Distributed European Infrastructure for Supercomputing Applications

DEISA - Cooperative Extreme Computing Across Europe

Ralph Niederberger, r.niederberger@fz-juelich.de
Achim Streit, a.streit@fz-juelich.de
Andreas Schott, schott@rzg.mpg.de
Paolo Malfetti, p.malfetti@cineca.it

Overview

Distributed European Infrastructure for Supercomputing Applications

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

r.niederberger@fz-juelich.de

DEISA objectives

Distributed European Infrastructure for Supercomputing Applications

- **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.**

r.niederberger@fz-juelich.de

Participating Sites

Distributed European Infrastructure for Supercomputing Applications

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

r.niederberger@fz-juelich.de

DEISA supercomputing environment

Distributed European Infrastructure for Supercomputing Applications

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

IBM AIX Super-cluster

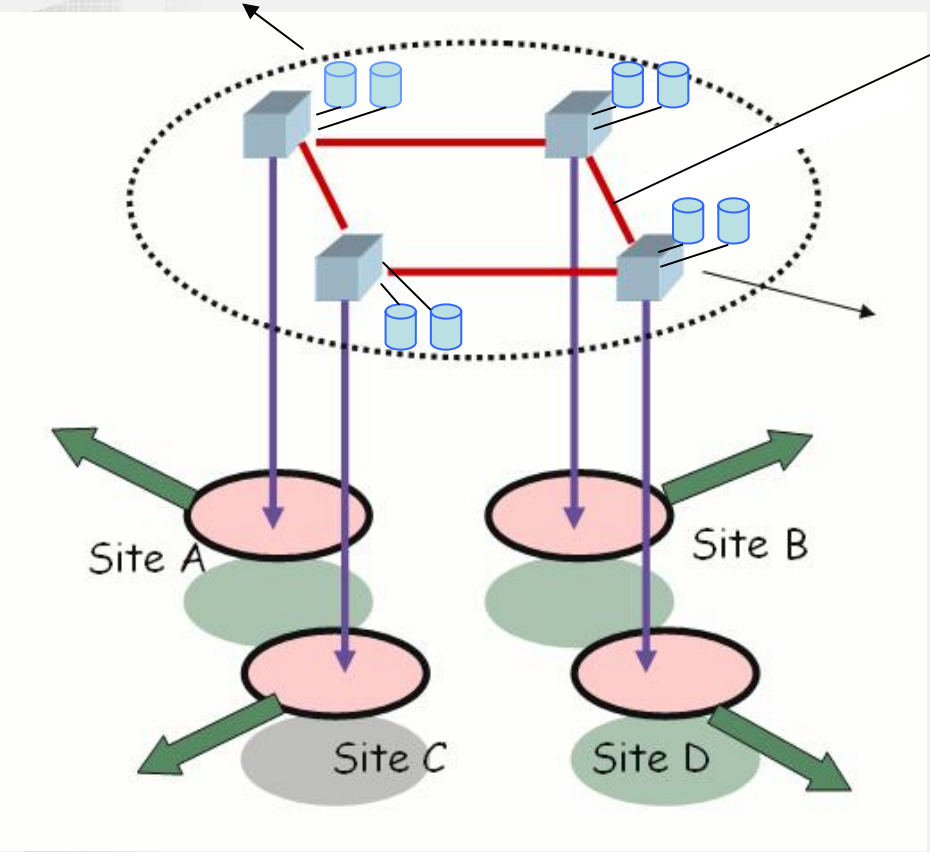
– FZJ - Juelich,	1312 processors,	*8,9 teraflops peak*
– RZG - Garching,	896 processors,	*4,6 teraflops peak*
– IDRIS,	1024 processors,	*6,7 teraflops peak*
– CINECA,	512 processors,	*2,6 teraflops peak*
– CSC,	512 processors,	*2,2 teraflops peak*
– ECMWF, 2 systems	2276 processors each,	*33 teraflops 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*

