

A Specialized Support Centre for Large International Grid Communities: **aka HEP SSC**

Jamie Shiers

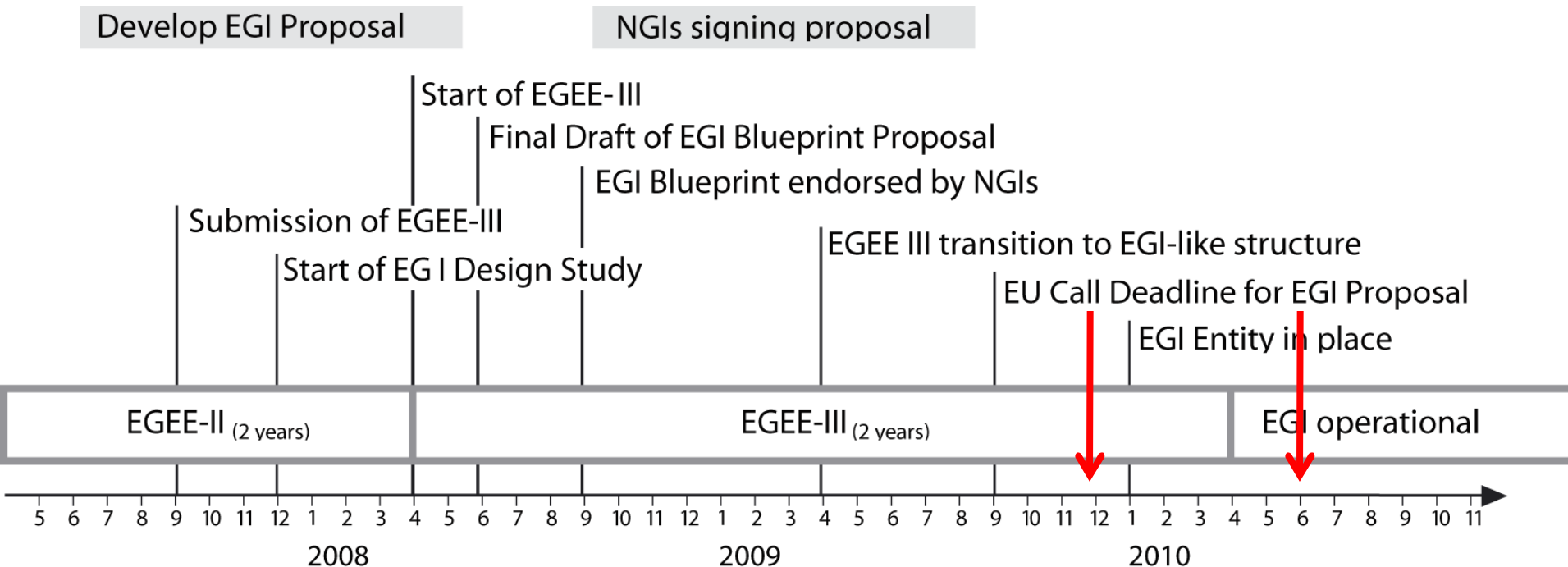
Grid Support Group

IT Department, CERN

From the EGI_DS DoW...

↳ **The establishment of EGI is guided by two basic principles:**

- 1. Build on the experience and successful operation of EGEE and related projects**
- 2. Make EGI operational before EGEE III ends**

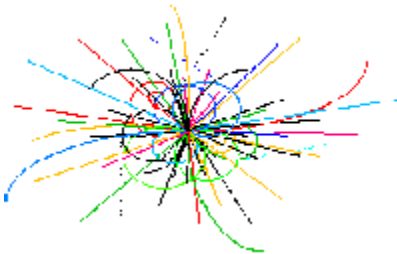


SSCs: A Brief History

- EGI_DS “[Blueprint](#)” document describes potential role of “Specialised Support Centres”
- (Somewhat) within the context of EGEE NA4, several preparation meetings have been held
- Most recently: May in Athens, Paris in July
 - See [Indico](#) for agendas and presentations
- In June there was an [Information Day](#) in Brussels which clarified the specific areas targetted by this call – as well as possible funds
- More information on “HEP SSC” was given to recent [OB meeting](#) & WLCG MBs...

