Life Sciences Specialized Support Center

V. Breton (CNRS-IN2P3)

breton@clermont.in2p3.fr

Version 1.1 (17 June 2009)

1 Background

This report describes a proposal for the Life Sciences Specific Support Center. This is a working document that will evolve with the discussions between the different stakeholders.

As described in the SSC guidelines document edited by Cal Loomis, the SSC must act as a focal point of a community's use of the European Grid Infrastructure. The SSC must provide stability and continuity for its community safeguarding the expertise of the community vis-à-vis the grid infrastructure and grid technology. As a long-term body, it may be appropriate for short-term projects to use the SSC as a repository of the results and knowledge of those projects to ensure that they are accessible to the SSC's user community after the projects have stopped and to later projects that can build on those projects.

The document is organized as follows:

- chapter 1 provides some background information.

- The following chapters describe the Life Sciences Specific Support Center (LSSSC) according to the recommendations of ref. 5: chapter 2 describes the user community, chapter 3 proposes a governance model, chapter 4 describes a work plan and the corresponding budget, chapter 5 discusses non-human resources and chapter 6 discusses links to EGI.

- In annex, the information collected from the NGI contacts points are provided.

1.1 Meetings

The current state of planning for the Life Sciences Specialized Support Center is the result of work within the EGEE Life Sciences cluster in collaboration with EGEE NA4 management and the EGI_DS project. A set of relevant past and future meetings can be found in Table 1 with Life Sciences specific meetings listed in red.

Table 1: Meetings and Workshops

WS 1	17 Dec. 2008	User Community Transition from EGEE to EGI (Orsay) http://indico.cern.ch/conferenceDisplay.py?confId=45736
LS WS1	5 March 2009	Life Sciences SSC session at EGEE User Forum 4
WS 2	5-6 May 2009	SSC Workshop: Preparing SSCs for EGI (Athens) http://indico.cern.ch/conferenceDisplay.py?confId=55097
LS WS 2	29 June 2009	Life Sciences SSC workshop at HealthGrid conference
WS 3	1 Jul. 2009	SSC Workshop (Orsay) http://indico.cern.ch/conferenceDisplay.py?confId=60563

WS 4	Sep. 2009	SSC Session at EGEE'09 (Barcelona)
		http://indico.cern.ch/conferenceDisplay.py?confId=55893
LS WS3	Sep. 2009	Life Sciences SSC workshop at EGEE'09 (Barcelona)
WS 5	3 Nov. 2009	SSC Workshop (CERN)
		http://indico.cern.ch/conferenceDisplay.py?confId=60564

1.2 Relevant Documents

A set of documents useful for the preparation of the Life Sciences SSC can be found in Table 2.

Table 2: Reference Documents

Ref. 1	EGI Blueprint
	http://web.eu-egi.eu/fileadmin/public/EGI_DS_D5_4_V300b.pdf
Ref. 2	Preliminary SSC Documents
	https://edms.cern.ch/document/983514/1
Ref. 3	EGI Functions Document
	http://knowledge.eu-egi.eu/knowledge/index.php/D32
Ref. 4	e-Infrastructures in FP7: Call 7 (WP2010)
	ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/e-infrastructure/e-infrastructures-in-fp7-call7_en.pdf
Ref. 5	SSC guidelines by C. Loomis
	https://edms.cern.ch/document/989620/1

1.3 Acknowledgements

We would like to acknowledge Diana Cresti important contribution to the preparation of the Life Sciences SSC. We also would like to thank Cal Loomis and Vangelis Floros for providing a framework for the preparation of the FP7 call.

2 LSSSC scientific user community

The LSSSC user community is the one which is targeted by EGEE-III Life Sciences cluster and covers all scientific activities relevant to biology and medicine. Part of the LSSSC user community extensively uses and deploys on the grid computational chemistry software (docking, Molecula Dynamics) for in silico drug dis-

covery which creates a potentially significant overlap between Life Sciences and Computational Chemistry SSCs.

2.1 Scientific scope

The LSSSC aims at enabling researches for understanding the laws of life and improving healthcare in developed countries but also at using the grid to address digital divide and improve Research and Development against tropical and neglected diseases.

