		-	7 RCs to be added
Portugal	100 CPUs	UPorto	90 CPUs, 1 TB		I ⁷	1 RC withdrawal
		UMinho	16 CPUs, 0.1 TB		I ⁷	
Spain	100 CPUs 20 TB	CETA-CIEMAT	112 CPUs		N/I	1 extra RC
		CIEMAT-TIC	90 CPUs, 35 TB		I ⁷	
		UNICAN ⁴	13 CPUs, 0.2 TB		I ⁷	
Uruguay	100 CPUs 10 TB	UdelaR	Not informed		N/I	
Venezuela	120 CPUs 6 TB	GRyDs-USB	2 CPUs		I	1 RC to be added
		ULA-MERIDA	48 CPUs, 1 TB		I	

⁴ Though not a GISELA partner, the RC administrator has agreed to accept some of the GISELA-supported VOs.

⁵ This site is operated by a non-GISELA entity: ROC_LA

⁶ This site is operated by a non-GISELA entity: NGI_IT

⁷ This site is operated by a non-GISELA entity: IBERGRID

Table 2 aggregates the information presented, providing an overall view of the pledged computational power status. One should note that although there are some *missing* resources, according to the DoW metrics, WP4 has until M12 to integrate about 800 more CPU cores, and then until M24 to integrate the remaining ones.

Table 2 - Resources Situation Summary in M03

<i>Status</i>	<i>CPUs</i>	<i>Storage (TB)</i>	<i>Resource Centres</i>
Pledged	2660	105	56
Integrated	1233	~ 63	21
Not Integrated	312	~ 3	15
Missing	1115	~ 39	20

3.2. CORE SERVICES

The core services encompass the underlying machinery required to run the prod.vo.eu-eela.eu VO services to users, as well as ROC_IGALC services for the operation of the Resource Centres that are under it. Table 3 illustrates the core services currently (M03) operational.

Table 3 – Core Services situation in M03

<i>Country</i>	<i>Resource Centre</i>	<i>Core Service</i>	<i>Scope</i>
Brazil	UFCG	Discovery Service	OurGrid
		Status Service	OurGrid
		Statistics Service	OurGrid
	UFRJ	Workload Management System	prod.vo.eu-eela.eu VO
		Proxy Server	prod.vo.eu-eela.eu VO
		Logical File Catalog	prod.vo.eu-eela.eu VO
		Logging and Bookkeeping	prod.vo.eu-eela.eu VO
		Top-Level BDII	prod.vo.eu-eela.eu VO
		User Interface	prod.vo.eu-eela.eu VO
		AMGA Server	prod.vo.eu-eela.eu VO
		Management System	prod.vo.eu-eela.eu VO
		Monitoring System	prod.vo.eu-eela.eu VO
		Central Storage Element for Monitoring	prod.vo.eu-eela.eu VO
		Top-Level BDII	IGALC
		Monitoring System	IGALC
Central Storage Element for Monitoring	IGALC		
Spain	CETA-CIEMAT	Workload Management System	prod.vo.eu-eela.eu VO
		Logging and Bookkeeping	prod.vo.eu-eela.eu VO
		User Interface	prod.vo.eu-eela.eu VO

3.3. OPERATIONS

This section highlights the WP4 efforts towards the seamless operation of the infrastructure described in sections 3.1 and 3.2, as well as the integration of the pledged resources.

As foreseen in the DoW, the Grid Support Centre (GSC) is currently centralising a number of responsibilities, until some of them can be delegated to the Grid Operation Centres (GOCs). The list of these responsibilities follows:

- *Resource Centre integration efforts:* a survey was conducted to raise the resources geography (see Section 3.1) and support is being provided to the first Resource Centres being integrated;
- *IGALC operations:* the GSC personnel is currently in charge of the OPS VO shifts – OPS is the monitoring system exclusive VO. The GSC also follows-up eventual tickets, providing support to the IGALC sites and handling the usual operation tasks;
- *Establishment of partnerships with other grid initiatives in LA:* WP4 is currently holding meetings with ROC_LA, in an attempt to establish agreements. The objective is to maximise the integration between the two bodies, thus providing higher quality services to the VRCs;
- *Requests from WP3 are being handled:* the enmr.eu VO, from the e-NMR project, is soon to be supported in several GISELA Resource Centres;
- *Two documents are currently under elaboration:* the IGALC Operational Level Agreement (OLA) for service grid Resource Centres and the GISELA Operations Manual, which defines the GISELA infrastructure. These two documents are very important, as they will provide to the involved parties an exact and clear understanding of the requirements, obligations and metrics in daily operations.
- *As part of the GISELA infrastructure definition,* the policy for claiming GISELA contribution to the number of grid jobs run has been defined (more details in Section 4). This is important for maintaining consistency throughout the future project Deliverables.