r.niederberger@fz-juelich.de

The DEISA facility

Distributed European Infrastructure for Supercomputing Applications

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



Dedicated bandwidth network:
GEANT, RENATER, DFN, GARR, …

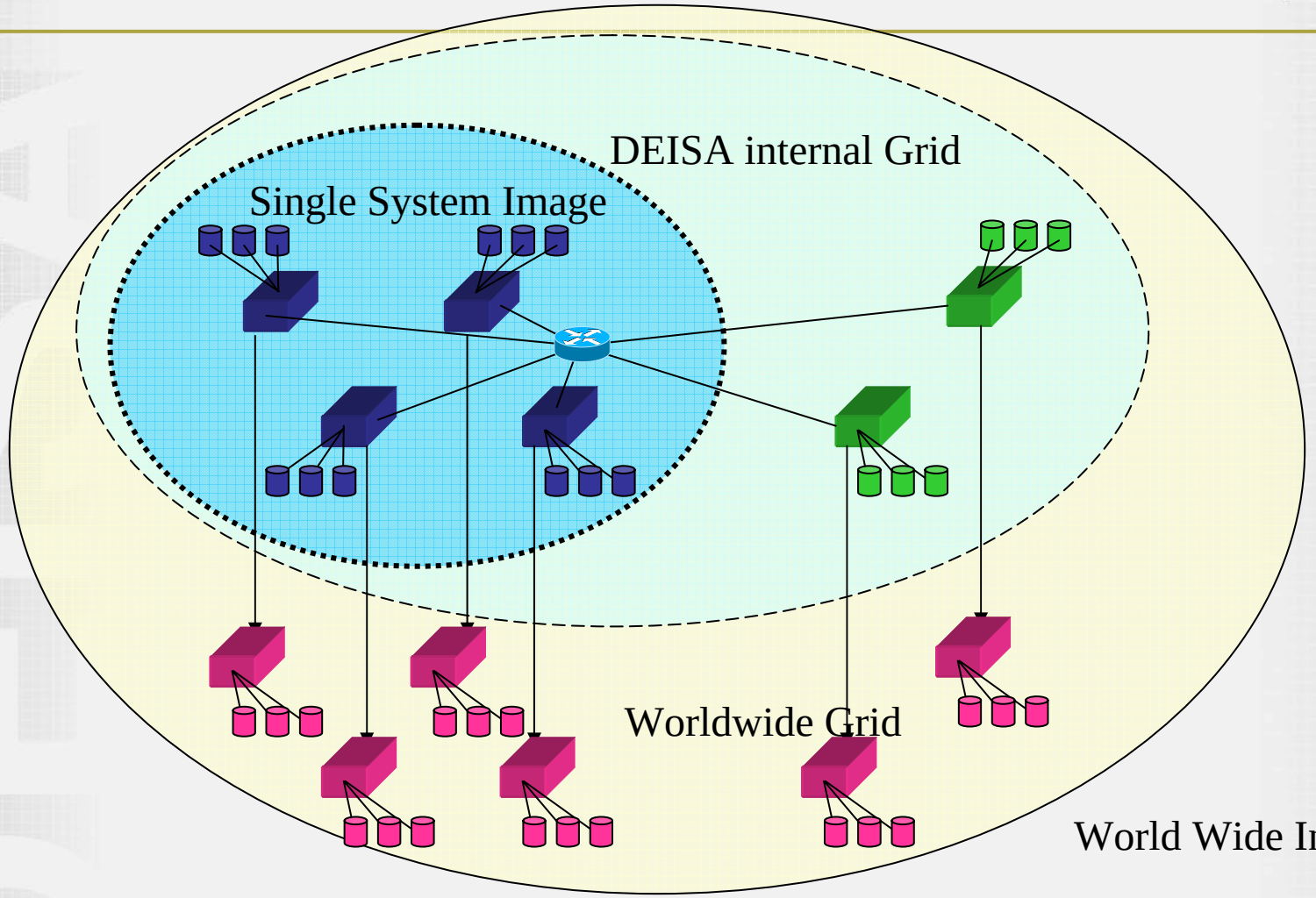
National supercomputing platforms:
IDRIS - France
JÜLICH - Germany
GARCHING - Germany
CINECA - Italy
…
SARA – The Netherlands
CSC - Finland

