

Production-Quality Grid Environments with UNICORE

Dietmar Erwin, Michael Rambadt,
Achim Streit and Philipp Wieder
Research Centre Jülich (FZJ)

GGF 14 Workshop on Grid Applications:
from Early Adopters to Mainstream Users

June 27, 2005, Chicago



Contents

- Setting the scene
- UNICORE in production
- OpenMolGRID: an application example
- Lessons learned: what users expect (& what they get)
- Conclusions



- Long project history (and future), currently DEISA, NaReGI, UniGrids, VIOLA, ...
- Features
 - GUI with single sign-on, X.509 certificates for AA and job/data signing, only one opened port in firewall required, workflow engine for complex multi-site/multi-step workflows, job monitoring, extensible application support, secure data transfer integrated, resource management, resources autonomy remains, production quality, ...
- Available as Open Source under BSD license on SourceForge at <http://unicore.sourceforge.net>

Production at FZJ

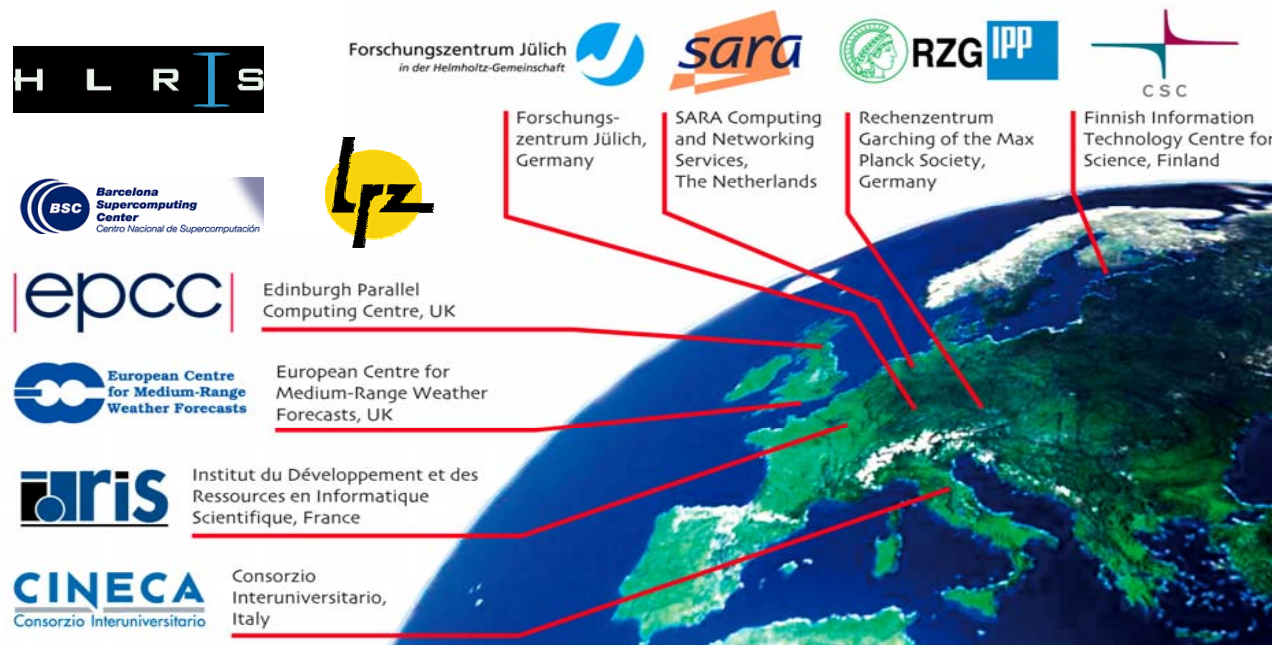


- National high-performance computing centre
“John von Neumann Institute for Computing”
- About 650 users in 150 research projects
- Access via UNICORE to
 - p690 IBM eServer Cluster 1600 (8,9 TFlops peak)
since June 2004
 - BlueGene/L (5,7 TFlops peak)
 - Cray SV1ex, Cray XD1
- 116 active UNICORE users
 - 72 external, 44 internal
- Resource usage (2004/05; CPU-hours)
 - Dec: 18.4%, Jan: 30.4%, Feb: 30.5%,
Mar: 27.1%, Apr: 29.7%, May: 39.1%