What is “HEP SSC”

- Google finds:



THE SUPERCONDUCTING SUPERCOLLIDER PROJECT

The Superconducting Super Collider Laboratory was a DOE supported facility. The U.S. House of Representatives decided in 1993 to halt the project after 14 miles of tunneling were completed and two billion dollars spent. The laboratory, located South of Dallas, TX, is in the final shutdown phase.

July 1995 lb
sscinfo@hep.net

SSC2

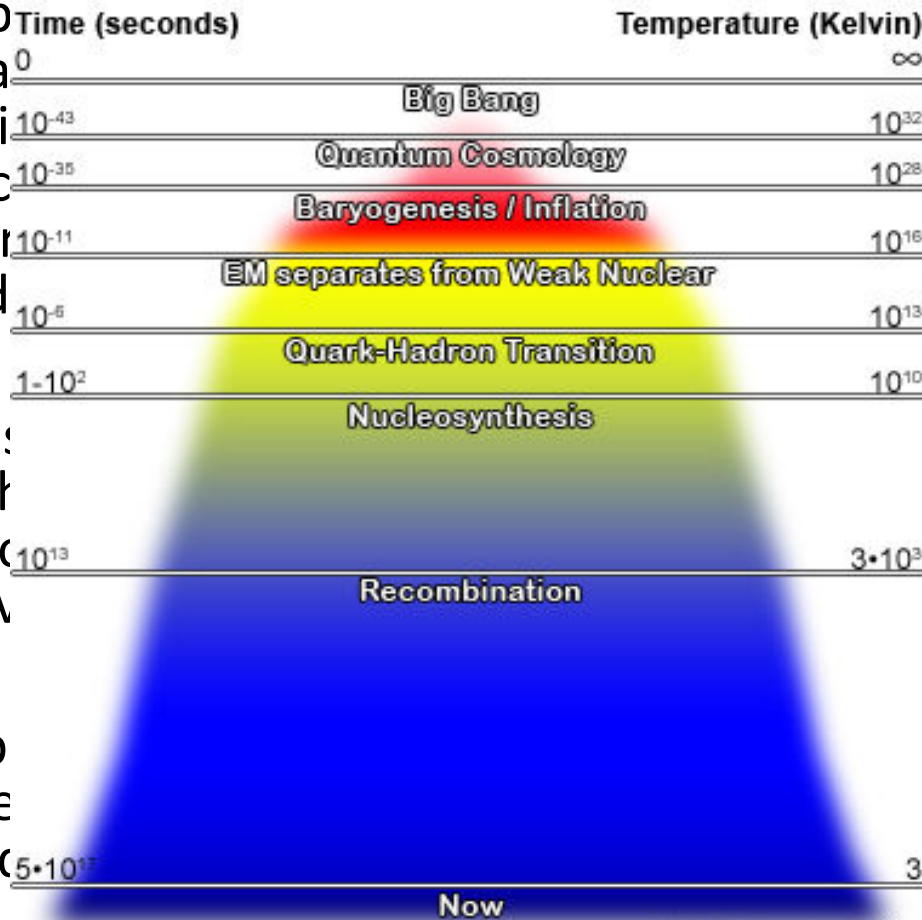
- SSC2 – aka “Isabella” is a new proton-*antiproton* collider secretly built by the US
- It has one single interaction point / experiment: “C0”
- It doesn’t use the grid – just a single “mainframe”



SSC2 – The Plot



Important Events in the Universe



- Isabella, born in Arizona and 15 miles from Stanton Leakey and 15 miles from the speed of light.

The goal is to probe the universe at a millionth of a second, the energy level of a proton.

The trouble is that the universe was once a plasma, and it's hard to probe such a hot, dense medium. Hence the need for a particle accelerator to generate power. So far, no results. Hence the need for a former monk, to go in and find out.

What? Where?



