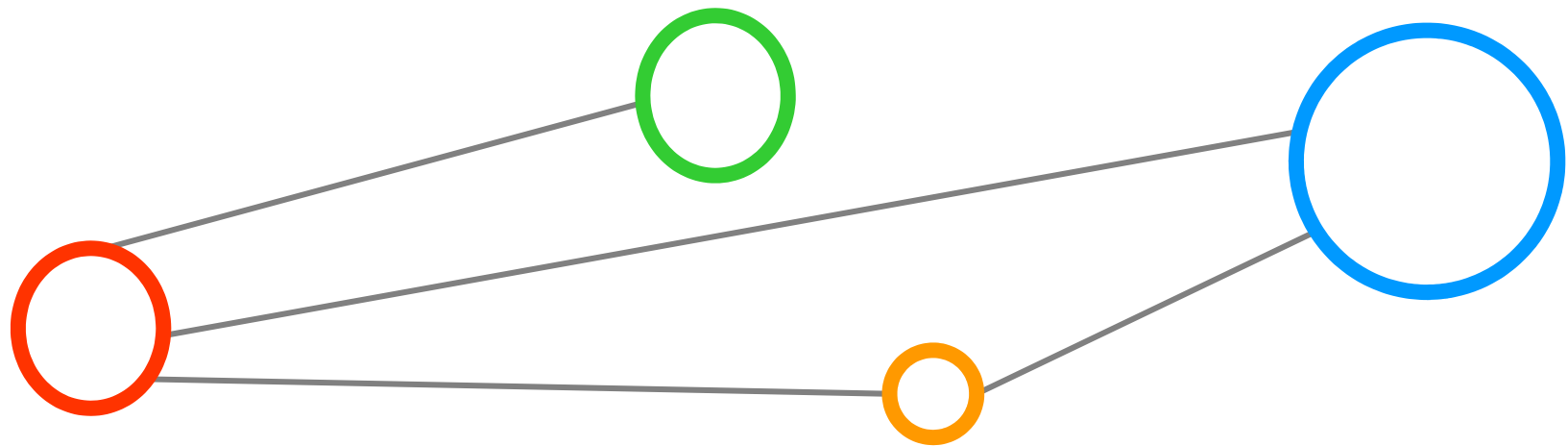**Open Grid Forum**

OPEN FORUM | OPEN STANDARDS

Interoperability of HTC & HPC e-Science Infrastructures

WISDOM - Improving the in-silico drug discovery process

Morris Riedel (DEISA & Jülich Supercomputing Centre)

Group Co-Chair Grid Interoperation Now & Production Grid Infrastructure

Outline

www.ogf.org

Outline

- Outline
- Scientific Project WISDOM
  - *Addressing ‚Project Description and Project Participants'*
- e-Infrastructure Setup
  - *Addressing ‚how is the infrastructure provided'*
- Interoperability Framework
  - *Addressing ‚what level of interoperability is required'*
- Other Scientific Applications
- Conclusions
- References

