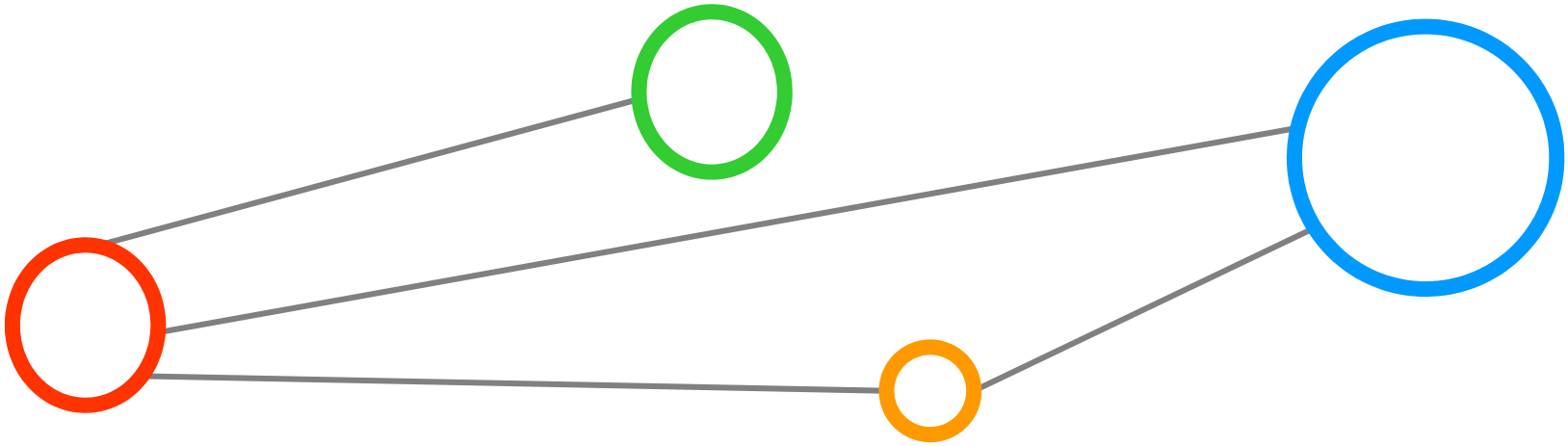
WSEAS Intern. Conferences, September 2009,Vouliagmeni Beach, Athens, Greece

Concepts and Design of an Interoperability Reference Model for Scientific- and Grid Computing Infrastructures

Morris Riedel (Juelich Supercomputing Centre, DEISA) et al.
Group Co-Chair Grid Interoperation Now & Production Grid Infrastructure

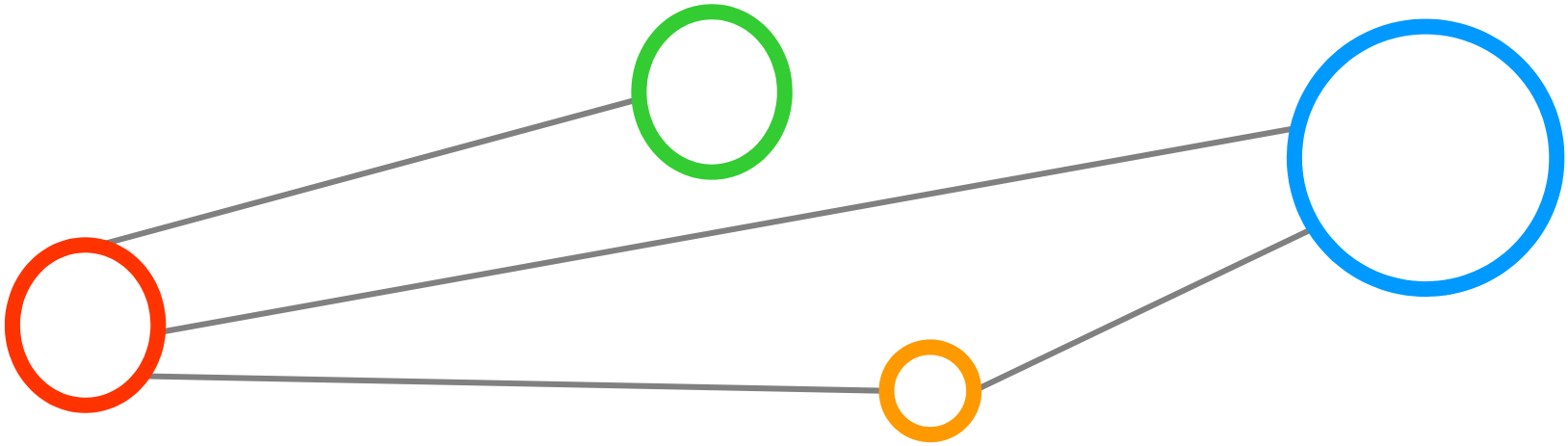
Outline



Outline

- e-Science 101
- Motivation for Interoperability
- Emerging Open Standards
- Interoperability Reference Model
- Computing Refinement Concepts
- Conclusion