LSSSC covers the scientific perimeter of the following ESFRI design studies :

- BBMRI, European Biobanking And Biomolecular Resources
- EATRIS, The European Advanced Translational Research Infrastructure
- ECRIN, Infrastructures For Clinical Trials And Biotherapy
- ELIXIR, Upgrade Of European Bioinformatics Infrastructure
- INFRAFRONTIER, Infrastructure for Phenomefrontier and Archivefrontier
- INSTRUCT, Integrated Structural Biology Infrastructure

These 6 ESFRI projects are complementary to each other as illustrated on figure 1 and need extensive distributed computing and storage resources.

Relevant to the LSSSC activity is also the field of biodiversity where the Life-Watch design study is also exploring the opportunities for using grid technology.
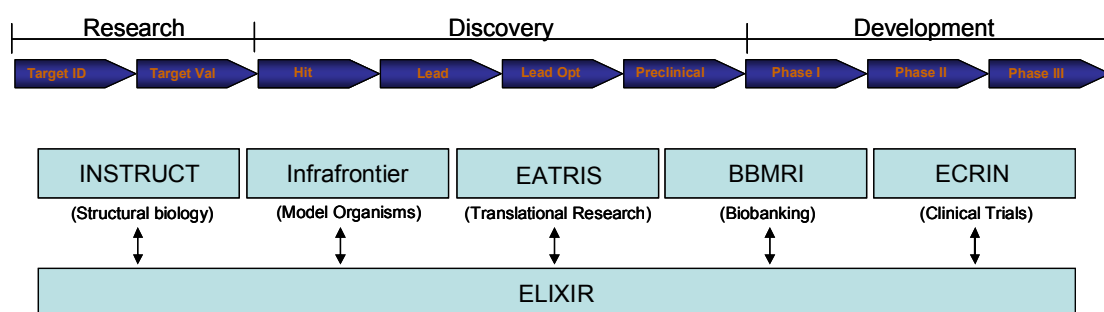
Research		Discovery				Development		
Target ID	Target Val	Hit	Lead	Lead Opt	Preclinical	Phase I	Phase II	Phase III

INSTRUCT	Infrafrontier	EATRIS	BBMRI	ECRIN
(Structural biology)	(Model Organisms)	(Translational Research)	(Biobanking)	(Clinical Trials)

ELIXIR

Figure 1: ESFRI design studies fitting into LSSSC scientific scope

2.2 Geographic scope

The LSSSC geographic scope is the world:

- drug discovery in the field of emerging and tropical diseases can not be conducted without the involvement of scientific collaborators from Africa, America and Asia.
- Public health grid projects must be designed at a worldwide level
- Addressing the digital divide requires setting-up and developing collaborations beyond Europe

In Europe, the LSSSC aims at reaching out to every country in Europe because they all are pursuing researches in the field of life sciences which require at least access to molecular biology databases.

2.3 Aims

The LSSSC goals are the following:

1. to offer services to the Life Sciences community for accessing EGI resources in a user-friendly way

2. to support the services offered by National Grid Initiatives to the Life Sciences community

3. to foster the adoption of grid technology by Life Sciences ES-FRI Research Infrastructures

4. to foster the adoption of grid technology at a world wide level for Public Health and for Research and Development on tropical and neglected diseases

We are now providing details on the strategy to achieve this goal.

2.3.1 Services to access EGI resources in a user friendly way

Contrary to the High Energy Physics community, the Life Sciences community is very heterogeneous in terms of computer skills. Beside highly skilled experts and excellence laboratories in Bio, Chemo and Medical informatics, the vast majority of healthcare professionals and biologists are not going to use the grid unless all technical complexities are completely hidden.

Existing grid portals like GENIUS and P-Grade are far too complex for these end-users. Designing very user-friendly interfaces to the grid is very challenging:

- getting a grid certificate is already a not so obvious thing

- each community has its own culture and there is a need to customize specific interfaces to specific projects or target audiences (biologists, healthcare professionals).

- EGI is going to federate resources operated by different middlewares. Access to these resources in a transparent way requires designing specific APIs.

Building upon the work done in several FP6 and FP7 European projects (Health-e-Child, NEUGrid, EGEE-II, EGEE-III, EMBRACE, OMII-Europe) as well as national projects, the strategy is to analyze the requirements and the state of the art of grid interfaces in order to design scientific gateways using a Service Oriented Architecture and a technology that allows customizing the interfaces.

The method to design scientific gateways as well as their template will be relevant to many other SSCs but also to the ESFRIs when they will explore the opportunities to use EGI resources.