With the prod.vo.eu-eela.eu Nagios (Monitoring system) now operational, all what remains is to train at least one GOC team for the operations to officially begin. The training of the first GISELA GOC is currently in its initial phase at UFRJ. It is expected that the team will be ready by January 2011.

4. INFRASTRUCTURE USAGE

This section shows some numbers related to the usage of the infrastructure encompassing the first three months of GISELA and some months before its beginning. Although is it too early to judge VRC usage by such numbers, they certainly can be used as the baseline for future reports. The current numbers can be found in Section 4.2.

Before that, it is necessary to define a policy for reporting the infrastructure usage. This need derives from the fact that most computational resources in the project are not exclusively dedicated to GISELA, in the sense that they are shared with various other projects. WP4 and WP3 have agreed on this policy, which is described in Section 4.1.

4.1. REPORTING POLICY

For a given VO v , in the set V of GISELA-supported VOs, and a RC r , in the set R of GISELA Partner Resource Centres, let $T_{r,v}$ be the amount of CPU power dedicated by r to v during a given reporting period. Let $W_{r,v}$ be defined as:

$$W_{r,v} = \begin{cases} 1 & \text{if } v \text{ is a VO operated by GISELA} \\ \frac{p_r}{q_r} & \text{for all other VOs,} \end{cases}$$

where p_r stands for the number of pledged CPU cores engaged for GISELA by r , and q_r is the total CPU count in r . The GISELA contribution to the global grid during the given reporting period may then be defined as follows:

$$T = \sum_{\forall r \in R} \sum_{\forall v \in V} W_{r,v} \times T_{r,v}$$

where T may be defined either in terms of executed jobs or CPU hours.

For Resource Centres that have not yet integrated all the pledged resources, situations where $w > 1$ may appear. In these cases, $w = 1$ is assumed.

4.2. STATISTICS

Following the policy described, the figures in this section depict the infrastructure usage between April and November 2010. One should note that the data for November 2010 are incomplete, since they were extracted before the end of the month.

Figure 1 and Figure 2 contain the gLite infrastructure usage numbers for the above-mentioned period. Both relate to the same jobs – which justify the fact that they have the same shape. The first one provides the breakdown per VO, while the second exposes the usage breakdown by Resource Centre.