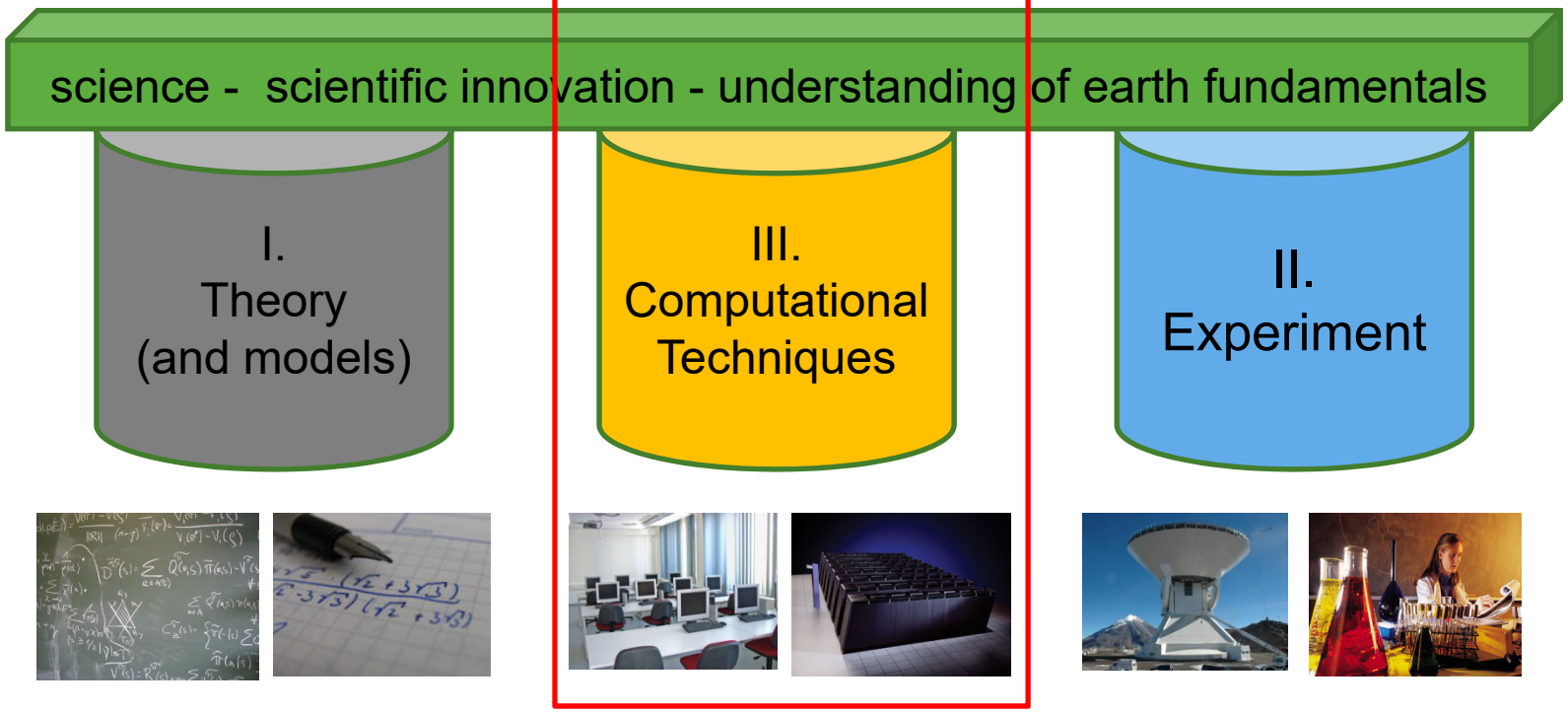
Motivation for Interoperability



Traditional Scientific Computing

OpenGridForum

*'Today, the natural sciences regard **computational techniques** as a third pillar alongside experiment and theory'*