- 1.2.1.1 “EGI”- including “generic” services & operation required by WLCG!
 - e.g. GGUS, etc – “the [usual list](#)”
- 1.2.1.2 **Services** for large existing multi-national communities
- The funding for 1.2.1.1 + 1.2.1.2 = EUR25M; a joint proposal is expected
 - **Some people say / think that there is EUR5M for 1.2.1.2 (AFAIK not written down anywhere) and that the EUR5M should be shared with at least 1 other (than WLCG) large community**
- 1.2.3 “Virtual Research Communities” = EUR23M
 - Currently 2-3 “SSC” proposals foreseen (ideally(?) 1)
 - P2: combining Astronomy & Astrophysics, Earth Science, and Fusion;
 - P1: combining the training, dissemination, business outreach;
 - **P0: combining the other scientific SSCs (high-energy physics, life science, computational chemistry & material science, grid observatory, and complex systems).**
 - **Our stated plan for the “HEP SSC” is for a EUR10M project over 3 years, 50% of funding coming from EU (dependant on details such as exact scope, partners etc.)**
- Also other possible areas of funding, e.g.
 - **1.2.1.3 m/w (separate (important) topic, not this talk);**
 - Others: probably too much fragmentation: focus on the above 2 (3) areas
- **Obviously, what we target in the sum of all 3 areas should be consistent and meet our global needs**

E-Infrastructures in FP7: Call 7

(WP2010)

- This call opened on July 30th and is due to close 17:00 Brussels time (for most components) on November 24th
- It has numerous elements – not all of which are relevant to us – or at least not to this talk
- ✓ **If we are successful in our bid we should be in a good shape to:**
 - 1. Successfully and efficiently exploit the scientific and discovery potential of the LHC (+ help others...)**
 - 2. Prepare a plan for sustainability**

Distributed Computing

Infrastructure: INFRA-2010-1.2.1

- 1.2.1.1 – The European Grid Initiative
 - 1.2.1.2 – Service Deployment
 - 1.2.1.3 – Middleware and repositories
 - 1.2.1.4 – Access to DCI platforms
 - 1.2.1.5 – Extension of DCI platforms
- **The EU gives guidelines on Expected Impact, Examples of Activities, sets the funding scheme and budget**

Funding Scheme

- Combination of Collaborative Projects and Coordination and Support Actions (CP-CSA)

¿ What does this mean?

- For a proposal writer it means that there are specific criteria for the evaluation of the proposal that must be addressed!
- Guide for Applicants: FP7-Infrastructures-2010-2

Evaluation Criteria

Scientific / Technical Quality	Implementation	Impact
Soundness of concept / objectives	Management structure / procedures	Contribution to European research
Progress beyond state-of-the-art	Quality and relevant experience of participants	Dissemination etc.
Networking activities	Quality of consortium	Socio-economic impact
Trans-national Access and/or services	Allocation / justification of resources	
Quality of Joint Research Activities		
≥3/5	≥3/5	≥3/5
≥10/15		

