

DEISA - Cooperative Extreme Computing Across Europe

Ralph Niederberger



Overview



- **DEISA objectives**
- **HPC environment, facilities and network overview**
- **Management and operation structures and tools**
- **Used middleware**
- **User environment and security**
- **Lessons learned and conclusion**

DEISA objectives

- contribute to a significant enhancement of capabilities and capacities of high performance computing (HPC) in Europe
 - ⇒ integration of leading national supercomputing infrastructures
- *deploy and operate a distributed multi-terascale European computing platform*, based on a strong coupling of existing national supercomputers not tied to any specific pre-established technology
 - ⇒ operate as a virtual European supercomputing center
- *contribute to the deployment of an extended, heterogeneous Grid computing environment for HPC in Europe*
 - ⇒ interfacing the DEISA research infrastructure with the rest of the European IT infrastructures.
- Enabling new science is the only criterion for success.

Participating Sites

BSC	<i>Barcelona Supercomputing Centre</i>	Spain
CINECA	<i>Consorzio Interuniversitario per il Calcolo Automatico</i>	Italy
CSC	<i>Finnish Information Technology Centre for Science</i>	Finland
EPCC/HPCx	<i>University of Edinburgh and CCLRC</i>	UK
ECMWF	<i>European Centre for Medium-Range Weather Forecast</i>	UK (int)
FZJ	<i>Research Centre Juelich</i>	Germany
HLRS	<i>High Performance Computing Centre Stuttgart</i>	Germany
IDRIS	<i>Institut du Développement et des Ressources en Informatique Scientifique - CNRS</i>	France
LRZ	<i>Leibniz Rechenzentrum Munich</i>	Germany
RZG	<i>Rechenzentrum Garching of the Max Planck Society</i>	Germany
SARA	<i>Dutch National High Performance Computing and Networking centre</i>	The Netherlands