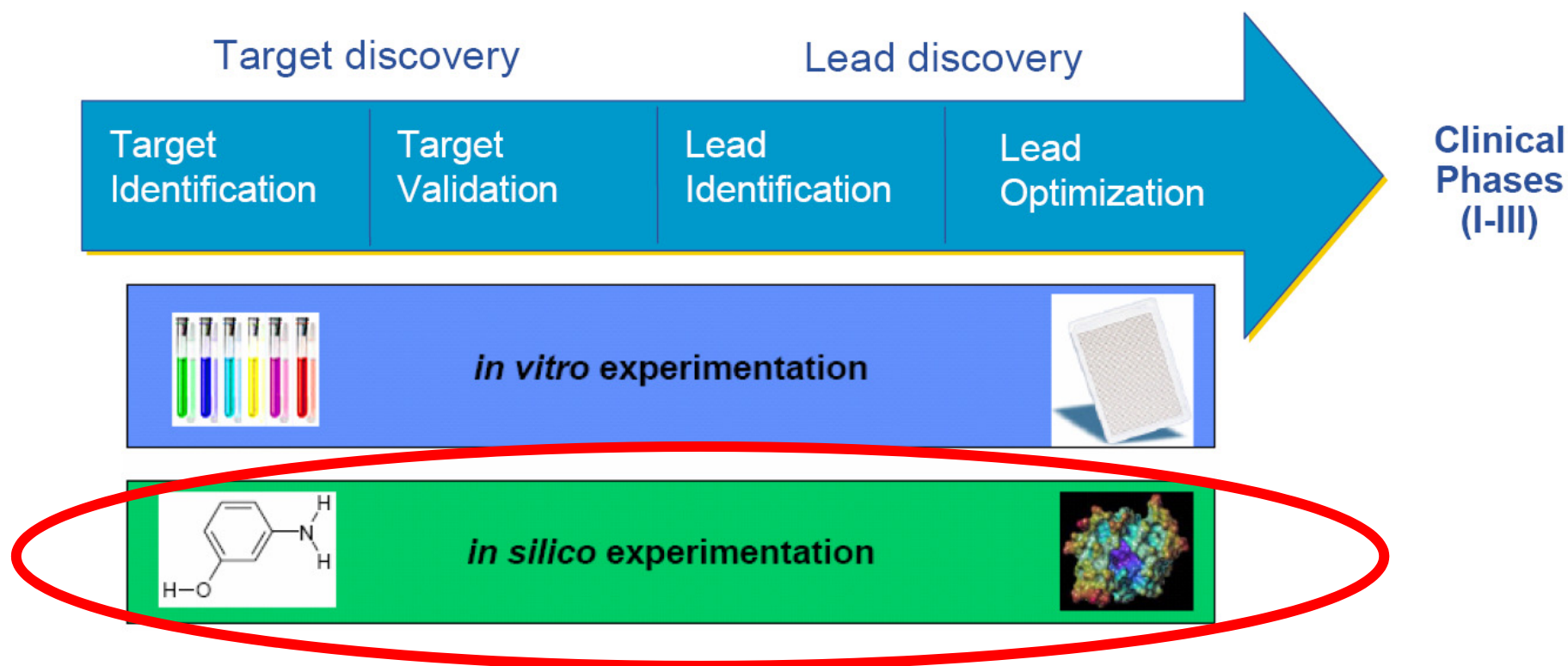# Scientific Project WISDOM

www.ogf.org
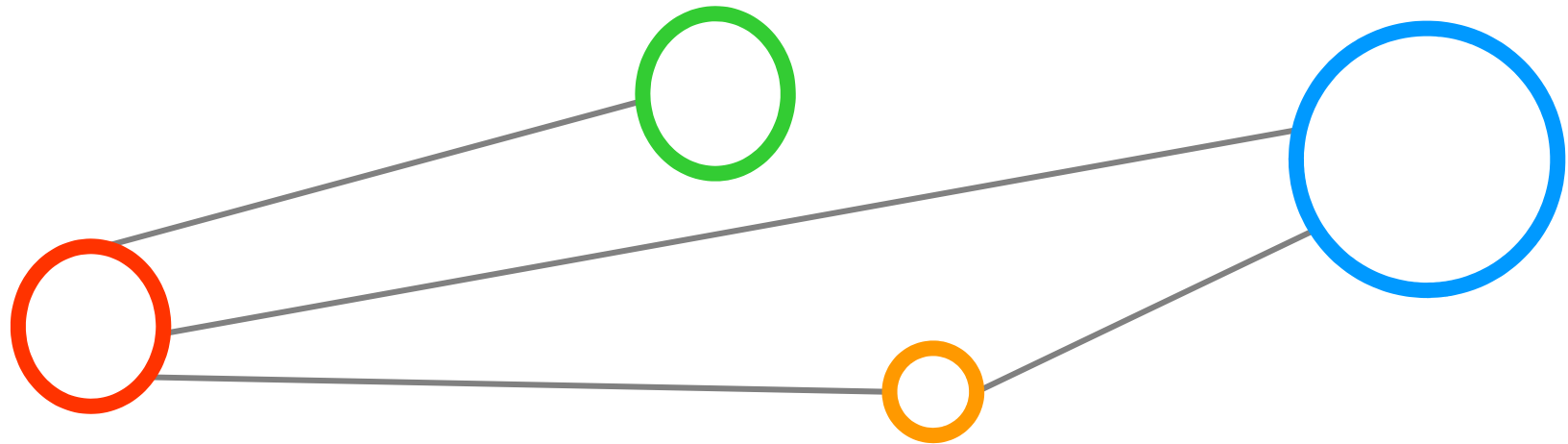
Scientific Project WISDOM

- Wide In Silicio Docking On Malaria (WISDOM)
  - Developing new drugs for neglected and emerging diseases with an initial particular focus on malaria
  - More recently 'WISDOM' stands for a broader science collaboration

- Fundamental goals
  - Accelerate research & drug development for emerging and neglected diseases (e.g. Malaria, Avian Flu, etc.)
  - Reduced research & drug development costs

- Basic project ideas
  - Use of computational-based so called 'in-silico methods' & push in-vitro afterwards
  - Intensive use of (European) e-science (i.e. Grid) infrastructures

In silico drug discovery

A human grid initiative.

- **Problem: development of a drug takes 12 to 15 years and costs approximately 800 million dollars**

Target discovery | Lead discovery

| Target Identification | Target Validation | Lead Identification | Lead Optimization | Clinical Phases (I-III) |

in vitro experimentation

in silico experimentation

[1] Dr. Nicolas Jacq

Project Members

OpenGridForum

eGee
Enabling Grids
for E-sciencE

[4] EGEE

BioSolveIT
CNR-ITB
CNRS
CEA
Healthgrid
IN2P3
LPC
SCAI Fraunhofer
Università di Modena e Reggio Emilia
Université Blaise Pascal
University of Pretoria
University of Los Andes

WISDOM
Initiative for grid-enabled drug discovery
against neglected and emergent diseases