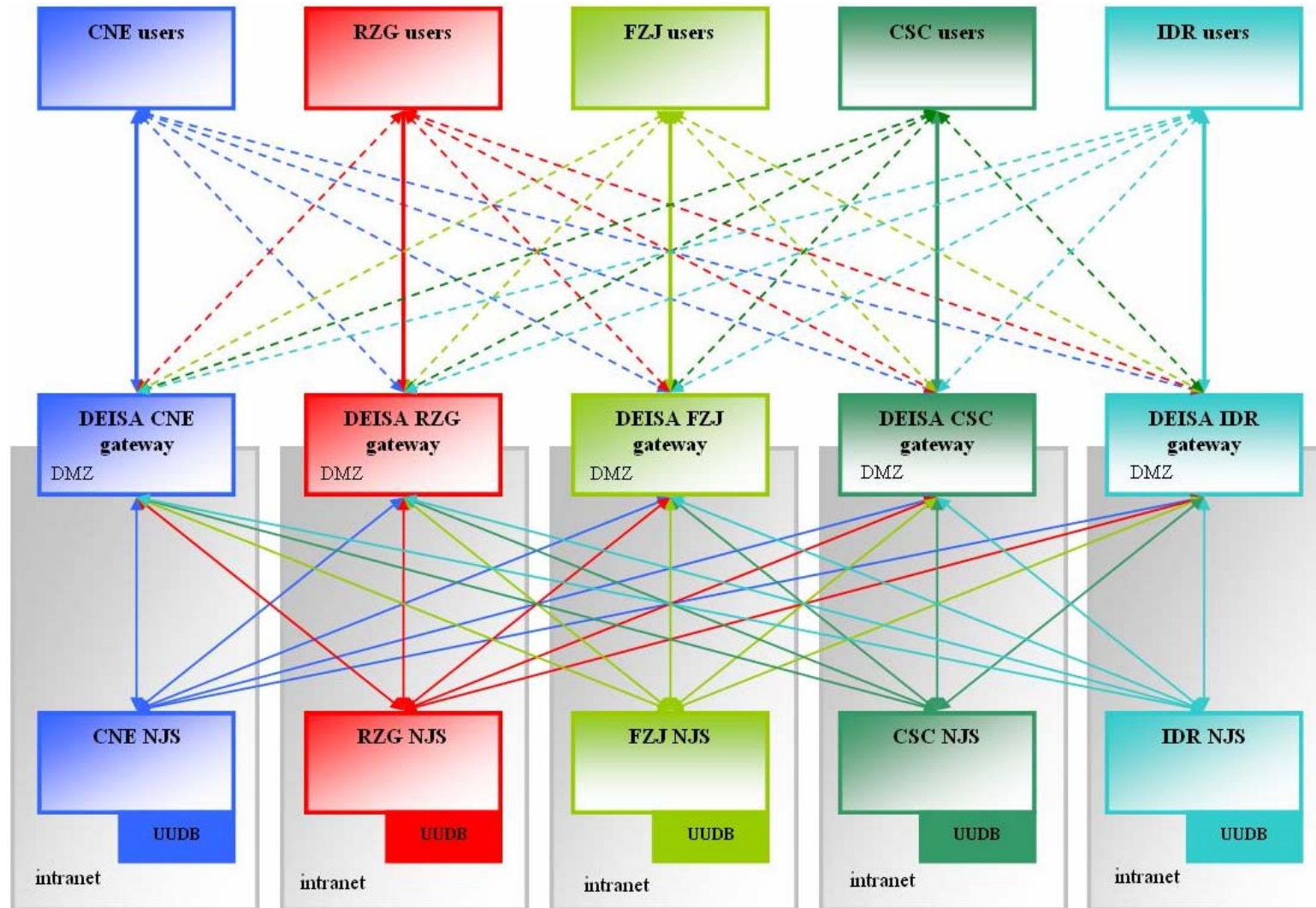


DEISA

- Status (Q1/2005): 4 sites / ~24 TFlops IBM Power4 / 1 Gbit / GPFS / LoadLeveler
- Goal (2006): 11 sites / ~140 TFlops / 10 GBit / GPFS / Metascheduler



DEISA configuration



Goal: "Speed-up, automate, and standardise the drug-design using Grid technology"