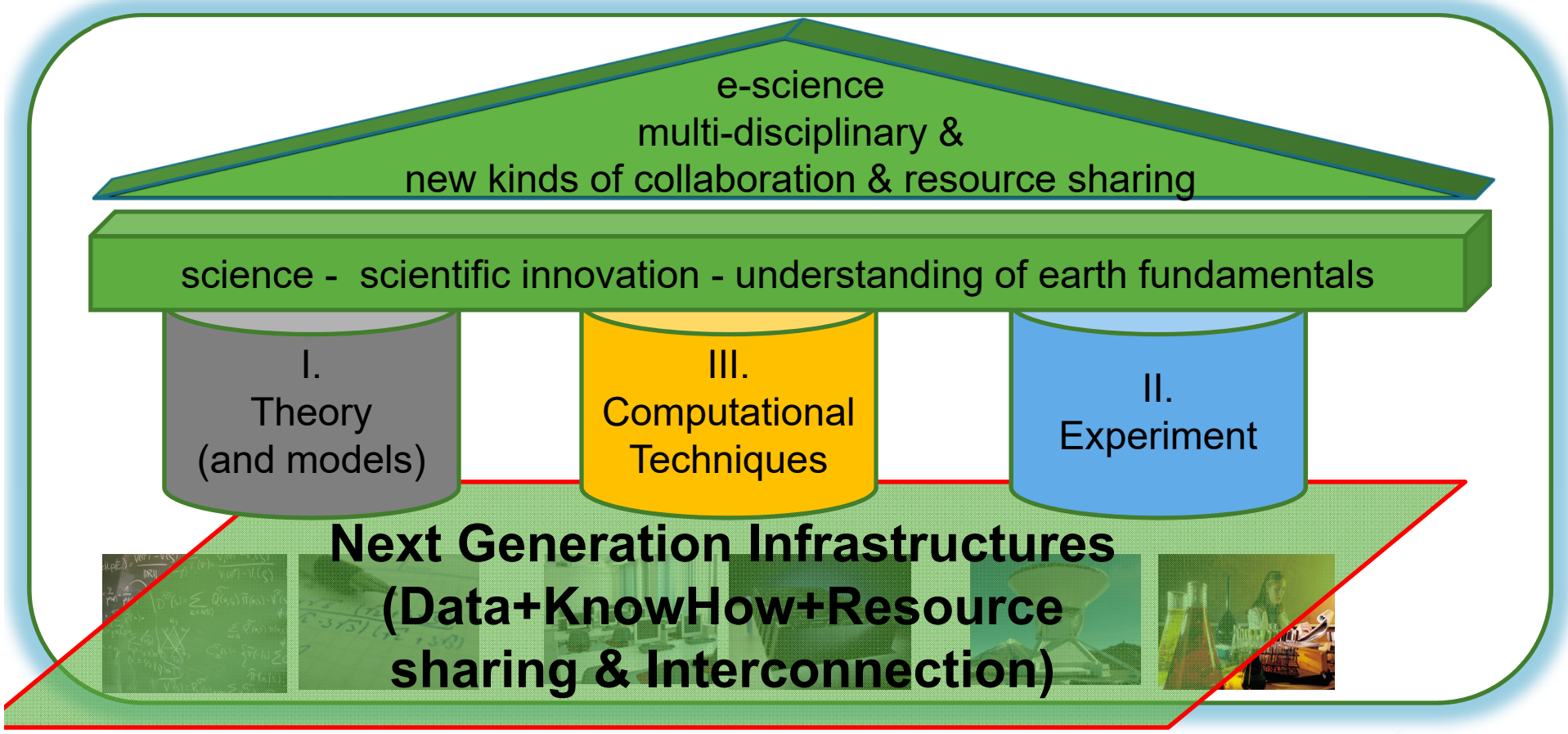


[1] Lippert et al., Interview *'Europe's biggest supercomputer'*

Enhanced Science (e-Science)

OpenGridForum

*'e-Science is about global collaboration in key areas of science and the **next generation infrastructure** that will enable it'*



[2] John Taylor, *'The definition of e-science'*

Production Grid Infrastructures

High Throughput Computing (HTC) Infrastructures

High Performance Computing (HPC) Infrastructures

Different Technologies

Open Science Grid

VDT Virtual Data Toolkit

NGS National Grid Service

omii-uk Software Solutions for e-Research

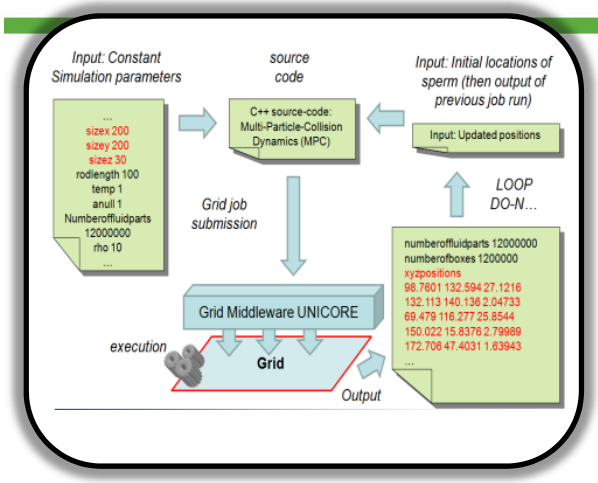
TeraGrid

the globus toolkit

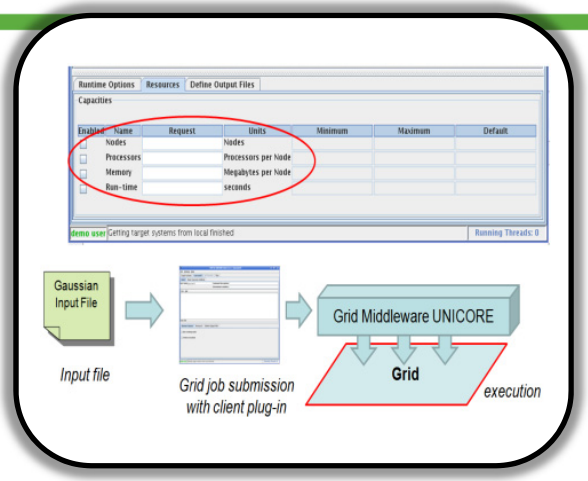
High Throughput Computing (HTC) Infrastructures