[5] WISDOM

JÜLICH
FORSCHUNGSZENTRUM

wisdom.healthgrid.org

- Jülich Supercomputing Centre as DEISA contact

Distributed
European
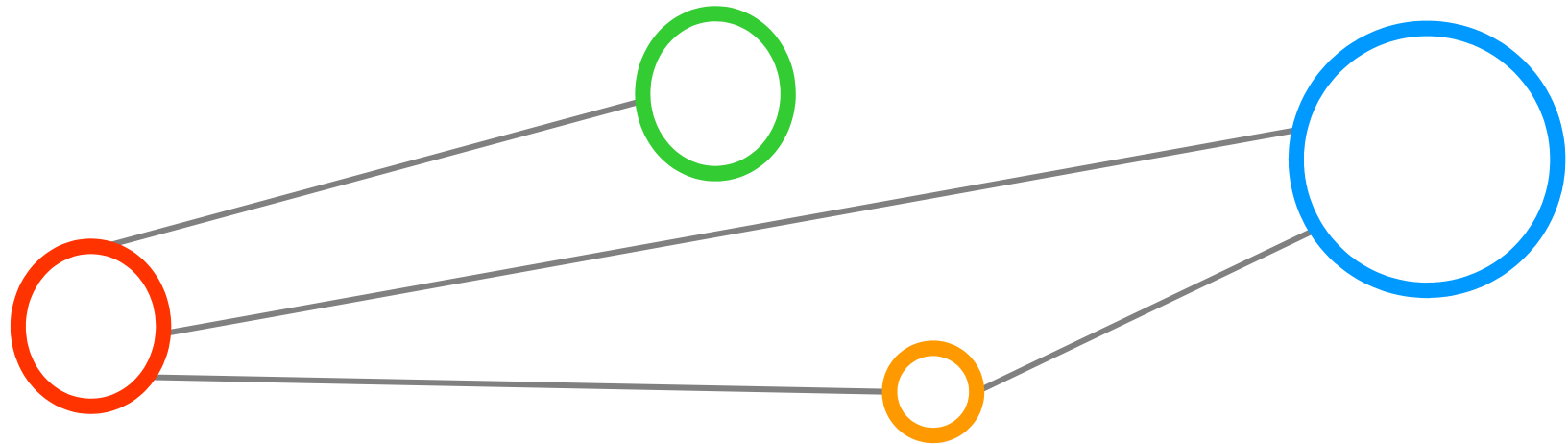Infrastructure for
Supercomputing
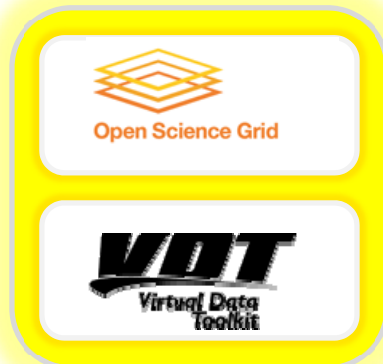Applications

[6] DEISA

7

www.ogf.org

e-Science Infrastructure Setup

www.ogf.org

e-Science Infrastructures

High Throughput Computing (HTC) Infrastructures | High Performance Computing (HPC) Infrastructures

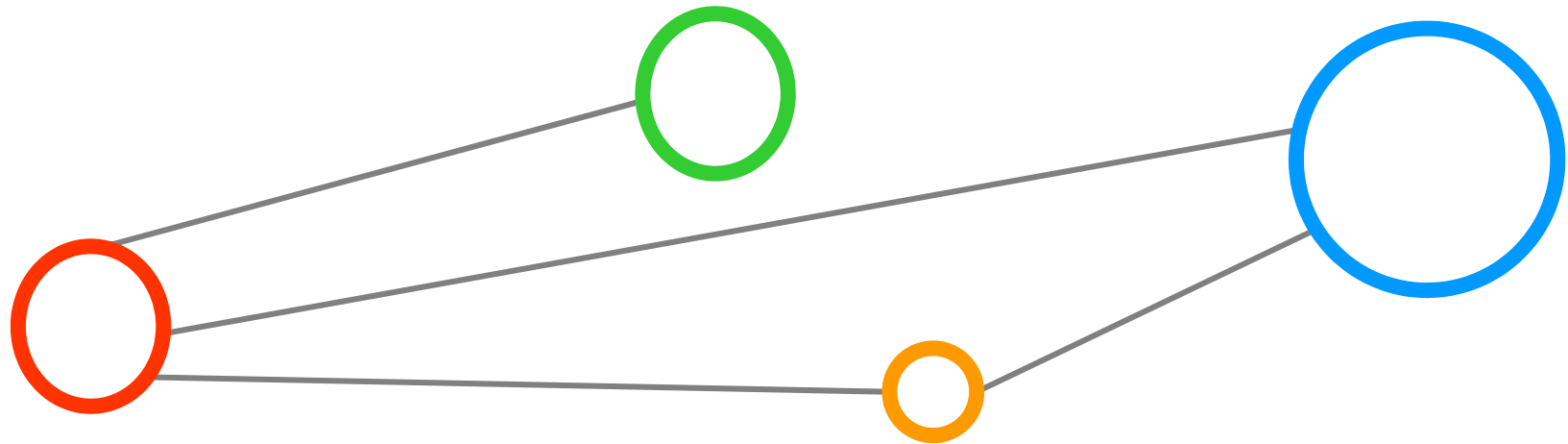www.ogf.org

e-Science Infrastructure Setup

- WISDOM uses EGEE for large scale in silicio docking
  - Computational method for prediction of whether one molecule will bind to another
  - Using AutoDock and FlexX software provided via gLite in EGEE
  - Output is a list of best chemical compounds (potential drugs)
  - That is not the final solution, only a potential list of drugs

- Refine best compound list via molecular dynamics (MD)
  - Simulate the docked compounds 'over time' for verification
  - Fast MD computations use AMBER (or scalable NAMD) in DEISA
  - AMBER (Assisted Model Building with Energy Refinement) , version 9