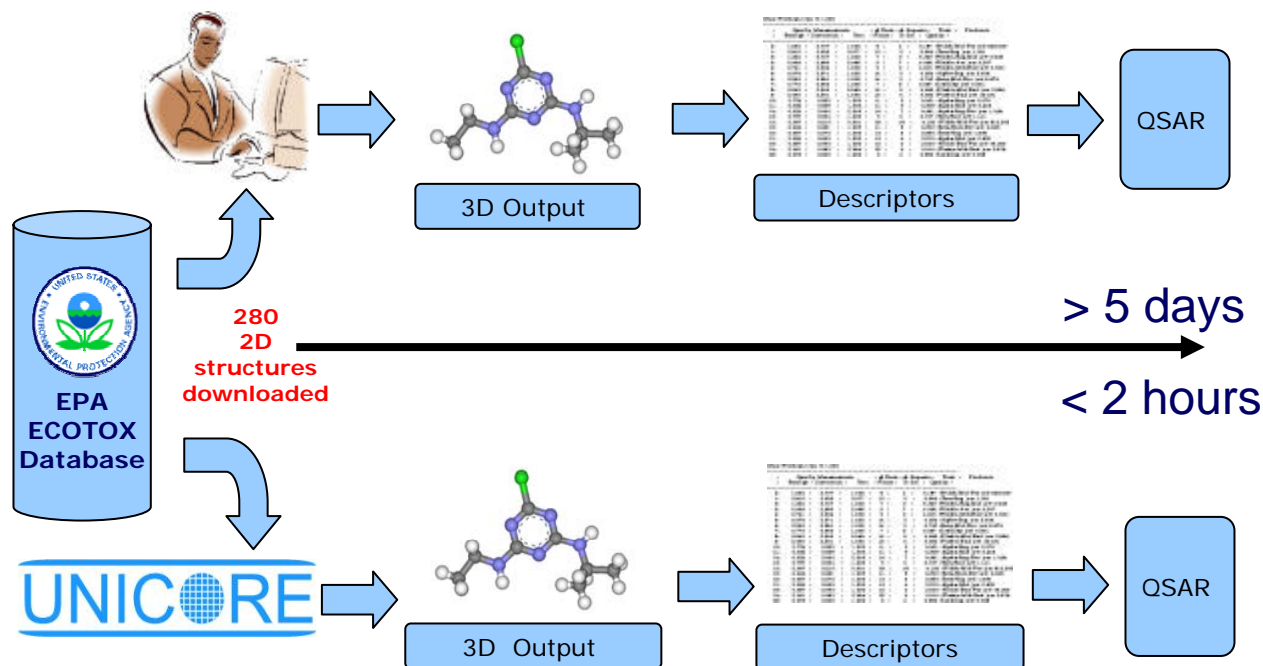
- Characteristics of the problem:
 - High calculation complexity
 - Large amounts of data
 - Integration of different applications
 - Secure access to distributed data
 - Collaborative work



A real Grid application

OpenMolGRID: Solution

- Usage of UNICORE, integration of applications
- Extended workflow support to automate processes
- Interface to databases
- API & command line interface



Lessons Learned

- Taught by end-users and software developers from
 - Industry: pharmaceutical, petrochemical, bio-molecular, weather prediction, automotive, engineering
 - Research: astrophysics, quantum physics, material science, medicine, biology, chemistry
- Deployment of new production software has to offer added value
 - Ease usage, increase effectiveness, decrease cost, ...
- Users have to be stimulated and encourage to
 - use Grid technology for applications, computations, data transfer and access to resources
 - adapt/integrate their applications to/into Grids
- Operation of production environments is costly
 - Certification authority, administrative tools, integration into site management, licenses, ...
 - Common production environment difficult to maintain



More lessons learned

- Fulfilment of functional requirements is not enough
- Users want
 - software of high quality, especially high reliability and resilience
 - help to overcome initial hurdles like
 - obtaining certificates, adapt applications, ...
 - 24/7 availability of the Grid infrastructure
 - 24/7 availability of the Grid experts
 - support hotline, help desk, mailing lists, ...
 - long-term commitment for continuous development and support
 - workshops, hands-on training, ...



Conclusions

- Production Grids are possible
 - Easy/unified usage, integration of legacy applications, cost reduction, ...
 - When will we see the WS-* impact?
- Users demand a fully-fledged product
 - Functions, but also support, support, support
 - Continuity is crucial
- Open Source distribution is the right way
 - Source for bug reports, requirements, ...
 - Higher visibility & community building