High Performance Computing (HPC) Infrastructures

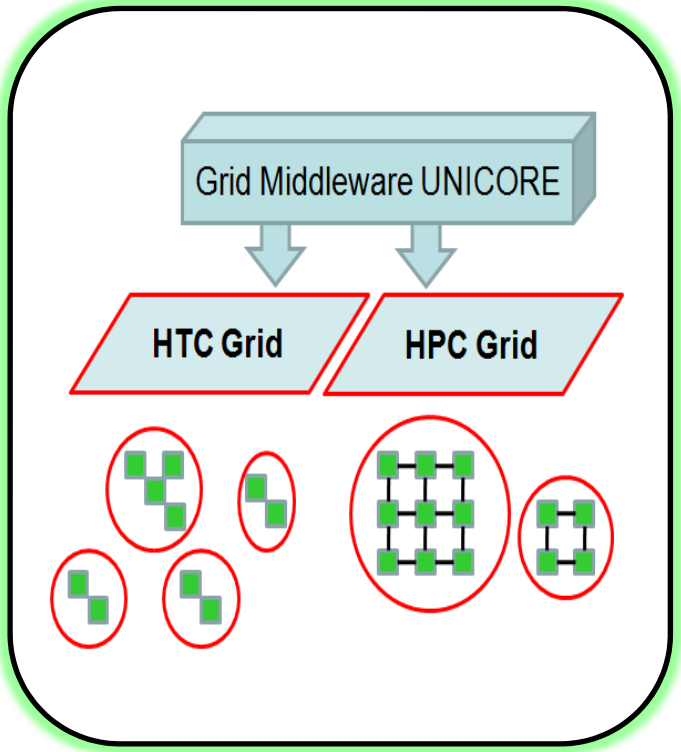
Different Approaches for e-Science



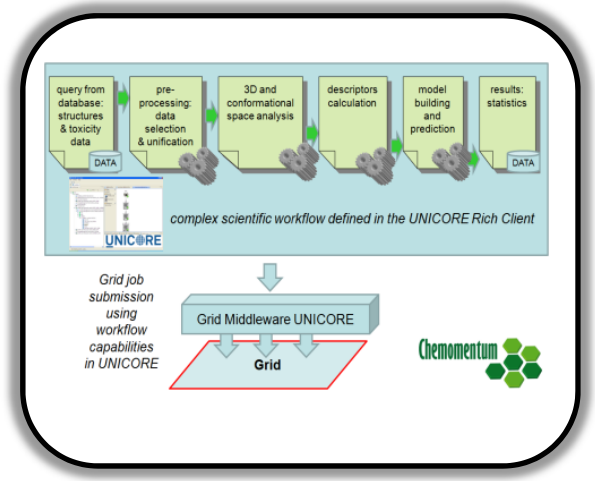
Simple Scripts & Control



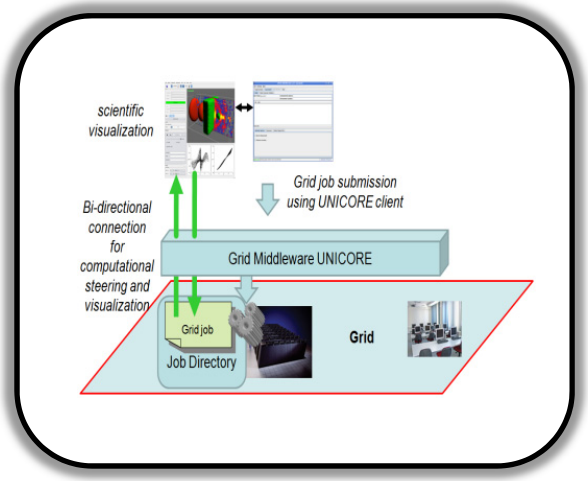
Application Plug-ins



Grid Interoperability