2.3.2 Support to National Grid Initiatives

National Grid Initiatives have reached different levels of maturity as well as adoption by the Life Sciences community. The LSSSC can play a significant role in supporting their efforts:

- to reach out to new users, by offering the services described in the previous paragraph for easy access to EGI and sharing the technology and tools of the scientific gateways

- to train Life Sciences users, by organizing regular biomedical grid schools in collaboration with relevant partners (training SSC)

- to share good practice and learn from other more successful NGIs by setting up the collaboration tools between the NGIs

- to connect to ESFRIs by acting as a mediator between NGIs and ESFRIs

2.3.3 Adoption of grids by ESFRIs

Most of the ESFRI design studies shown on figure 1 are currently evaluating if and how grid infrastructures can provide the computational resources required to reach their research objectives. All of them have planned for significant investment to buy hardware.

Convincing ESFRIs to adopt grids and to extensively rely on EGI resources involves a number of challenges:

- they are afraid to lose money because they will get less money for hardware

- they are afraid of losing control over the resources if they are just customers of EGI or even if the resources are shared within EGI

- they are reluctant to move to a technology that is still hard to use

The strategy to involve them is the following:

- provide information on what EGI is and is not

- identify pilot applications on EGI where grid impact could be demonstrated

- Identify key partners in the EGI who can act as mediators

- Provide training and user support

- Help customize the ESFRI scientific gateway

2.3.4 Promotion of grid for public health and to address digital divide

Public health as well as research on tropical diseases is coordinated at an international level. Grids provide a great added value for alerting of pandemic risk and monitoring epidemics as well as supporting research and development for tropical diseases. Collection of data on the field and provision of services to the researchers in the developing countries are key elements of the international agencies (WHO, Gates foundation) strategies to fight infectious diseases.

The LSSSC strategy to achieve the adoption of grids to fight these world-wide threats includes the following actions:

- promote the grid technology to international agencies
- foster the design and deployment of pilot applications to improve the current practices for epidemics monitoring
- organize biomedical grid schools in developing countries with relevant partners (regional grid projects, training SSC) to create a user community

2.4 Legitimacy

The LSSSC is about setting up and operating services for the Life Sciences community. All LSSSC partners have a track record of operating services for the Life Sciences community. Not all of them are doing science in the field of Life Sciences but all have been enabling research in that field. As the Specific Support Center is dedicated to operating services, the partners have been selected on their experience with the life sciences community.

The LSSSC must have legitimacy with respect to its stakeholders which are:

- the major research institutes in the field of life sciences and healthcare
- EGI.org
- the National Grid Initiatives
- the Regional Grid initiatives
- the EU-funded projects falling into the LSSSC scientific field
- the ESFRI research infrastructures

3 Governance model

3.1 Boards

3.1.1   Overview board

LSSSC overview board is made of the structures funding the SSC:

- Representatives of the research institutes in the field of life sciences and medicine supporting LSSSC activities
- Representatives of the National Grid Initiatives supporting LSSSC activities
- Representatives of the Regional Grid initiatives supporting LSSSC activities
- Representatives of the EU-funded projects supporting LSSSC activities
- Representatives of the ESFRI research infrastructures supporting LSSSC activities

Members of LSSSC overview board are revised on a yearly basis depending on the renewed commitment.

3.1.2 Executive board

LSSSC Executive Board is made of the institutes mandated by the Overview Board to operate the services offered by the LSSSC. In a transition period, they are partners of the LSSSC European Project funded to set-up and start operating specific services to reach the goals listed in chapter 2. Their commitment to the SSC translates into a financial contribution to the SSC operation as a contribution to the European Project where LSSSC is involved. As a consequence, they are also involved in the Overview Board.

3.1.3 Consultative board

LSSSC consultative board is made of the structures willing to contribute to the SSC activities without making resources available to the LSSSC:

- the research institutes in the field of life sciences and medicine supporting LSSSC activities
- the National Grid Initiatives supporting LSSSC activities
- the Regional Grid initiatives supporting LSSSC activities
- the EU-funded projects supporting LSSSC activities
- the ESFRI research infrastructures supporting LSSSC activities

3.1.4 Informed observers

A circle of informed observers involves the potential stakeholders which are not yet involved:

- the major research institutes in the field of life sciences and medicine without an LSSSC representative
- the National Grid Initiatives without an LSSSC representative
- the Regional Grid initiatives without an LSSSC representative
- the EU-funded projects without an LSSSC representative
- the ESFRI research infrastructures without an LSSSC representative

3.2 LSSSC management

The LSSSC will start with the same institutes involved in the Overview and Executive board. In the period of transition covered by the EU funding, the services will be developed and set up: their relevance will be demonstrated so that, by the end of the project, stakeholders willing to support the services will be identified.