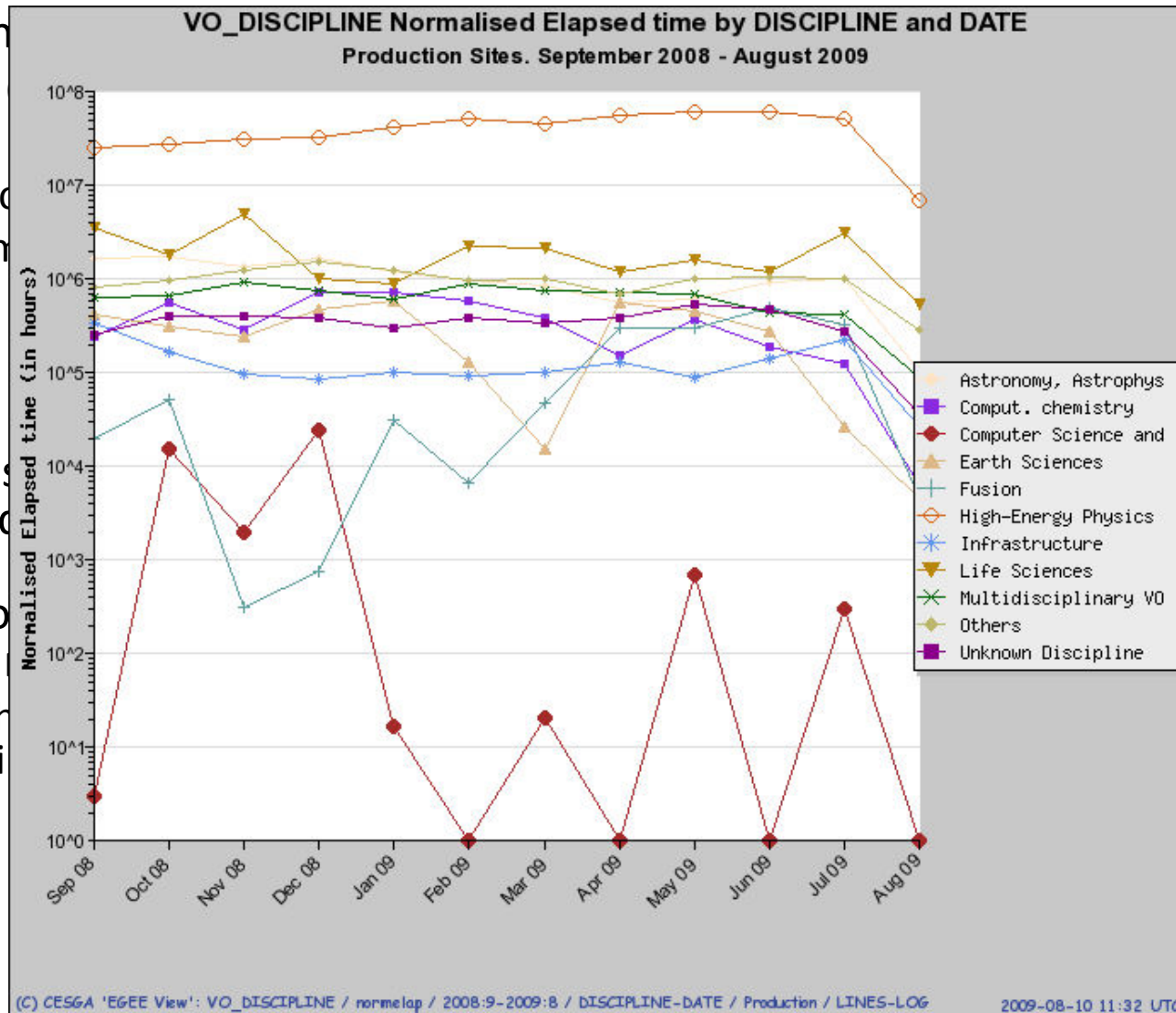
0	Fail
1	Poor
2	Fair
3	Good
4	Very good
5	Excellent – minor short-comings only

1.2.1: more details

- 1.2.1.1 (EGI)
 - Set up organization of sustainable grid services, including user support and m/w repository, operation and certification
 - Stakeholders: NGIs, others? (e.g. CERN)
 - Seamless transition
 - Promote international collaboration etc.
- 1.2.1.2 (Service Deployment)
 - Services for user communities that are heavy users of DCIs and have multinational dimension
 - (Sounds like an exotic search at SSC2)

1.2.1.2 – Service Deployment

- The Communities identified
 - High Energy Physics
 - Life Sciences (LS)
 - Astronomy and Astrophysics
 - Computational Chemistry
 - Earth Sciences (ES)
 - Fusion (F)
- Besides the massive use of the infrastructure and especially HEP and LS, the communities made decisive contributions to the deployment of the infrastructure and services
 - It is expected now that the services pertaining to the infrastructure will be used by the communities



1.2.1.2 – “SA4”

- There are currently 3 tasks in “SA4” that are of particular interest to us (subject to change)
 - TSA4.2: service deployment / operation of LFC, FTS
 - Request: 2 FTE@CERN, 1 FTE@each of 7 European T1s
 - TSA4.3: VO-specific services
 - Request: 10.5 FTE (ALICE:2, ATLAS:3, CMS: 3.5, LHCb: 2)
 - *Why? – explained next...*
 - TSA4.4: Dashboards, Ganga
 - Request: DB (4); GG (2)
- N.B. all of these numbers are total FTE – assume co-funding of 50% EU, 50% partner