DEISA supercomputing environment



23.460 processors and 205 Teraflop in March 2007, but changing constantly

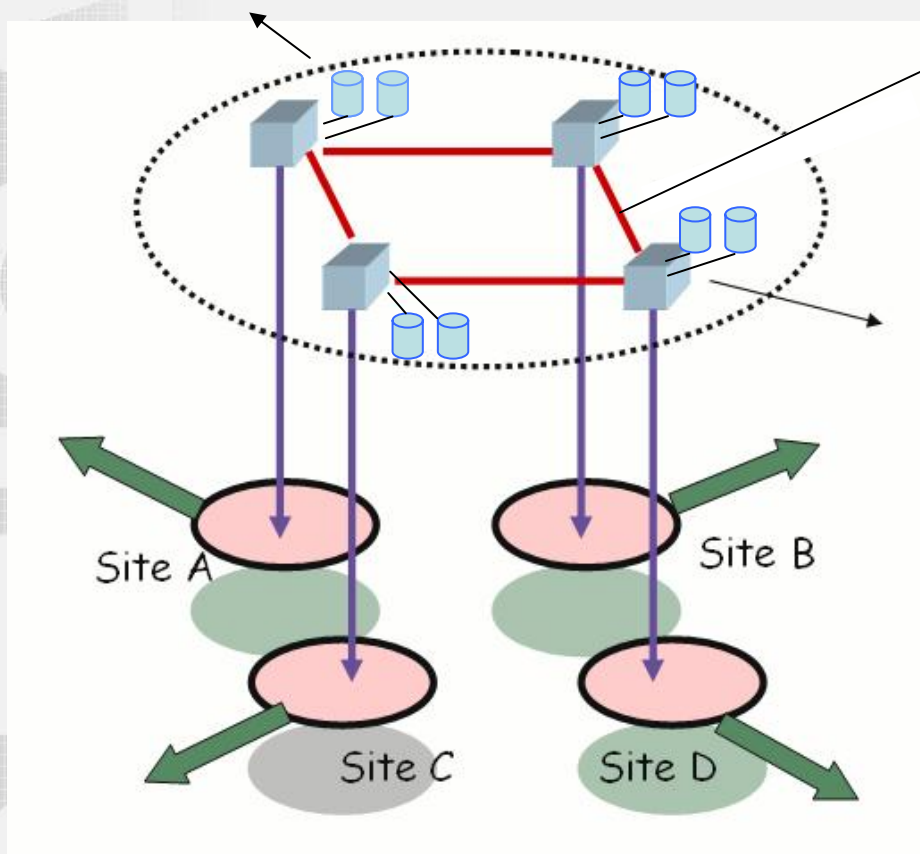
IBM AIX Super-cluster

- FZJ - Juelich, 1312 processors, *8,9 teraflops peak*
- RZG - Garching, 896 processors, *4,6 teraflops peak*
- IDRIS, 1024 processors, *6,7 teraflops peak*
- CINECA, 512 processors, *2,6 teraflops peak*
- CSC, 512 processors, *2,2 teraflops peak*
- ECMWF, 2 systems 2276 processors each, *33 teraflops peak*
- HPCx, 1600 processors, *12 teraflops peak*
- **BSC, IBM PowerPC Linux system (MareNostrum)**
10240 processors, *94 teraflops peak*
- **SARA, SGI ALTIX Linux system**
416 processors, *2,2 teraflops peak*
- **LRZ, SGI ALTIX system** 4096 processors, *26,2 teraflops peak*
in 2007 *> 60 teraflops peak*
- **HLRS, NEC SX8 vector system** 576 processors, *12,7 teraflops peak*

The DEISA facility

Global distributed, high performance file system with continental scope (GPFS).

**Dedicated bandwidth network
GEANT, RENATER, DFN, GARR, ...**



National supercomputing platforms:

IDRIS - France

JÜLICH - Germany

GARCHING - Germany

CINECA - Italy

...

SARA – The Netherlands

CSC - Finland

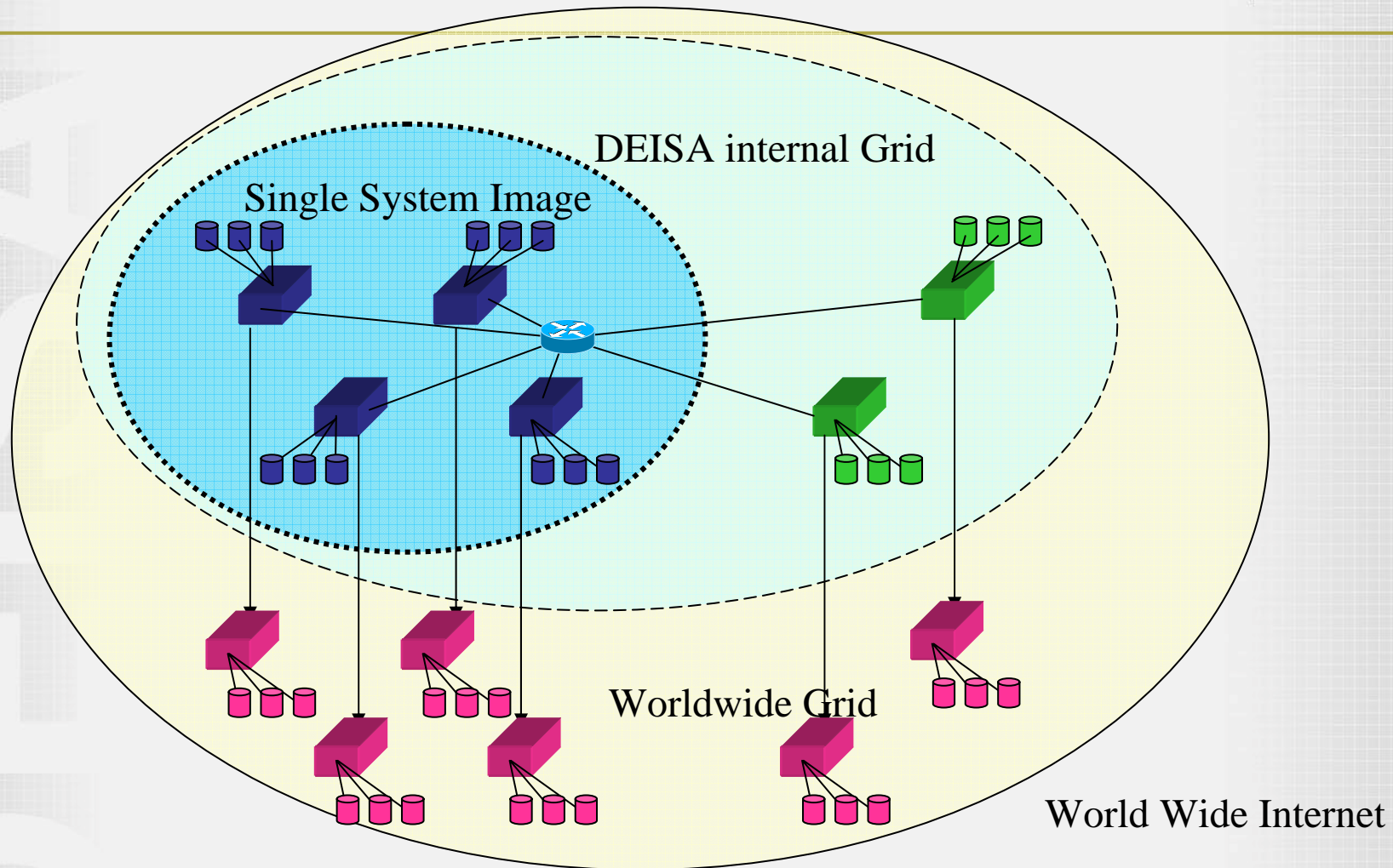
Extended Grid services :