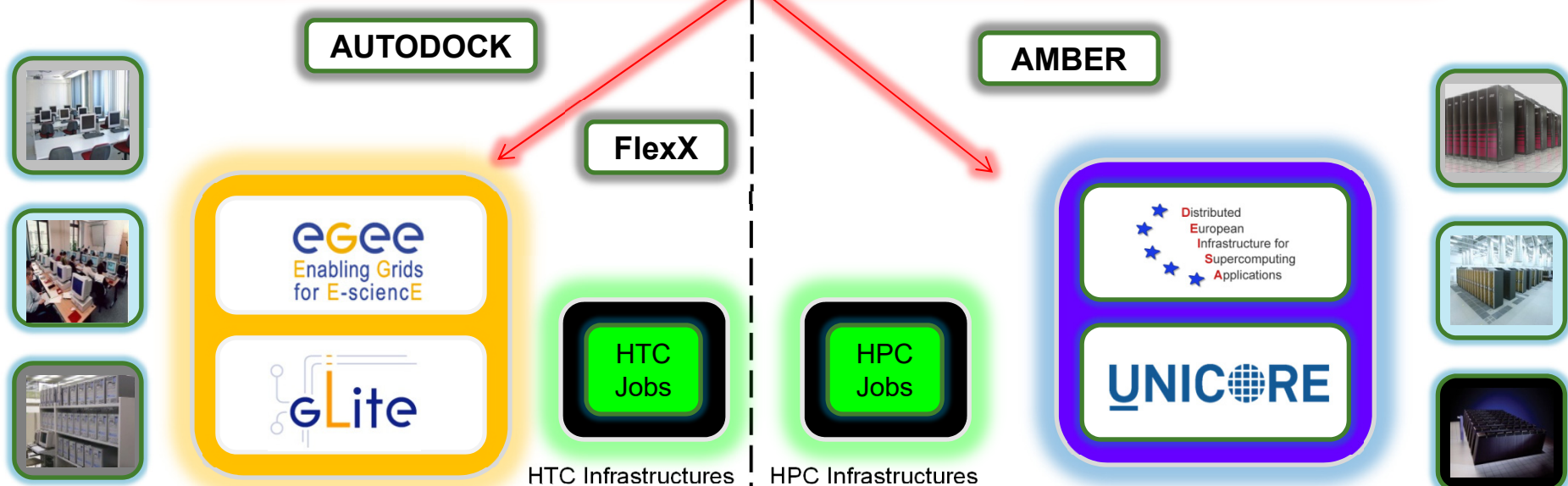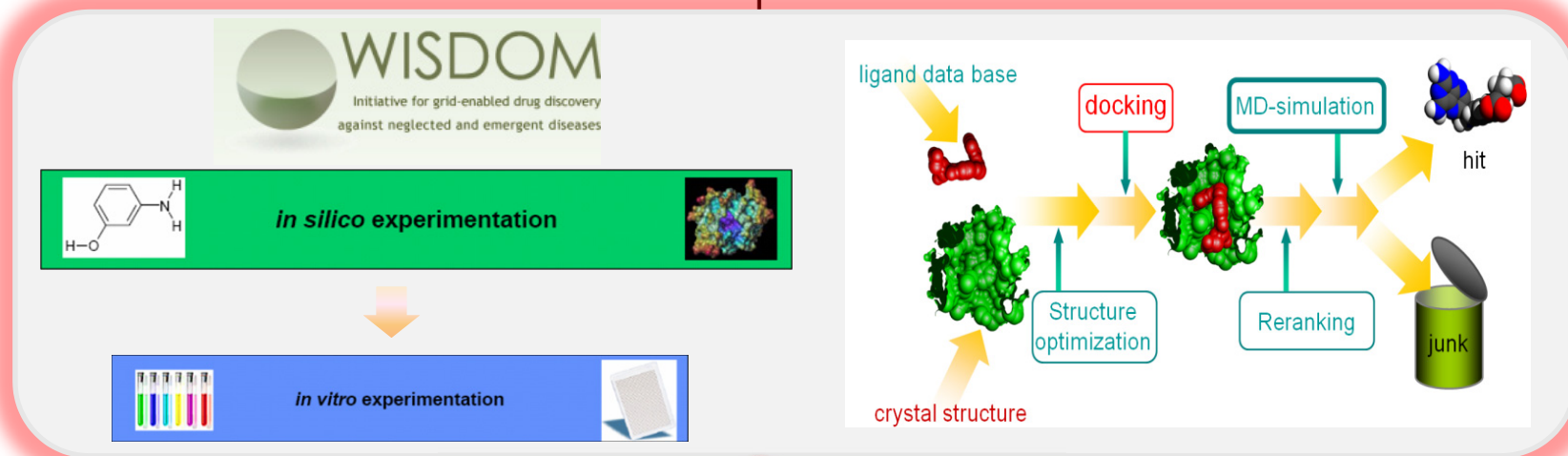- **Goal: Accelerate drug discovery using EGEE + DEISA**