👉 This is still too high wrt WLCG/HEP share of EUR5M

1.2.1.2 – “SA4”

- There are currently 3 tasks in “SA4” that are of particular interest to us (subject to change)
 - TSA4.2: service deployment / operation of LFC, FTS
 - FTS: several GB/s, every second of the day, many days in succession
 - FTS: up to 1PB/day over several days (beyond immediate LHC requirement)
 - **SUPPORTED AS GENERAL EGI INFRASTRUCTURE? TBC...**
 - TSA4.3: VO-specific services
 - *Why? – explained next... (really)*
 - TSA4.4: Dashboards, Ganga
 - Dashboards: strategic; more details in DB session and experiment talks
 - Ganga: adopted not only by ATLAS & LHCb but also many other HEP/non-HEP VOs
 - **SUPPORTED ALSO FOR OTHER HUCs**

Requirement for VO Services

- From SA4 draft:
 - *Building on the powerful generic infrastructure of the underlying grids that they use, the LHC experiments have developed important complementary services particularly in the areas of data and workload management, as well as in support for analysis services. Such services, which extend the capabilities of the infrastructure by exploiting knowledge of the experiment's computing model, data placement policies and/or information in metadata repositories, allow these massive international communities to maximise the benefit of the grids that they use. For example, PhEDEx, the CMS data movement system, is able to source files belonging to a larger dataset (a concept that does not exist at the underlying FTS layer) from alternative sites, leading to additional robustness and performance. As much as 50% of the data – possibly more – may be retrieved from such a source: functionality that cannot – by design – be provided at the FTS layer.*

VO Services & Frameworks - Manpower

VO	FTE requirement	Tasks
ALICE	2	AliEn services, covering workload management, data management (built upon standard components), integration of these services into WLCG services, VO box services and support.
ATLAS	3.5	Distributed Data Management system, built upon underlying services such as FTS and LFC and monitored via the Dashboard framework.
CMS	3.5	PhEDEx Data Service, CMS Remote Analysis Builder (CRAB) and related workload management and data services.
LHCb	2	DIRAC workload management and data management services.

Framework	Manpower Requirement
Dashboards	4 FTEs
Ganga	2 FTEs

1.2.3 – Virtual Research Communities (“SSCs”)

- In addition to 1.2.1.2, there is a call targetting “virtual research communities”
 - General objective:
 - Enable an increasing number of users and research communities from all science and engineering disciplines to access and use e-Infrastructures
 - Remove the constraints of distance, access and usability as well as the barriers between disciplines for a more effective scientific collaboration and innovation
 - More specifically:
 - Deployment of e-Infrastructures in research communities to enable multi-disciplinary collaboration
 - Deployment of end-to-end e-Infrastructure services and tools for integrating and increasing research capacities
 - This call – target budget EUR23M – is not limited to “EGI”
 - Some 10 “vertical” SSCs plus a couple of horizontal ones are foreseen
- **Given the size of each VO in our community and the scale of grid usage, something in the range “3 – 4 standard SSCs” seems to be reasonable**

ROSCOE - RObust Scientific Communities for EGI

- This includes the following disciplines:
 - [High Energy Physics](#)
 - [Life Science](#)
 - [Grid Observatory](#)
 - [Comp. Chem. and Material Science Tech.](#)
 - [Complex Systems](#)
 - Cal Loomis (CNRS) is overall coordinator
- Others include:
 - EGI SSC P1: Astron/Astrophysics, Earth Science, Fusion (Claudio Vuerli, INAF, IT)
 - CUE: Dissemination, Training, Outreach to business (Roberto Barbera, INFN, IT)

Requirements

- These proposals must include:
 - A Service Activity (SA)
 - A Network Activity (NA)
 - A Joint Research Activity (JRA)
- “Our” proposal (HEP input into “ROSCOE”) currently includes:
 - 3 WPs for WLCG/HEP (1 SA, 1 NA & 1 JRA)
 - 1 WP for **F**acility for **A**ntiproton and **I**on **R**esearch
 - Other WPs were foreseen but not included by time of deadline (31st August)