Portals, Web-like services, ...

Interfacing the core platform to other virtual organizations.

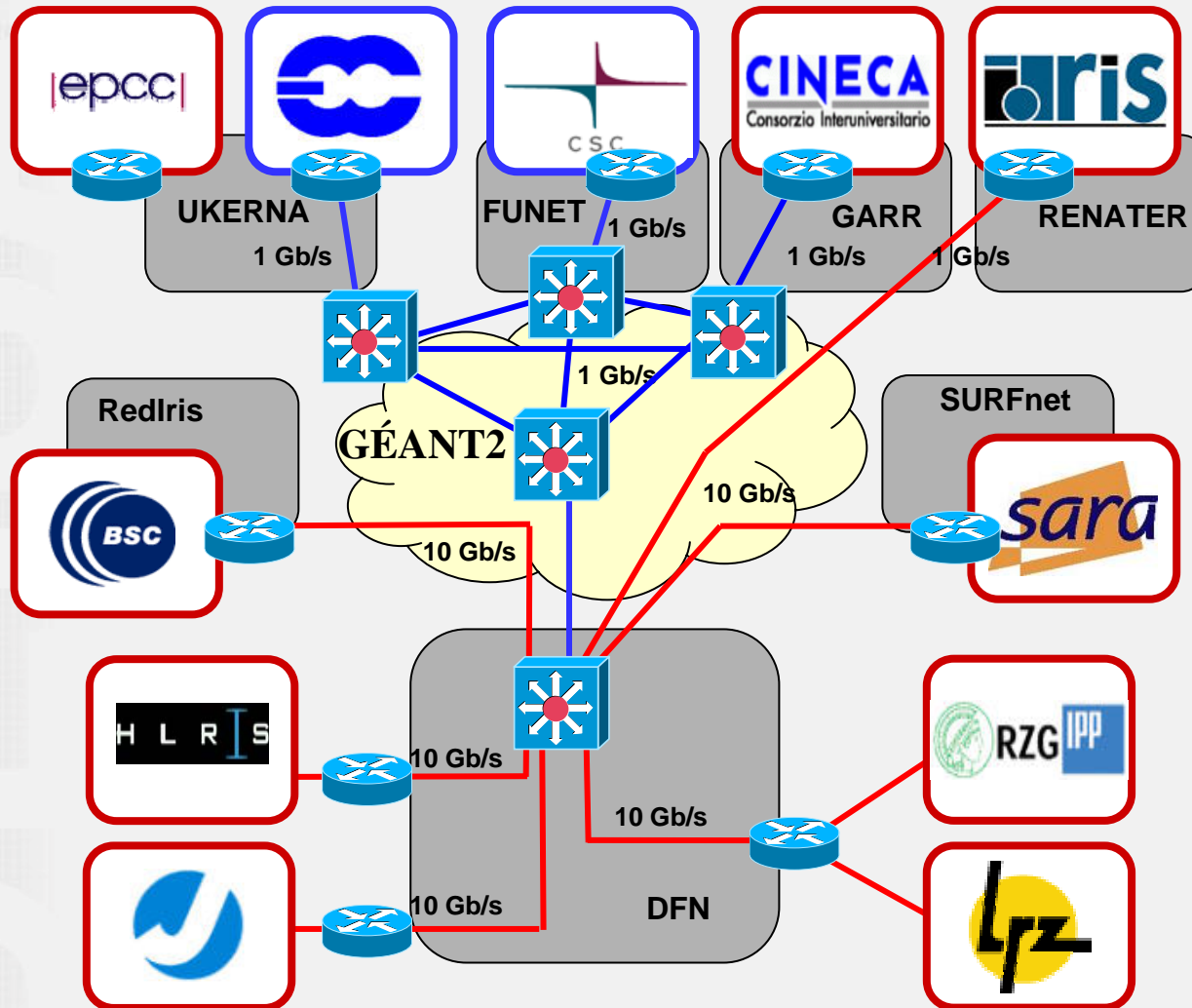
Grid-middleware Unicore hiding complex environments from end users