egee
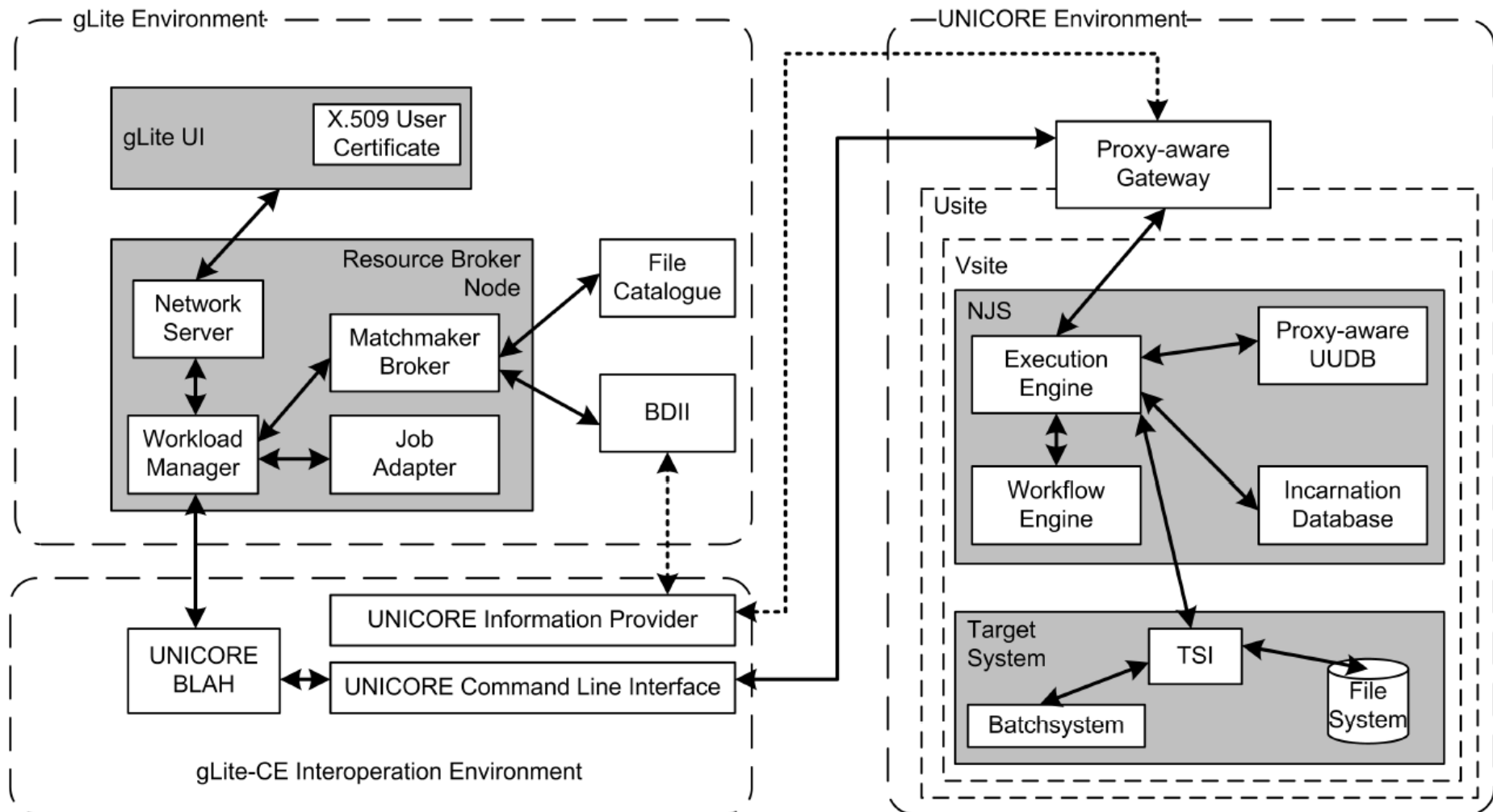**Enabling Grids for E-sciencE**

Distributed European Infrastructure for Supercomputing Applications

www.ogf.org

[2] Riedel et al., *'Improving e-Science with Interoperability of the e-Infrastructures EGEE and DEISA'*

Interoperability Framework

Interoperability Framework

AUTODOCK

AMBER

FlexX

HTC Jobs

HPC Jobs

HTC Infrastructures

HPC Infrastructures

www.ogf.org

[2] Riedel et al., *'Improving e-Science with Interoperability of the e-Infrastructures EGEE and DEISA'*

Initial ‚Proprietary' Approach

[7] Riedel et al., *'Interoperation of world-wide production e-science Infrastructures'*
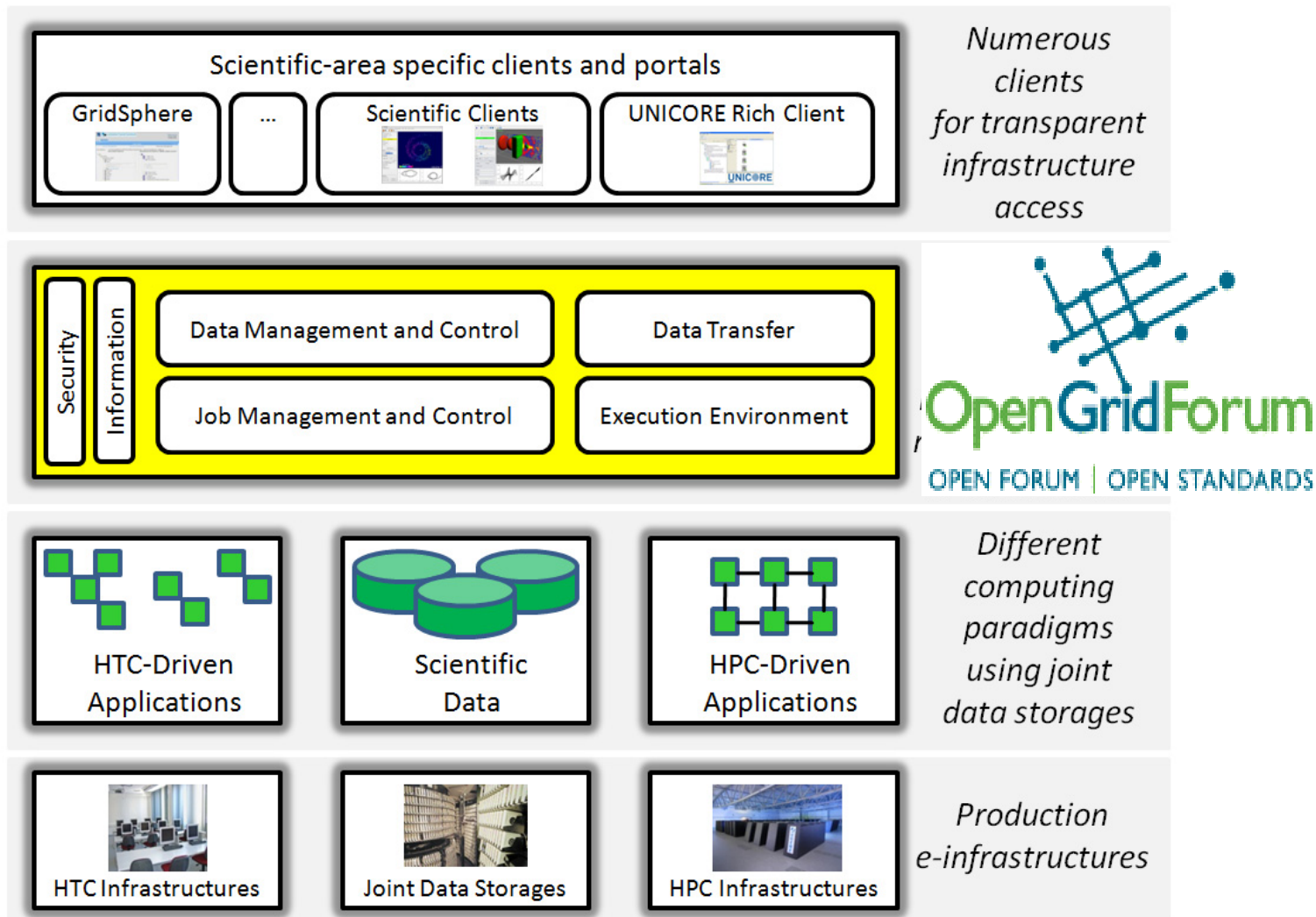
Interoperability Components (1)