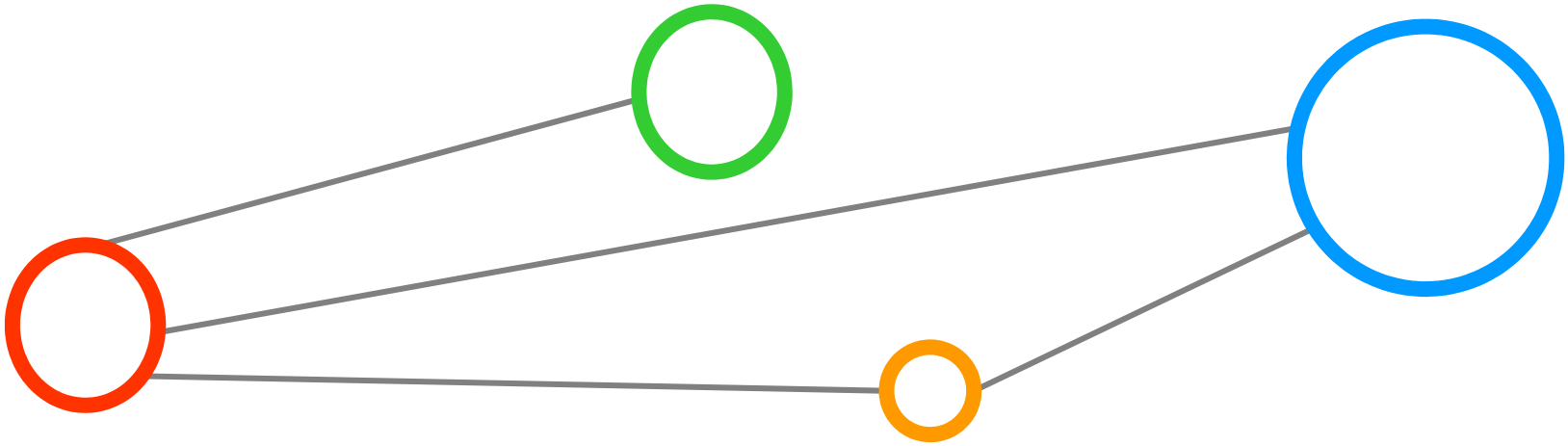
Complex Workflows



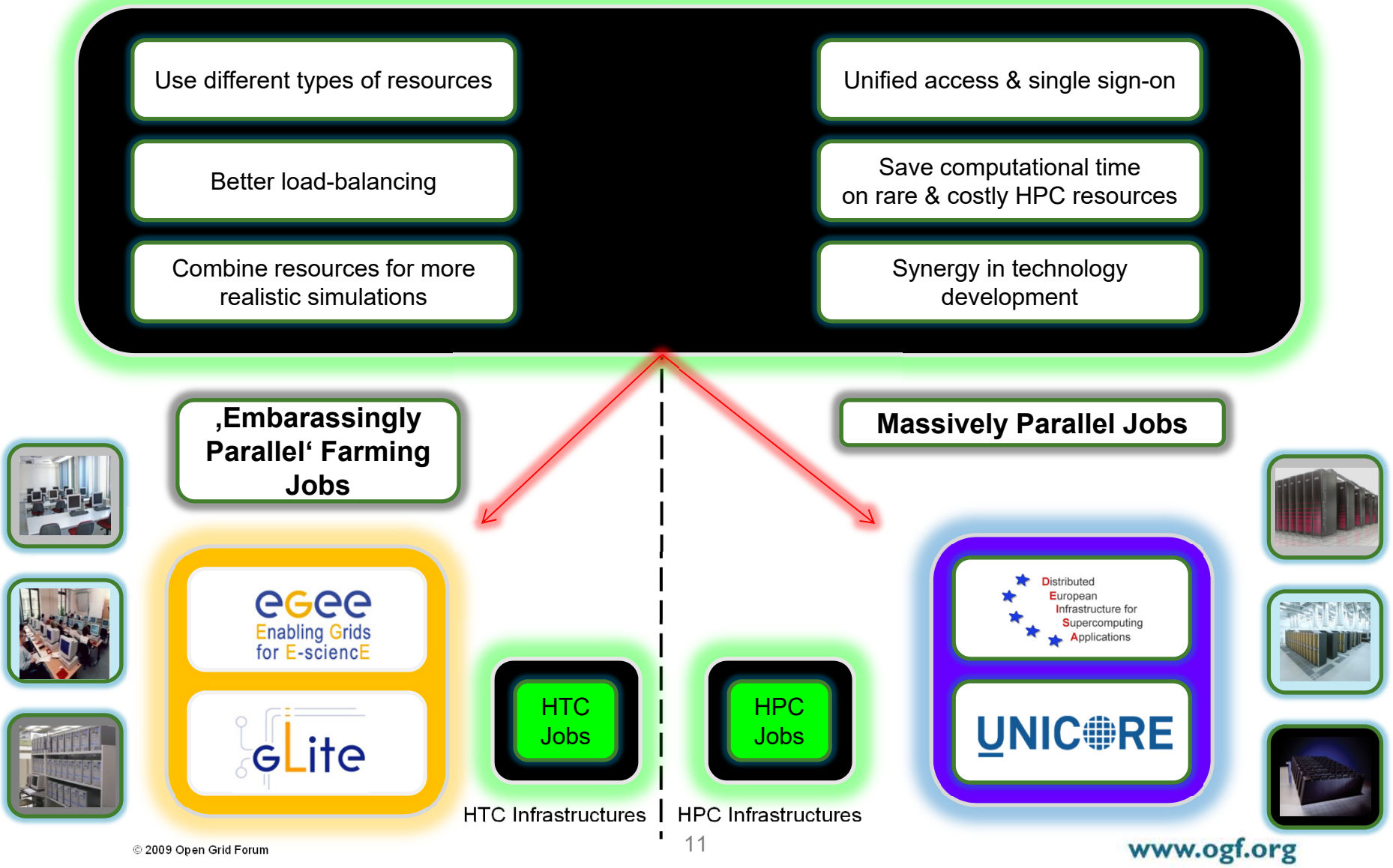
Interactive Access

[3] Riedel and Kranzlmueller et al.,
'Classification of Different Approaches for e-Science Applications in Next Generation Computing Infrastructures '