DEISA and the Grid



DEISA network infrastructure

intermediate phase



Dedicated
10 Gb/s
wavelength

1 Gb/s LSP

DEISA management structure

Definition of a number of executive and working teams provides operation and management within DEISA

DEISA Executive Committee (DEC) –

managing staff members of all centers, dealing with higher level project management issues (technological strategies, scientific management, provision of computational resources, external relations)

DEISA Policy Committee (DPC) –

representatives of organizations defining and implementing national policies in HPC and providing funding of DEISA resources

Advisory Technology Committee (ATC) –

HPC and Grid technology experts from partners. Acts as an external consulting board on technology issues

Advisory Scientific Committee (ASC) –

Number of experts in computational sciences. Acts as an external consulting board on scientific issues

Advisory User Committee (AUC) –

constituted of scientists managing DEISA supported projects. Feedback on analysis of operation and quality of services

DEISA Infrastructure Management Committee (DIMC) –

Executive extension of DEC (project director, leaders SA & Operations team)

DEISA

service and joint research activities

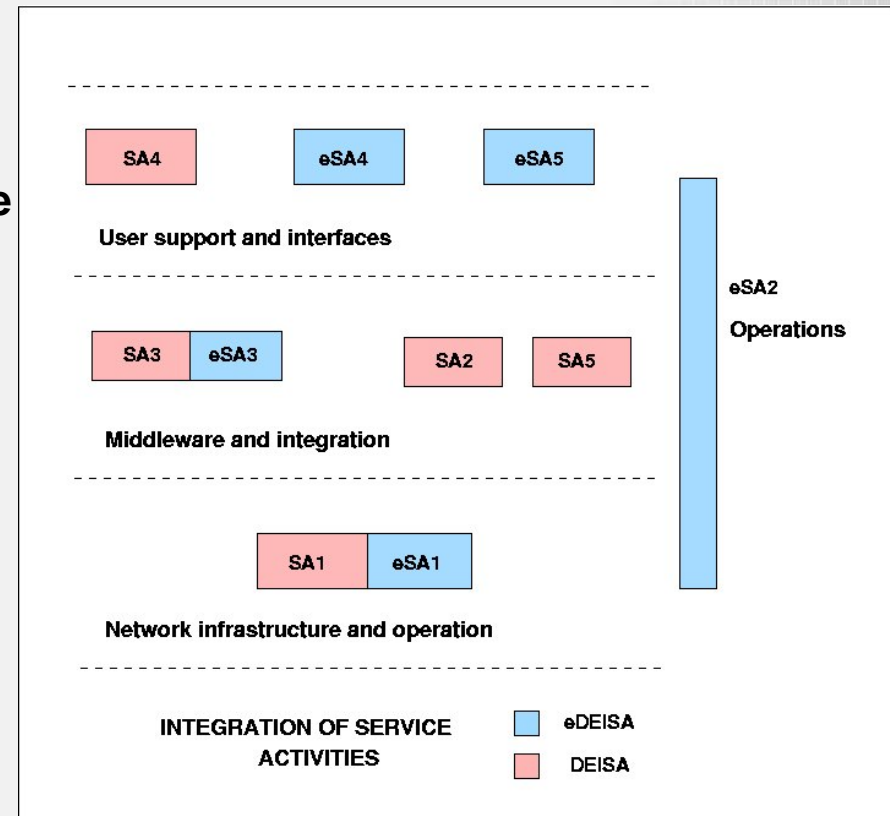


Service Activities

- SA1: Network Operation and Support
- eSA1: Extended Network Infrastructure
- SA2: Data Management with GPFS
- eSA2: Operation of the Grid Infrastructure
- SA3: Resource Management
- eSA3: Extended Resource Management
- SA4: Applications and User Support
- eSA4: Application Enabling
- SA5: Security
- eSA5: User Interfaces