Service Activities

- To provide specific research infrastructure related services to the scientific community. This may include (non exhaustive list):
 - procurement and upgrading communication infrastructure, network operation and end-to-end services;
 - Grid infrastructure support, operation and management; integration, test and certification; services deployed on top of generic communication and computing infrastructures to build and serve virtual communities in the various scientific domains;
 - deployment, quality assurance and support of middleware component repositories;
 - data and resources management (including secure shared access, global scheduling, user and application support services) to foster the effective use of distributed supercomputing facilities; federated and interoperable services to facilitate the deployment and wide use of digital repositories of scientific information.
 - vertical integration of the different services in support of specific virtual research communities, including virtual laboratories for simulation and specific workspaces.

Network Activities

- To foster a culture of co-operation between the participants in the project and the scientific communities benefiting from the research infrastructures and to help developing a more efficient and attractive European Research Area. Networking activities could include (non exhaustive list):
 - joint management of access provision and pooling of distributed resources;
 - strengthening of virtual research communities;
 - definition of common standards, protocols and interoperability; benchmarking;
 - development and maintenance of common databases for the purpose of networking and management of the users and infrastructures;
 - spreading of good practices, consultancy and training courses to new users;
 - foresight studies for new instrumentation, methods, concepts and/or technologies;
 - promotion of clustering and coordinated actions amongst related projects;
 - coordination with national or international related initiatives and support to the deployment of global and sustainable approaches in the field;
 - dissemination of knowledge; internal and external communication;
 - promotion of long term sustainability, including the involvement of funders and the preparation of a business plan beyond the end of the project.

Joint Research Activities

- These activities should be innovative and explore new fundamental technologies or techniques underpinning the efficient and joint use of the participating research infrastructures. To improve, in quality and/or quantity, the services provided by the infrastructures, these joint research activities could address (non exhaustive list):
 - higher performance methodologies and protocols, higher performance instrumentation, including the testing of components, subsystems, materials, techniques and dedicated software;
 - integration of installations and infrastructures into virtual facilities;
 - innovative solutions for data collection, management, curation and annotation;
 - innovative solutions for communication network (increasing performance, improving management, exploiting new transmissions and digital technologies, deploying higher degrees of security and trust) and introduction of new end-to-end services (including dynamic allocation of resources and innovative accounting management);
 - novel grid architecture frameworks and policies, innovative grid technologies, or new middleware solutions driving the emergence of high level interoperable services;
 - advanced Service Level Agreements and innovative licensing schemes, fostering the adoption of e-Infrastructures by industry;
 - innovative software solutions for making new user communities benefit from computing services.

HEP SSC – JRA (Data Management)

- Investigation of innovative solutions for data management, targeting not only high-throughput multi-stream random-access style usage (typical of end-user analysis) but also the integration of new industry standards and solutions into end-to-end data management solutions covering catalog, file transfer and storage aspects;
- This is not a “rewrite everything” activity – it is a small but essential (3 FTE x 3 year) R&D work package
- It may well lead to a larger activity – or be superseded by one

Partner Acronym	Effort in Person-Months
CERN	36
DESY	36
INFN	36

HEP SSC – NA (Coordination)

- Dissemination of the progress and achievements of the e-infrastructure within the scientific and technical community and to the wider public
- Liaison with middleware providers; testing and collaborative deployment ...
- Liaison with EGI operations and user support and their counterparts in other grids
- Organization of regular workshops and conferences (WLCG Collaboration & Topical workshops etc.)
- **Overall WLCG Service Coordination and liaison with other WLCG structures and bodies;**
- Tier2 coordination, Network coordination;
- Address the long-term sustainability of this activity

Partner Acronym	Effort in Person-Months
CERN	108
Oslo	36
INFN	36
OSG (non-funded)	72

HEP SSC – SA (Service)

- This is essentially the *raison d'être* of the proposed support centre. In particular, one of its main goals is to support the High Energy Physics and related communities at this critical phase of LHC startup and exploitation.
- This involves approximately 10,000 researchers worldwide who need to access and analyze data 24x7 using worldwide federated grid resources.
- The service and user support to this community – enabling them to maximize the scientific and discovery potential of the LHC machine and the detectors that will take data at it – is a fundamental goal.