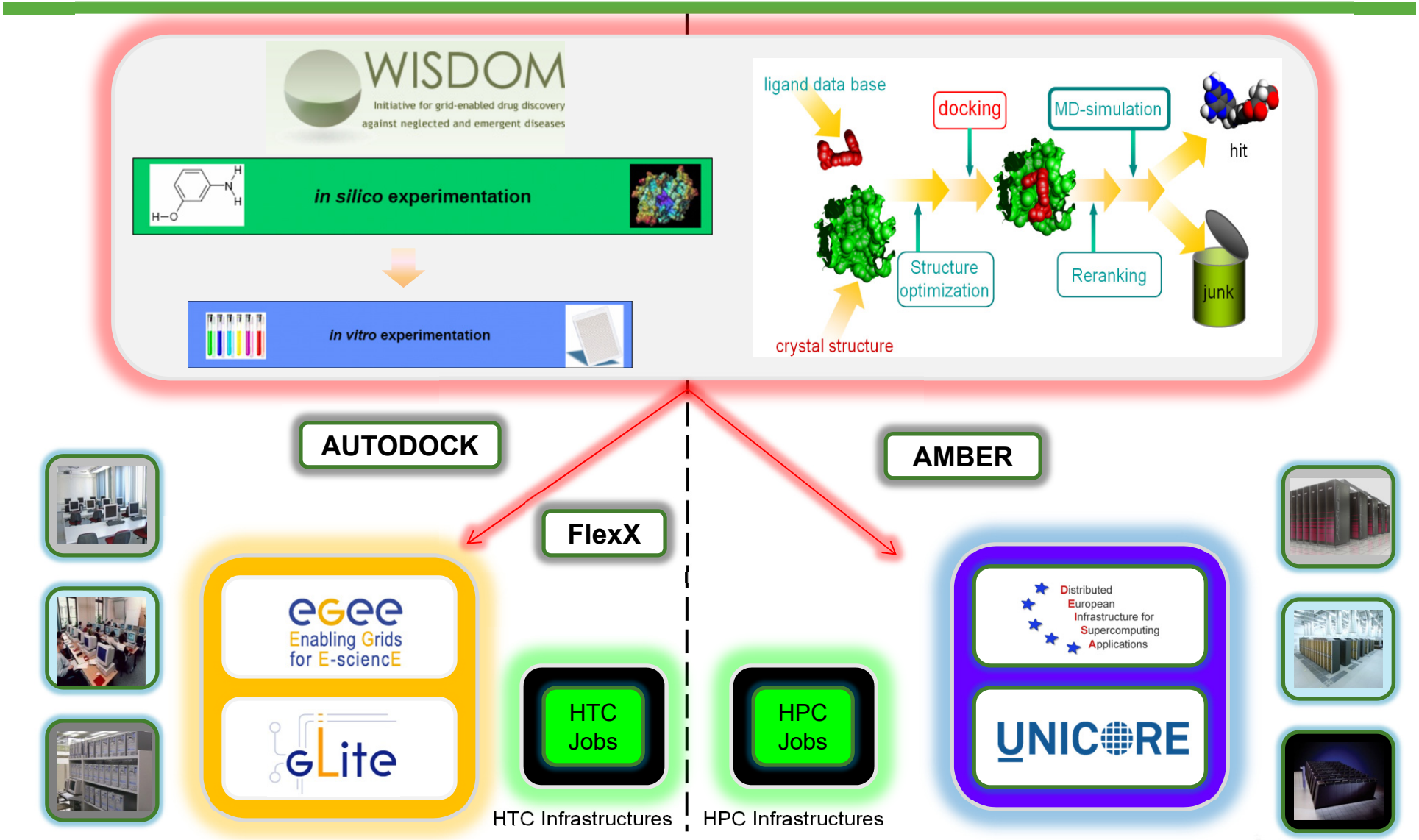
Motivation for Interoperability



Motivation



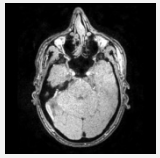
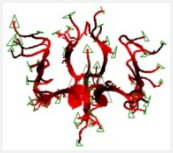
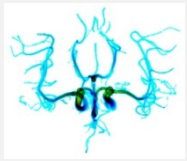
e-Health Use Case HTC/HPC



www.ogf.org

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

e-Health Use Case HPC/HPC

Virtual Physiological Human
network of excellence

Quantify uncertainties, reduce time-to-solution, different job runs with same code ‚same time'

HEMELB

HEMELB



NGS
National Grid Service

omii-uk
Software Solutions
for e-Research

HPC Infrastructure

JÜLICH
FORSCHUNGSZENTRUM

rzg
RECHEN-
ZENTRUM
GARCHING

Distributed
European
Infrastructure for
Supercomputing
Applications

UNICORE

HPC Infrastructure



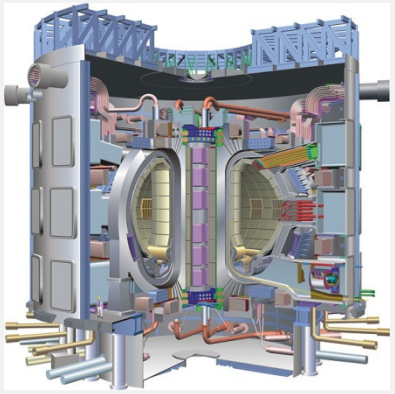
[7] Manos & Riedel et al., *DEISA Newsletter December 2008*

Fusion Use Case Example

OpenGridForum

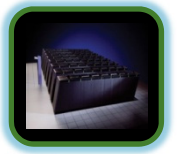
EUFORIA

Advanced cross-computational paradigm simulation of future power generating power plants