Joint Research Activities

- JA1: Material Sciences
- JA2: Cosmology
- JA3: Plasma Physics
- JA4: Life Sciences
- JA5: Industrial CFD
- JA6: Coupled Applications
- JA7: Access to Resources in Heterogeneous Environments



DEISA

operation and management



- each SA has its own well defined communication and operation channels
- interactions between SAs as cases arise
- overall coordination done by operations team
- “operations team” often also “Emergency response team”
- used communication channels:
 - phone conferences (on demand)
 - video conferences (biweekly)
 - on site discussions (on demand)
 - training sessions (quarterly or on demand)
 - email lists for every SA and special issues

DEISA and its operation tools

- **TTS (Trouble Ticket System)**
developed by LRZ. Documents problems potentially crossing site boundaries. Regularly checked and updated.
- **RMIS (Resource Management Information System)**
delivers up to date and complete resource management information of supercomputing system to administrators and end users.
- **Ganglia monitoring tool**
provides monitoring of distributed systems and large-scale clusters. Ganglia information fed into MDS2/Globus component.
- **MDS2/Globus**
provides standard mechanism for publishing and discovering of resource status and configuration information via a uniform flexible interface to data. Information is displayed by the RMIS system.
- **INCA system implementation**
provides overview and manages DEISA CPE. Runs periodically a number of validation scripts (*reporters*), collecting installed versions and checking correct operation. INCA displays all software components used in CPE allowing to check where software can be run.

DEISA and its grid middleware

Challenging task is to run bigger and more demanding applications.

“Grid enable” applications is not the right way to do (old term “metacomputing” failed because of latency problems)

DEISA uses a different strategy:

- **Load balance computational load** across national borders. Huge, demanding apps can be run because of reorganizing workload (freeing resources) by transferring smaller jobs to other sites (**MC-LoadLeveler**)
- Transparent file sharing through IBM’s **GPFS**
- **UNICORE** used for accessing the heterogeneous set of computing resources and managing workflow applications
- “first generation” Co-allocation service based on **LSF Multi Cluster** from Platform Computing
- “second generation” Co-allocation service will be deployed that is vendor independent
- New middleware will be evaluated in future (e.g. **Unicore6 & GTK4**)
- **Gridftp** to access and store data on other non-DEISA storage resources
- Provide high performance access to distributed data sets (DB management software OGSA-DAI or grid storage software SRB)

DEISA and its users



- **Help desks** and **Dispatch** are available at any DEISA site for local users support
- Documentation for using the DEISA facilities
 - **DEISA primer**
 - for *DECI (DEISA Extreme Computing Initiative)* users
 - and DEISA standard users (DSU)
 - **Frequently Asked Questions** list on DEISA web
 - **DEISA public deliverables** on DEISA web
 - **DEISA Training sessions** and **DEISA Symposium**
 - Quarterly **DEISA newsletter**
 - **DEISA press releases**

DEISA and security

DEISA constituted by 11 European organizations connecting supercomputers, grid systems, servers, management stations and network equipment includes also **11 different security policies**

- In this sense DEISA makes up a “**Virtual Organization**”.
- Partners need:
 - **transparent access to Grid resources**
 - **control of local site’s resource usage**
 - **security**
- One UID/GID per DEISA user realized by distributed **LDAP service** updated locally once per day
- Single sign on via **X.509 certificates** (CA’s accredited by EuGridPMA)
- **Dedicated network** only reachable via supercomputer and server systems, which implies a “net of trust”
- Local DEISA **site firewalls**
- Local DEISA sites CERTs form the **global DEISA CERT**

DEISA lessons learned

- Operation of an DEISA like infrastructure opens new management challenges
- Staff members dealing with a problem are thousands of miles far away
- No short cuts, no office next door
- Every small software or hardware modification requires
 - agreement on all sites
 - may lead to dependencies not directly obvious
- Task scheduling, installations, maintenance, network infrastructure changes have to be planned in advance and agreed on
- An **operations team is mandatory** to handle all this issues and to decide on further progress in case of disagreement

Conclusion



- Three years of successful operation have shown that the concept implemented in DEISA proceeded very well
- This does not preclude that organizational structures of DEISA may change over time
- But general idea of DEISA will sustain
- Next step will be to establish an efficient organization embracing all relevant HPC organizations in Europe
- Being a central player within European HPC, DEISA intends to contribute to a global infrastructure for science and technology furthermore
- Integrating leading supercomputing platforms with Grid technologies and reinforcing capability with shared petascale systems is needed to open the way to new research dimensions

Questions ???