4 Work Plan/Budget

4.1 LSSSC activities

Activity name	Tasks	Resources required in FTE
NA1 - coordination	Management of EU project Relationship with EGI (User Forum representative) Relationship with NGIs Relationship with ESFRIs Relationship with major research institutes in the life sciences community Quality control	2
NA2 – dissemination	Reach out to the potential stakeholders (see 3.1.4) Coordination with EGI dissemination SSC Maintenance of knowledge base	1.5
NA3 – training	Coordination with EGI training SSC Organization of yearly biomed Grid schools Coordination with regional grid infrastructures for organizing biomed grid schools in developing countries	1
SA1 – resource operation	Virtual organization management Relationship with EGI resource operation (Grid planning officer to participate in EGI MCB meetings) Relationship with NGIs resource operation	1
SA2 – operation of data and access services	Science gateway operation – Maintenance of data repositories – collaboration with ESFRIs for service provision	2
SA3 – user support	User support – Application porting - Coordination with EGI User Support and Application Porting SSC	2
JRA1 - scientific gateways	Scientific gateways – Design of template and development	2
JRA2 – pilot	Pilot applications for medical development in de-	1

applications	veloping countries	
TOTAL		12.5 FTE

4.2 List of deliverables and milestones for the first 15 months

A first list of deliverables is proposed for the first 15 months of the project.

Deliverable or milestone	Deadline	Content description
MSA1.1	PM1	Biomed VO operational
DNA1.1	PM2	Detailed workplan
DNA2.1	PM3	List of requirements for LSSSC web site - Dissemination strategy to potential stakeholders
DJRA1.1	PM3	List of requirements and technical design of second generation scientific gateways
MSA2.1	PM3	Kick-off of first generation Scientific Gateway
DNA2.2	PM9	LSSSC web site
DJRA2.1	PM9	List of requirements for selected pilot application for medical development
DJRA1.2	PM12	First template of second generation scientific gateway
MSA2.1	PM15	Deployment of second generation scientific gateway

4.3 Budget

The budget for running the activities described in section 4.1 is roughly estimated to 1M€ per year, corresponding to 12.5FTE at an average 80.000€/year.

5 Non-Human Resources

Non human resources required	Source providing the resources	Allocation policy

Computational resources	Biomed Virtual Organization: >20.000 CPUs worldwide	Grid planning officer in collaboration with NGI representatives
Data resources	No data resource at the beginning - To be discussed with ESFRIs on a case by case basis	Free access to the data resources
Collaborative resources (Indico, mailing lists, scientific portals, chat servers, knowledge base)	CNRS – HealthGrid	Free access to the collaborative resources
Yearly community meeting	HealthGrid is organizing an annual conference	
Yearly community training school	ITB-CNR is organizing a yearly biomed grid school	

LSSSC builds upon the work done by several projects and on the resources already made available to the life sciences research community:

- The Biomed Virtual Organization is providing access to more than 20.000 CPUs worldwide. Its access is freely granted to users from the life sciences community. It is also open to users from adjacent communities (computational chemistry, biodiversity) and should remain so.
- The HealthGrid association is already proposing collaborative resources to the life sciences community (mailing lists, knowledge base)
- The HealthGrid conference organized yearly since 2003 will be a relevant yearly event for the LSSSC community as it reaches out to the same community
- The biomed grid school organized yearly by ITB-CNR since 2007 will be a relevant yearly training event for the LSSSC community

6 Sustainability Plan

LSSSC sustainability plan builds upon the experience of the HealthGrid initiative. The HealthGrid association, funded in 2003, aims at fostering the adoption of grids for health. Since 2003, it has been steering up the community through the provision of services on its web site (knowledge base) and the organization of events like the HealthGrid conference.

6.1 Links for EGI

Links for EGI are named people acting as SSC/EGI contact points who are not funded by the EC but by the NGIs supporting the LSSSC.