Fusion HTC code suite

Fusion HPC code suite



egee Enabling Grids for E-sciencE

gLite

HTC Jobs

HPC Jobs

Distributed European Infrastructure for Supercomputing Applications

UNICORE

HTC Infrastructures

HPC Infrastructures

www.ogf.org

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

Challenges

One Client (command-line, portal, application with integrated API)

Embarassingly Parallel Jobs

Massively Parallel Jobs

Different job description languages

Different Data Transfer Techniques

Different security setups

Different job submission interfaces & protocols

Different Storage Access Techniques

Different information semantics

eGee Enabling Grids for E-sciencE

gLite

HTC Jobs

HPC Jobs

HTC Infrastructures

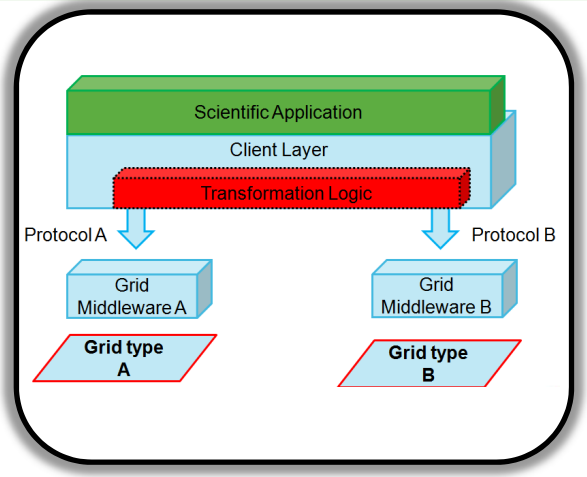
HPC Infrastructures

Distributed European Infrastructure for Supercomputing Applications

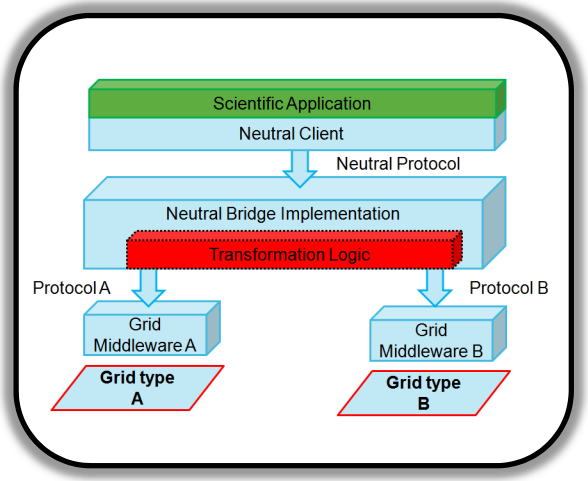
UNICORE

[5] Riedel et al., *'Experiences and Requirements for Interoperability between HTC- and HPC-driven e-Science Infrastructures'*

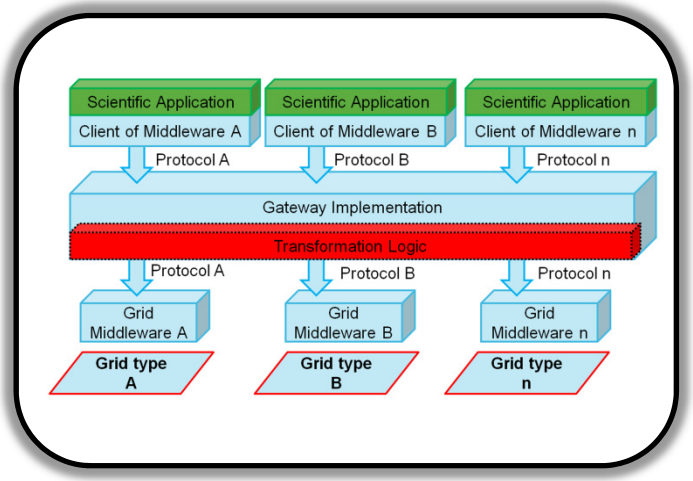
Different Approaches for Interoperability