Extended Grid services :
Portals, Web-like services, …
Interfacing the core platform to other virtual organizations.
Grid-middleware Unicore hiding complex environments from end users

r.niederberger@fz-juelich.de

DEISA and the Grid

Distributed European Infrastructure for Supercomputing Applications

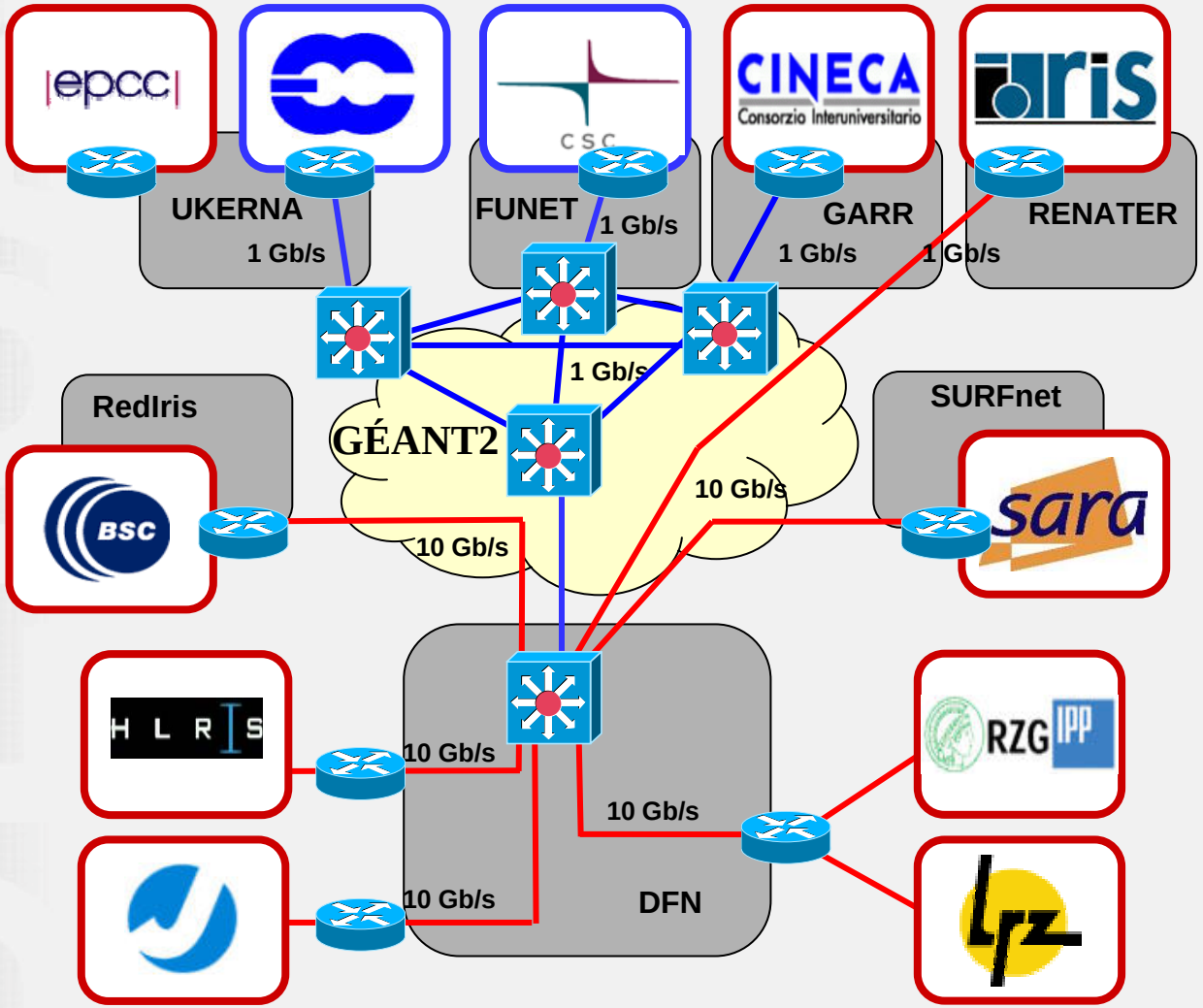


r.niederberger@fz-juelich.de

DEISA network infrastructure

intermediate phase

Distributed European Infrastructure for Supercomputing Applications



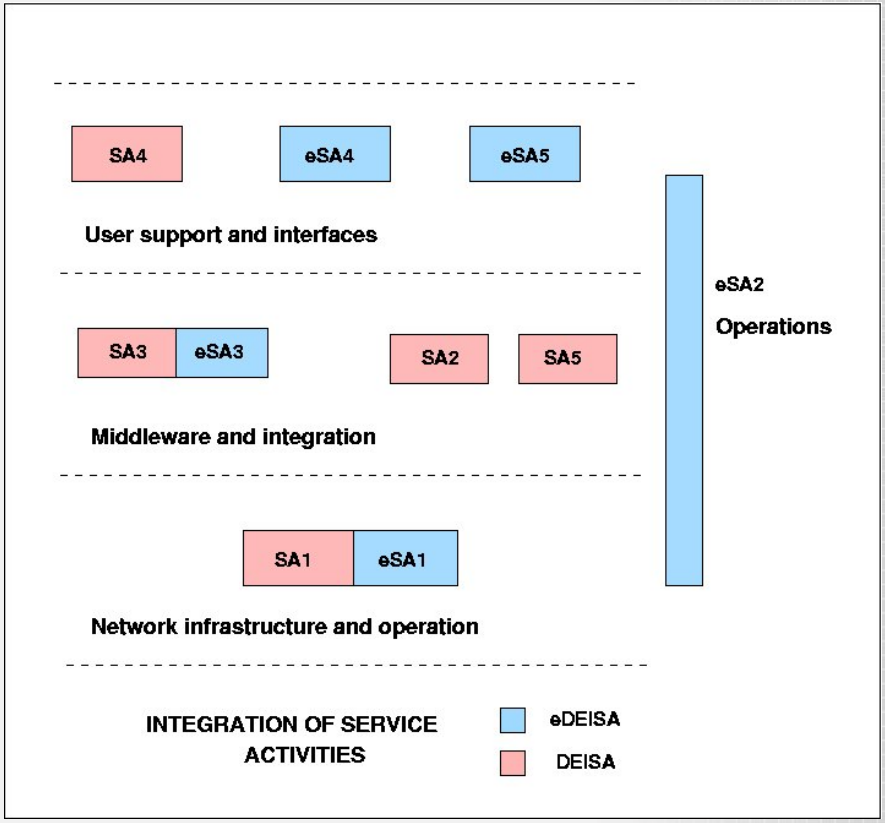
r.niederberger@fz-juelich.de

DEISA
service and joint research activities

Distributed European Infrastructure for Supercomputing Applications

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

r.niederberger@fz-juelich.de

DEISA
operation and management

Distributed European Infrastructure for Supercomputing Applications

- definition of a number of executive and working teams provides operation and management within DEISA (policy, executive, technology, science, user)
- 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

r.niederberger@fz-juelich.de

DEISA and its operation tools

Distributed European Infrastructure for Supercomputing Applications

- **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.

r.niederberger@fz-juelich.de

DEISA and its middleware

Distributed European Infrastructure for Supercomputing Applications

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 boarders. 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)

r.niederberger@fz-juelich.de

DEISA and its users

Distributed European Infrastructure for Supercomputing Applications

- **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**

r.niederberger@fz-juelich.de

DECI:
enabling leading computational science

Distributed European Infrastructure for Supercomputing Applications

- DECI is basic service providing model for scientific users
- Identification, deployment and operation of a number of « flagship » applications requiring the infrastructure services, in selected areas of science and technology.
- European Call for proposals in May-June every year. Applications selected on scientific excellence, innovation & relevance with collaboration of HPC national evaluation committees.
- 29 projects in operation in 2005 – 2006 (23/5 in 2206-2007)
- Supported by Applications Task Force. ATASKF activities:
  - enable and deploy the Extreme Computing applications
  - Hyperscaling of huge parallel applications, data oriented applications, Workflows and coupled applications
  - Production of an European Benchmark Suite for HPC systems (in collaboration with the HPC-EUR initiative, to be used in future procurements of European supercomputers).