Partner Acronym	Effort in Person-Months
CERN	288 (= 8FTEs for 3 years)
INFN	216
GridPP	72
Oslo	36
FZU	36

SA – Objectives & Tasks

- User and application support services, including support for grid integration, production data processing and end-user analysis;
- Grid infrastructure / service deployment and support, including monitoring of resource usage and service availability / reliability, service coordination, debugging of complex middleware service problems and feedback to service / middleware providers;
 - 1. Integration Support**
 - 2. Operation Support**
 - 3. Distributed Analysis Support**

SA Task 1 – Integration Support

- HEP experiments have developed elaborate computing frameworks on top of the grid middleware(s) which now operate in full production. However, the experiments will need to adapt their infrastructures to exploit new middleware functionalities, cope with issues that will inevitably arise during data taking and improve the current operational model to increase automation and reduce the need for manual intervention.
 - Testing of new middleware features and functionality in pre-production environments, as well as stress testing of key components following experiment requirements. This includes negotiation of service setups with various NGIs and middleware providers, definition of the test environment, scenarios and metrics, development of the test framework, test execution and follow up.
 - Integration of experiment specific information in high level monitoring frameworks. The 4 main LHC experiments – ALICE, ATLAS, CMS and LHCb – developed specific monitoring frameworks for both workload and data management; the aim is to provide a general view of the experiments activities oriented to different information consumers: sites, other experiments, WLCG coordination.
 - Development of experiment specific plug-ins to existing frameworks. WLCG relies on complex frameworks such as Service Availability Monitoring (SAM), Service Level Status (SLS) and NAGIOS to measure site and service availability and reliability and to implement automatic notification and alarms. The experiments can benefit from a common infrastructure, developing specific plug-ins.
 - Further developments oriented to integration of middleware with the application layer. This includes maintenance of end-user distributed analysis tools and frameworks and their related VO-specific plug-ins.
 - Provision of a scalable and sustainable distributed support framework to support large user communities on all grid infrastructures used by a given VO.

SA Task 2 – Operation Support

- Many day by day operational tasks need grid expertise and such requests for specialized operational support will increase with the first LHC data when experiment computing models will need to react promptly to various use cases and scenarios.
 - Offer general grid expertise for identification and solution of middleware issues as well as site configuration and setup problems. This includes a possible risk analysis and definition of action plans to prevent escalation of criticality.
 - Development of experiment specific operational tools. Such tools include intelligent mining of grid monitoring data (for both workload and data management), automation of workflows and procedures, enforcement of data consistency across various services (storage and catalogs).
 - Support for the integration of experiment specific critical services into the WLCG infrastructure. This includes service deployment, definition of escalation procedures and support models.
 - Development and operation of tools which facilitate end-to-end testing of analysis workflows, including functional testing which is integrated with SAM and stress testing to investigate site- and VO-specific bottlenecks.

SA Task 3 – Distributed Analysis Support

- Scientists have been running analysis on the WLCG distributed infrastructure since many years. In addition, experiments successfully went through several challenges to test the readiness of the infrastructure and tools to massive end user analysis. However, with the arrival of the first LHC data, chaotic access is expected to scale up by an order of magnitude and attract inexperienced Grid users. Therefore, it should not only be foreseen to have a dedicated effort for maintenance and further development of analysis tools (already accounted for in Task 1 and Task 2 above), but also a focused end-user support structure, consisting of the following activities:
 - Investigation and deployment of tools which enable effective user-to-user and user-to-expert interaction.
 - Coordination of support providers, namely experts from the VO user communities.
 - Coordination of general and VO-specific training for end-users and support providers.
 - Partners: CERN(4?), GridPP(2/3), INFN(1), NDGF(1), FZU(1), (others?)

What? Where?