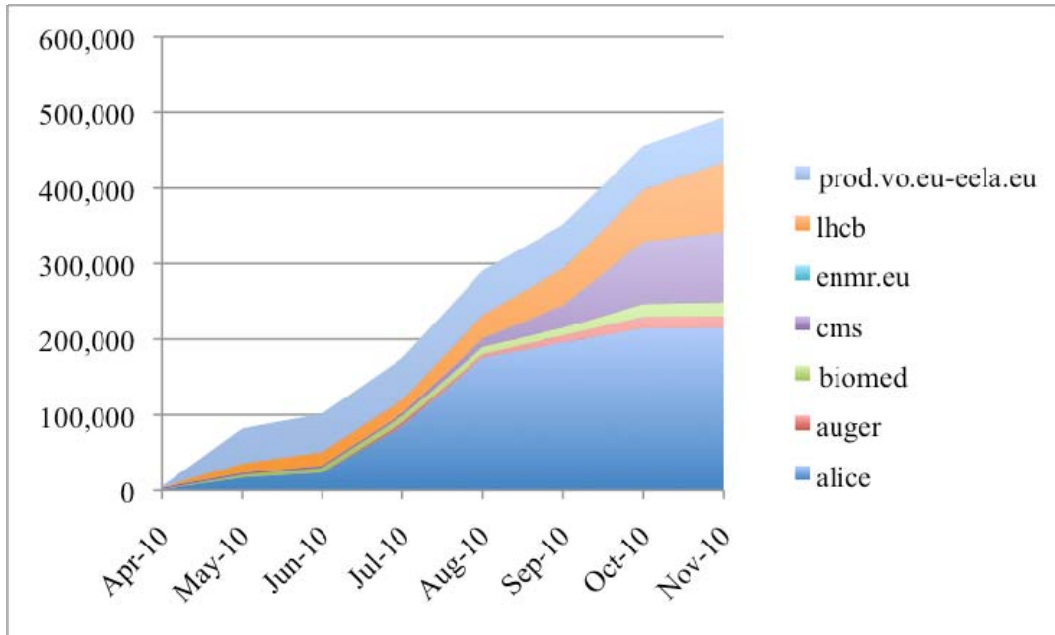


Figure 1 – Accumulated normalised CPU Hours – gLite (date x VO)

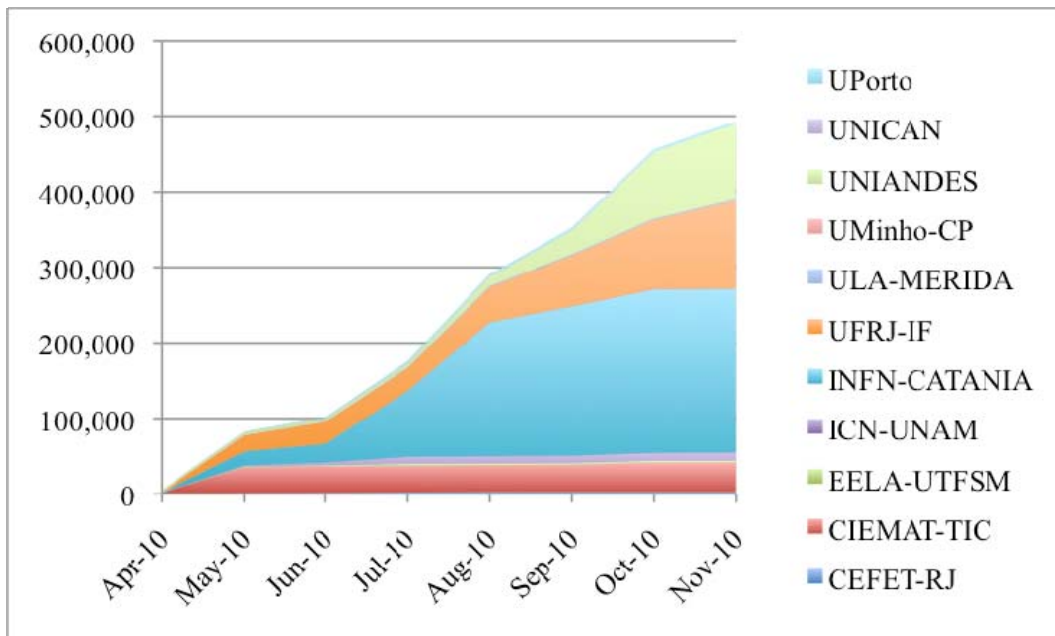


Figure 2 – Accumulated normalised CPU Hours – gLite (date x RC)

5. CONCLUSION

It is worth noting that the WP4 work in GISELA is not beginning from scratch, since there was a significant amount of resources ready and available to the VRC users very early in the project (EELA-2 heritage). This benefits GISELA as a whole, since new users do not have to wait for the establishment of GOCs or any other organisational structure to try their hands at grid. Furthermore, the availability of RCs indicates a certain level of middleware expertise within the Project, which will be helpful for the establishment of the GOCs, as well as for the operation metrics compliance.

The GSC role as a catchall GOC has been proven to be effective, as it has been able to handle the initial tasks satisfactorily. This does not eliminate the need for Regional Structures. Indeed, they will have to disseminate grid knowledge in LA and the GSC will not be able to operate the entire infrastructure, as it grows further.

Finally, the infrastructure usage (~500,000 CPU hours) confirms that the VRCs are profiting from the infrastructure computational power, even though many Resource Centres have not yet been integrated.