r.niederberger@fz-juelich.de

DECI status, June 2006

Distributed European Infrastructure for Supercomputing Applications

Acronym	Scientific Discipline	Principal Investigator(s) and Affiliation	cpu-h	Tot mem	Sta-tus	Home site	Exec Site
SIMU-LU	Computational Cosmology	Gustavo Yepes, Universidad Autónoma de Madrid, Spain	1 600 000		P	BSC	BSC
LFI-sim	Cosmology, Space Missions	Fabio Pasian, INAF – O.A. Trieste, Italy	500 000		1/4	CIN	Core
HORIZON	Astrophysics, Cosmology	Julien Devriendt, Centre de Recherche Astronomique de Lyon, France	1 400 000		P	IDR	
STARS	Astrophysical Fluid Dynamics, Stellar MDT	Allan Sacha Brun, CEA – CNRS – Univ. Paris 7, UMR7158, France	220 000		1/4	IDR	RZG
FEARLESS	Astrophysics	Jens Niemeyer, University of Würzburg, Germany	300 000		1/2	LRZ	SARA
GIMIC	Astrophysics, Cosmology	VIRGO Consortium via Simon White, Max Planck Institute for Astrophysics, Germany, and Carlos Frenk, University of Durham, UK	600 000		P	EPCC	EPCC
SUPERNOVA	Astrophysics, Supernova Research	Wolfgang Hillebrandt, Max Planck Institute for Astrophysics, Germany	240 000		F	RZG	EPCC
REIONIZATION	Astronomy	Garrelt Mellema, Astron, The Netherlands	300 000		P	SARA	
TASEC	Bioinformatics	Modesto Orozco, INB, Spain	1 400 000		P	BSC	
BET	Computational Biophysics	Paolo Carloni, SISSA Trieste, Italy	200 000		F	CIN	IDRIS
SNARE	Computational Chemistry, Biology	Marc Baaden, UPR9080 CNRS, Labo. De Biochimie Theorique, France	256 000		P	IDR	RZG
LCDIS	Physical Chemistry of Materials	Claudio Zannoni, Universita' di Bologna, Italy	200 000		P	CIN	
NATURE	Nanoscience, Materials Science	Risto Nieminen, Helsinki University of Technology, Finland	1 000 000		P	CSC	
LIAMS	Large Scale Molecular Dynamics	Peter V. Coveney, University College London, UK	600 000		P	EPCC	Core
IQCS	Quantum Computing	Hans de Raedt, University of Groningen, The Netherlands	100 000		P		
CAMP	Materials Science	Michele Parrinello, ETH Zurich, Switzerland	300 000		1/2	RZG	FZJ
POLYRES	Materials Science	Kurt Kremer, Max Planck Institute for Polymer Research, Germany	350 000		1/2	RZG	CIN
QCfarm	Quantum Chemistry	J.H. van Lenthe, Utrecht University, The Netherlands	1 000 000		P	FZJ	
Channel-2000	Fluid Mechanics	Javier Jiménez, Universidad Politécnica de Madrid, Spain	819200		F	BSC	BSC
HEAVY	Turbulence, Fluid Dynamics	Federico Toschi Instituto per le Applicazioni del Calcolo – CNR, Italy	400 000		P	CIN	
FOCUS	Combustion	Olivier Gicquel, Ecole Centrale Paris – CNRS / France	300 000		P	IDR	
SJN	Computational Fluid Dynamics	Jorn Sesterhenn, Technical University of Munich, Germany	300 000		P	LRZ	
GLORIA	Earth Science – Air Quality and Climate Change	José M. Baldasano, Barcelona Supercomputing Center (BSC-CNS) - Earth Science Division, Spain	450 000		P	BSC	
SSSC	Earth System Modelling - Atmospheres	Heiski Järvinen, Finnish Meteorological Institute, Finland	300 000		P	CSC	
ESSENCE	Climate Research	H.A. Dijkstra, IMAU, Utrecht University, The Netherlands	1 200 000		P	SARA	HLRS
GROM	Physical Oceanography	Pierre Bahurel, GIP Mercator Océan, France	450 000		P	IDR/ ECMWF	ECMWF
GYROKINETICS	Plasma Physics	Karl Lackner, Max Planck Institute of Plasma Physics, Germany	500 000		P	RZG	LRZ
HDM	Particle Physics, Nuclear Physics	Kari Rummukainen, University of Oulu, Finland	350 000		1/4	CSC	
TMQCD	Lattice Field Theories	Karl Jansen, DESY-Zeuthen, Germany	350 000		1/2	FZJ	Core

r.niederberger@fz-juelich.de

DEISA and security

Distributed European Infrastructure for Supercomputing Applications

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**

r.niederberger@fz-juelich.de

DEISA lessons learned

Distributed European Infrastructure for Supercomputing Applications

- 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

r.niederberger@fz-juelich.de

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 eInfrastructure 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

r.niederberger@fz-juelich.de

Distributed European Infrastructure for Supercomputing Applications

Questions ???

r.niederberger@fz-juelich.de