- 1.2.1.1 “EGI”- including “generic” services & operation required by WLCG!
 - e.g. GGUS, etc – “the [usual list](#)”
- 1.2.1.2 **Services** for large existing multi-national communities
- The funding for 1.2.1.1 + 1.2.1.2 = EUR25M; a joint proposal is expected
 - **Some people say / think that there is EUR5M for 1.2.1.2 (AFAIK not written down anywhere) and that the EUR5M should be shared with at least 1 other (than WLCG) large community**
- 1.2.3 “Virtual Research Communities” = EUR23M
 - Currently 2-3 “SSC” proposals foreseen (ideally(?) 1)
 - P2: combining Astronomy & Astrophysics, Earth Science, and Fusion;
 - P1: combining the training, dissemination, business outreach;
 - **P0: combining the other scientific SSCs (high-energy physics, life science, computational chemistry & material science, grid observatory, and complex systems).**
 - **Our stated plan for the “HEP SSC” is for a EUR10M project over 3 years, 50% of funding coming from EU (dependant on details such as exact scope, partners etc.)**
- Also other possible areas of funding, e.g.
 - **1.2.1.3 m/w (separate (important) topic, not this talk);**
 - Others: probably too much fragmentation: focus on the above 2 (3) areas
- **Obviously, what we target in the sum of all 3 areas should be consistent and meet our global needs**

Sustainability

- Given the proposed structure of the HEP SSC, this mainly concerns the Service Activity
 - Some level of JRA and NA are required but these are a small fraction of the total
- Different approaches may be required for different activities
 - E.g. the proposed model for Distributed Analysis Support – distributed across a small number of expert sites – appears reasonable and is likely to be sustainable
 - “Integration support” must decrease – but not to zero – with time
 - “Operation support” will continue – a responsibility both of main labs (tiers) plus experiments
 - VO services: core infrastructure – host lab(s); services themselves – experiments (rotational basis for 1-2 years)
- WLCG has always been a production deployment – aka service activity
- It has **never** been a research project!

Summary

- In order to address our two primary goals (below) proposals are being prepared in two areas: Services (1.2.1.2) + HEP SSC (1.2.3)
 1. **Successfully and efficiently exploit the scientific and discovery potential of the LHC (+ help others...)**
 2. **Prepare a plan for sustainability**
- ✓ **If approved, there will be a number of short-term positions (fellows, staff, students, others) ~mid-2010**
- There are detailed sessions on EGI in general, SSCs specifically and many other issues at EGEE'09 in Barcelona, e.g. complementary proposals
 - **Success in this area is of great importance to WLCG and other experiments in HEP & beyond**

More information: HEP SSC wiki

- <https://twiki.cern.ch/twiki/bin/view/LCG/HEPSSCPreparationWiki>
- From this page you can find pointers to the Indico category, the mailing list and all documents and presentations + “work area”
 - Most files simply “uploaded” – check comments!



EGEE'09 - Uniting our strengths to realise a sustainable European grid

21-25 September 2009 Hotel Barcelo Sants

- Overview
- Scientific Programme
- Timetable
- Contribution List
- Author index
- Book of abstracts [PDF]

Home > **Timetable** > Session details

A Specialised Support Centre for Large International Scientific Communities: Grid Services and Support for the WLCG, HEP and Related Communities

This session exposes in detail the status of discussions and planning for a proposed Specialised Support Centre (SSC) that would serve large existing scientific communities. The tentative list includes

- HEP
- Astro-particle physics
- Nuclear physics
- International initiatives - e.g. EnviroGRIDS, UNOSAT, ...

These communities are 'linked' by either the science that they perform and / or the technologies that they use - e.g. Ganga, Dashboards, AMGA, ...

More information can be found at the [Wiki](#) for this activity

Place: *Hotel Barcelo Sants*
Barcelona

Room: Barcelona B

Dates: Wednesday 23 September 2009 14:30

Conveners: [Dr. Shiers, Jamie](#)

[Contribution List](#) **Time Table**

Wednesday, 23 September 2009	
14:00	[177] Overview of the main areas of "EGI" relevant to Heavy Multi-National Users (Barcelona B: 14:30 - 15:15)
15:00	[178] Services for Heavy Users of DCIs (Barcelona B: 15:15 - 16:30)
16:00	[179] Specialised Support Centre: communities supported & workplan (Barcelona B: 17:00 - 18:30)
17:00	[180] Next steps in Proposal Preparation & Wrap-up (Barcelona B: 18:30 - 19:00)