[3] Riedel et al., 'Research Advances by using Interoperable e-Science Infrastructures – The Infrastructure Interoperability Reference Model applied in e-Science '

www.ogf.org

Interoperability Components (2)

Scientific-area specific clients and portals

GridSphere · ... · Scientific Clients · UNICORE Rich Client

Numerous clients for transparent infrastructure access

Security · Information · Data Management and Control · Data Transfer · Job Management and Control · Execution Environment

HTC-Driven Applications · Scientific Data · HPC-Driven Applications

Different computing paradigms using joint data storages

HTC Infrastructures · Joint Data Storages · HPC Infrastructures

Production e-infrastructures

[3] Riedel et al., 'Research Advances by using Interoperable e-Science Infrastructures – The Infrastructure Interoperability Reference Model applied in e-Science '
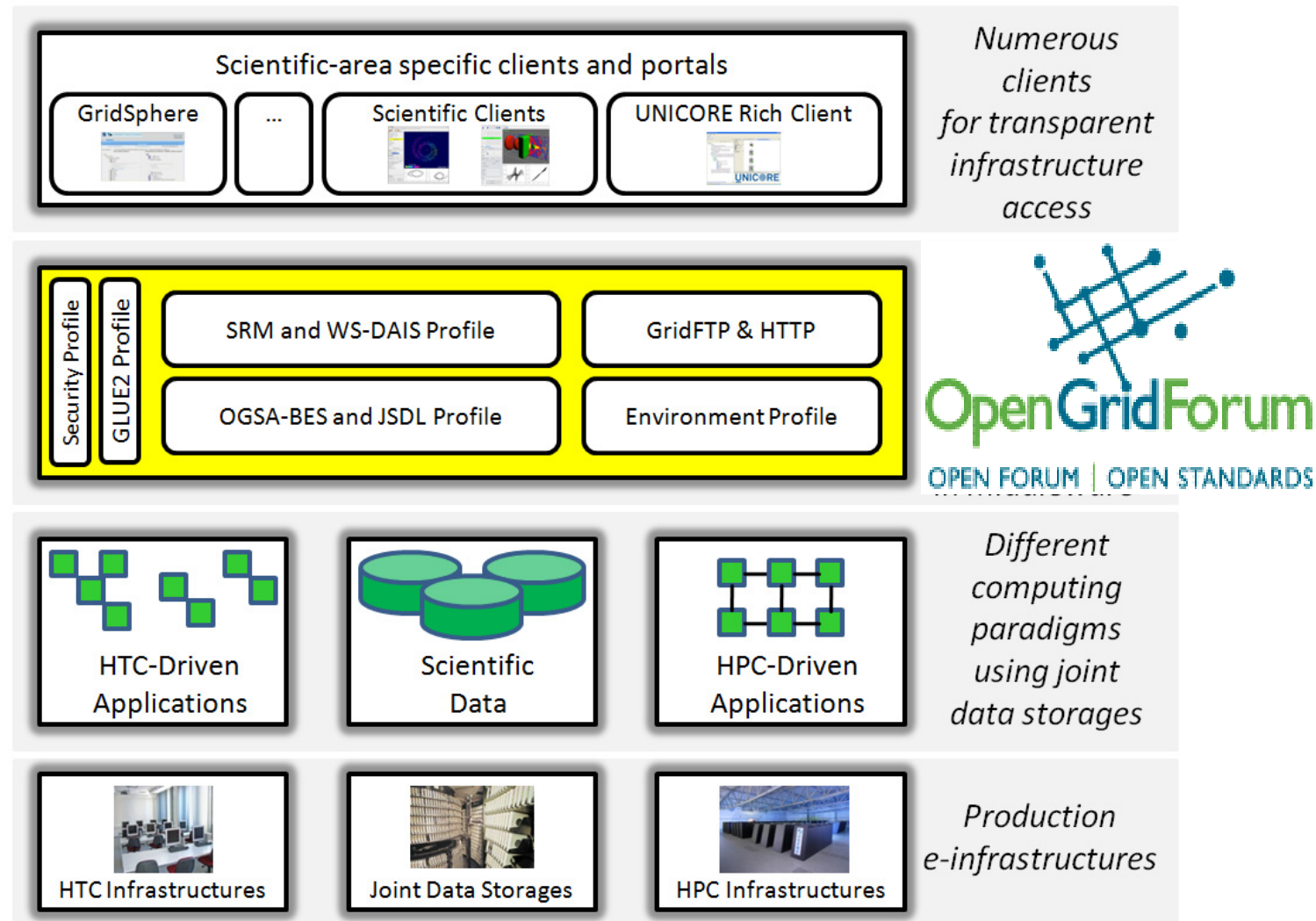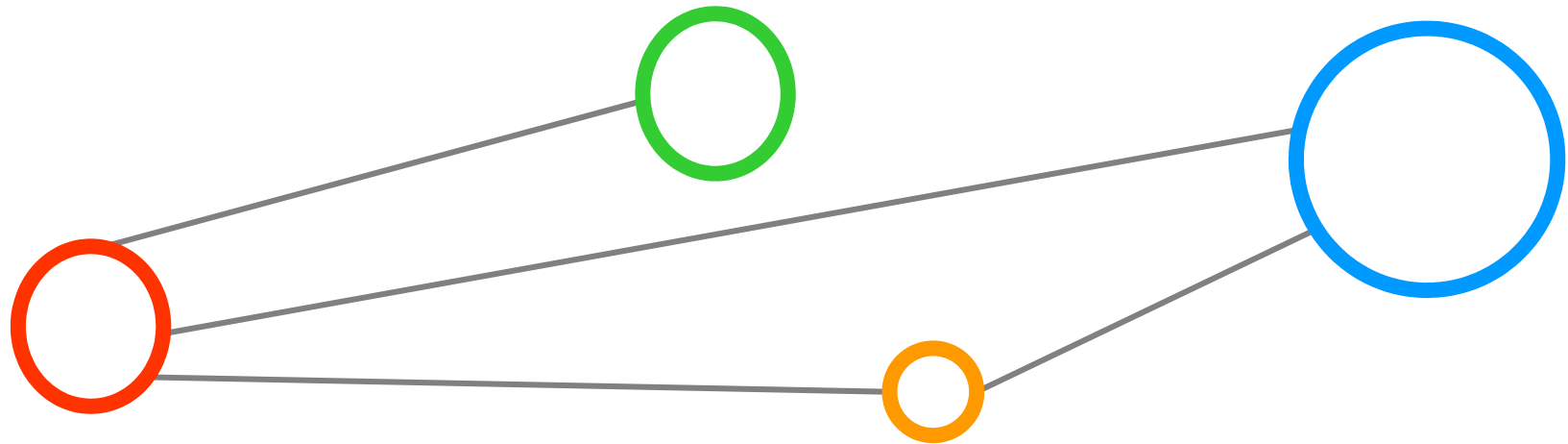
Interoperability Components (3)