7 Annexes

7.1 Life Sciences research activities on National Grid Infrastructures in Europe

This section builds upon the data collected by the NGI Life Sciences contact points identified

Country	Statistics on life sciences users and resource usage	Main life sciences grid projects	Most active players in the field
Finland	20% of CSC users 4% of CSC resources	EMBRACE	CSC
France	16% of EGEE biomed VO resources (see section 3.2)	ACGT, *AGIR*, BioinfoGRID, *Décrypthon*, EGEE, EMBRACE, ELIXIR, *GWENDIA*, NeuGrid, *Neurolog*	CNRS (Clermont-Ferrand, Lyon, Nice), INRIA, HealthGrid
Greece	240 CPUs	ACGT, ELIXIR	IBRB/NHRF
Italy	12% of EGEE biomed VO resources (see section 3.2)	BioinfoGRID, EGEE, e-NMR, EMBRACE, EU-ChinaGrid, INSTRUCT, *LIBI*, *LITBIO*, NeuGrid	CNR-ITB, CIRMMP – Firenze, Consorzio Cometa, Fatebenefratelli Brescia, Bologna Biocomputing Group, University of Roma3, University of Modena,
Netherland	3590 CPU cores 34 M CPU hours	e-NMR, *LSGRID*, NeuGrid, *VLEMED*, VIROLAB	Free University Medical Center, University of Amsterdam, Utrecht University
Switzerland	-	Swiss Bio Grid, SystemsX.ch,	Swiss Institute of Bioinformatics, University Hospital of Geneva

7.2 Status of Biomed Virtual Organization on EGEE

The Biomed Virtual Organization represents about 3% of the overall EGEE activity in terms of number of users, normalized CPU time used and number of jobs submitted.

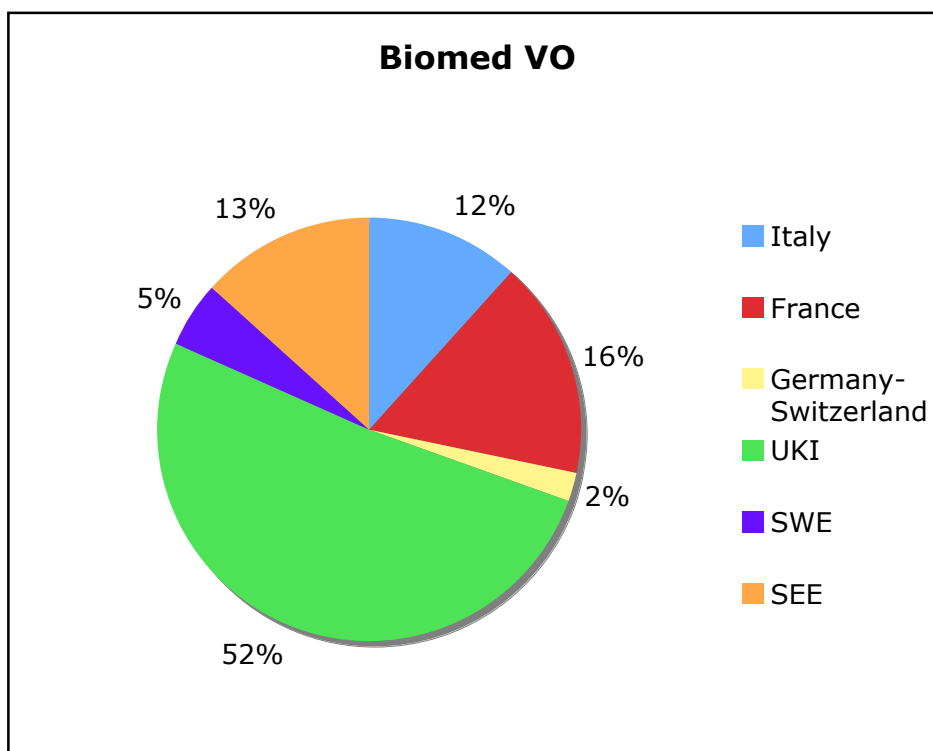| Parameters	Parameter value	Fraction of EGEE activity
Number of users	389	2,99%
Normalized CPU time in the last 12 months (June 08 – May 09)	12,558,680 units 1K.SI2K.Hours	3.39%
Number of jobs submitted in the last 12 months (June 08 – May 09)	3,035,350	2,85%

Figure 1: relative weight of the 7 regions providing the largest computing resources to the Biomed Virtual Organization