[3] Riedel et al., 'Research Advances by using Interoperable e-Science Infrastructures – The Infrastructure Interoperability Reference Model applied in e-Science '

www.ogf.org

Other Scientific Applications

www.ogf.org

Other Scientific Application

- European Fusion for ITER Applications
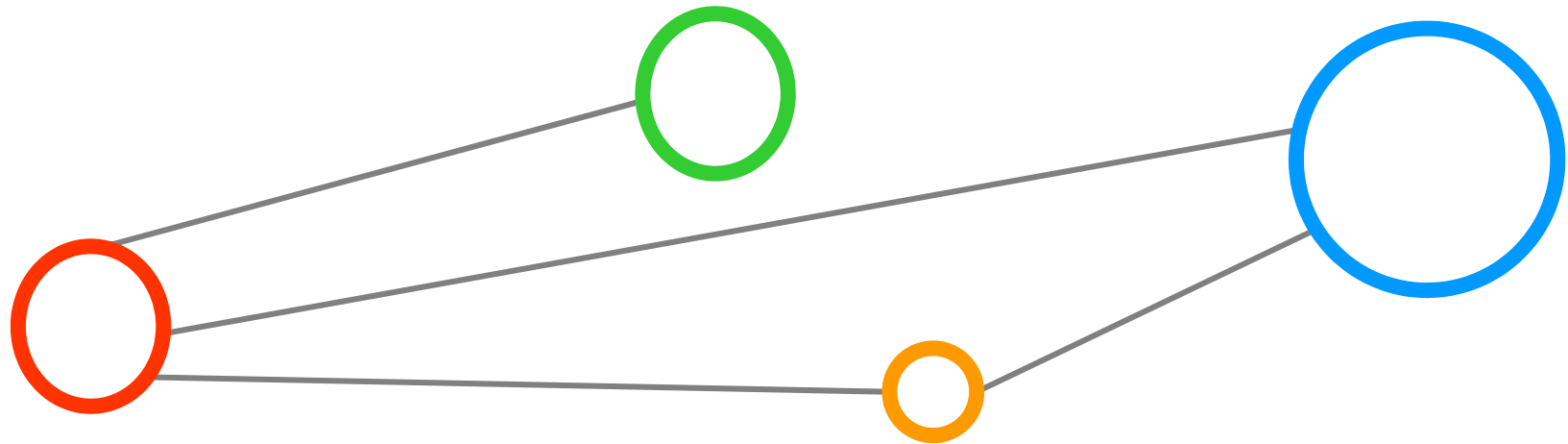 - Towards future electricity producing power plants

- Goal: Provide a comprehensive framework/infrastructure for core and edge transport and turbulence simulation using HTC/Grid and HPC to support the modelling community
  - GIN Demonstration at Supercomputing 2009

- Production e-science infrastructures
  - EGEE for HTC codes
  - DEISA for HPC codes

- Scientific software
  - Kepler workflow tool orchestrating HPC/HTC
  - ~80 codes for HTC and HPC computations

www.ogf.org

Conclusions

www.ogf.org
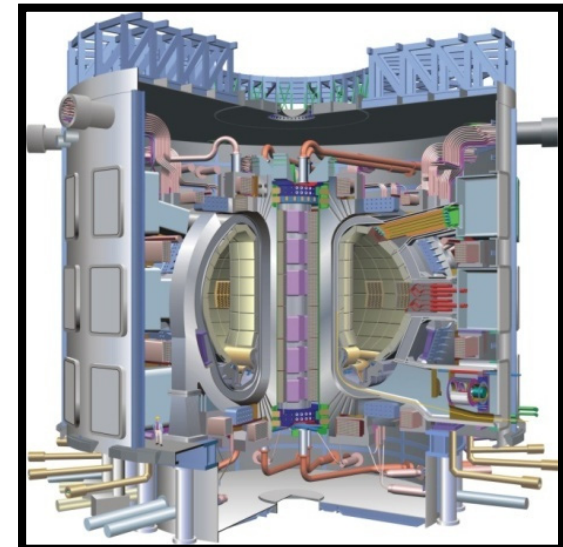
Conclusions (1)

- More and more projects evolve that require interoperability between HTC- & HPC-driven e-science infrastructures

- Interoperability requirements with respect to policies&usage…

- WISDOM is one project that can take advantage of this
  - However the access to computational paradigms (i.e. HTC, HPC) must be significantly easier
- More work needs to be done with licenses
  - E.g. FlexX is good, but commercial – licenses per user vs. per sites
- More work needs to be done in usage policies / access
  - HPC systems overbooked, applying for time with proposals not easy
  - Step from HTC infrastructure to HPC is ‚a very big one' requiring experience in coding (i.e. MPI) in order to achieve a ‚speed-up'

20

Conclusions (2)

- Interoperability requirements with respect to technology…

- Transparency for end-users is the key – one client only(!)
  - Different security models are major showstoppers

- Open Standards already provide a first good step into the right direction – but some need to be refined / improved
  - E.g. High Performance Computing – Basic Profile (HPC-BP)
  - E.g. Storage Ressource Manager (SRM), GridFTP
  - E.g. GLUE2 Information Model

- WISDOM use case experience significantly contributes to the Production Grid Infrastructure (PGI) Working Group
  - Improvements and feedback to common open standards
  - Missing links between OGSA-BES, JSDL, SRM, GridFTP, GLUE2,…

References

www.ogf.org
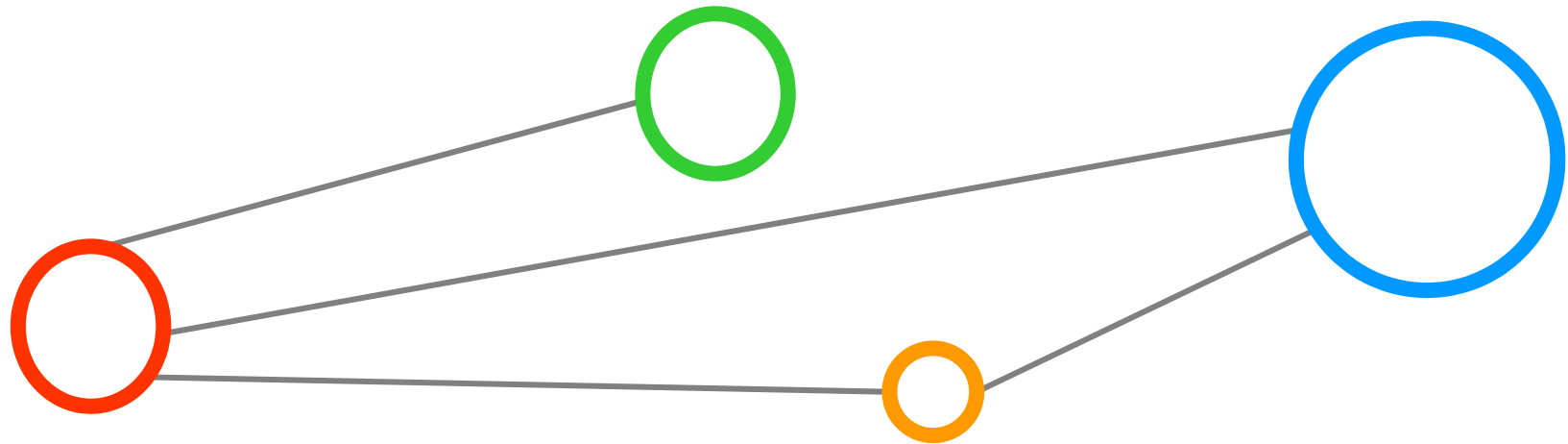
References

[1] Dr Nicolas Jacq, International Symposium on Grids for Science and Business 12 June 2007, http://events.ibbt.be/grid2007/pdf/Jacq.pdf?PHPSESSID=51537714797e1362c6c1e33f554e2dd5

[2] M. Riedel et al., *Improving e-Science with Interoperability of the e-Infrastructures EGEE and DEISA,* Proceedings of the 31st International Convention MIPRO, Conference on Grid and Visualization Systems (GVS), May 2008, Opatija, Croatia, Croatian Society for Information and Communication Technology, Electronics and Microelectronics, ISBN 978-953-233-036-6, pages 225 – 231

[3] M. Riedel, F. Wolf, D. Kranzlmüller, A. Streit, T. Lippert, *Research Advances by using Interoperable e-Science Infrastructures - The Infrastructure Interoperability Reference Model applied in e-Science,* Journal of Cluster Computing, Special Issue Recent Advances in e-Science

[4] Enabling Grids for e-Science (EGEE), http://www.egee-eu.eu

[5] WISDOM Web-page: http://wisdom-healthgrid.org

[6] Distributed European Infrastructure for Supercomputing Applications (DEISA), http://www.deisa.org

[7] M.Riedel and E. Laure et al., *Interoperation of World-Wide Production e-Science Infrastructures*, Concurrency and Computation: Practice and Experience, 21 (2009) 8, 961 – 990, DOI: 10.1002/cpe.1402

Full Copyright Notice

Copyright (C) Open Grid Forum (2009). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works.

The limited permissions granted above are perpetual and will not be revoked by the OGF or its successors or assignees.