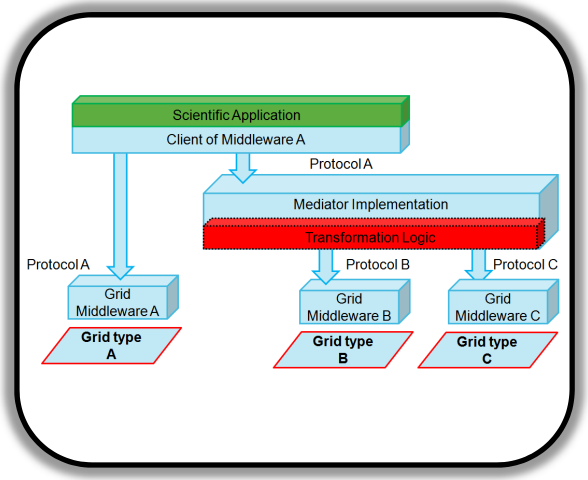
Client Layer Approach



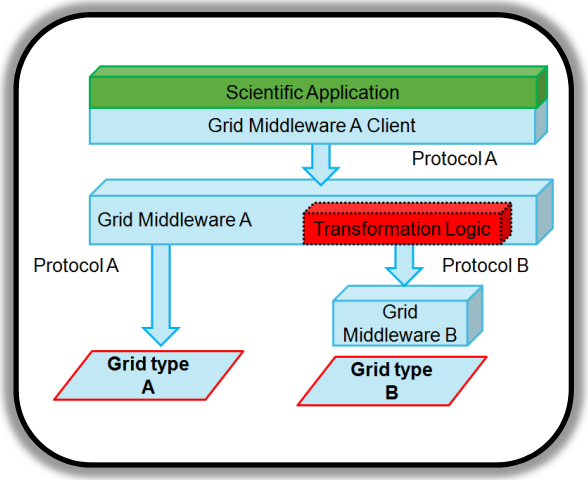
Neutral Bridge Approach



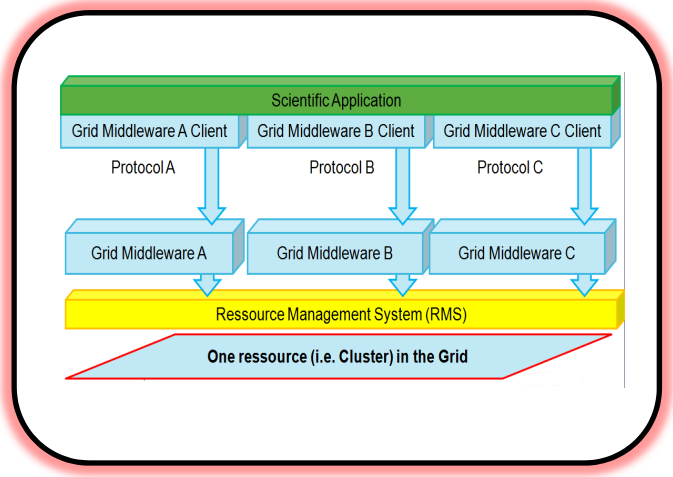
Gateway Approach



Mediator Approach



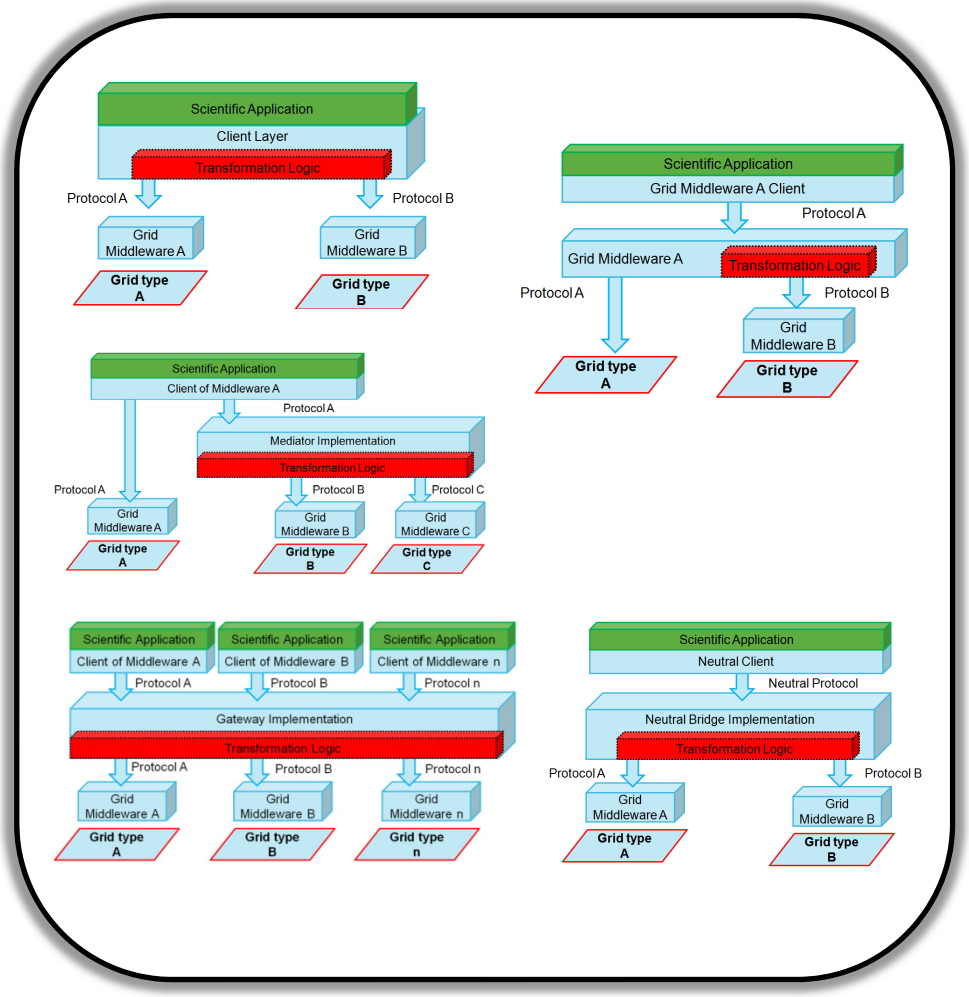
Adapter Approach



Middleware Co-Existence

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

Transformation Logic



Approaches that require transformation logic

protocol A or schema A

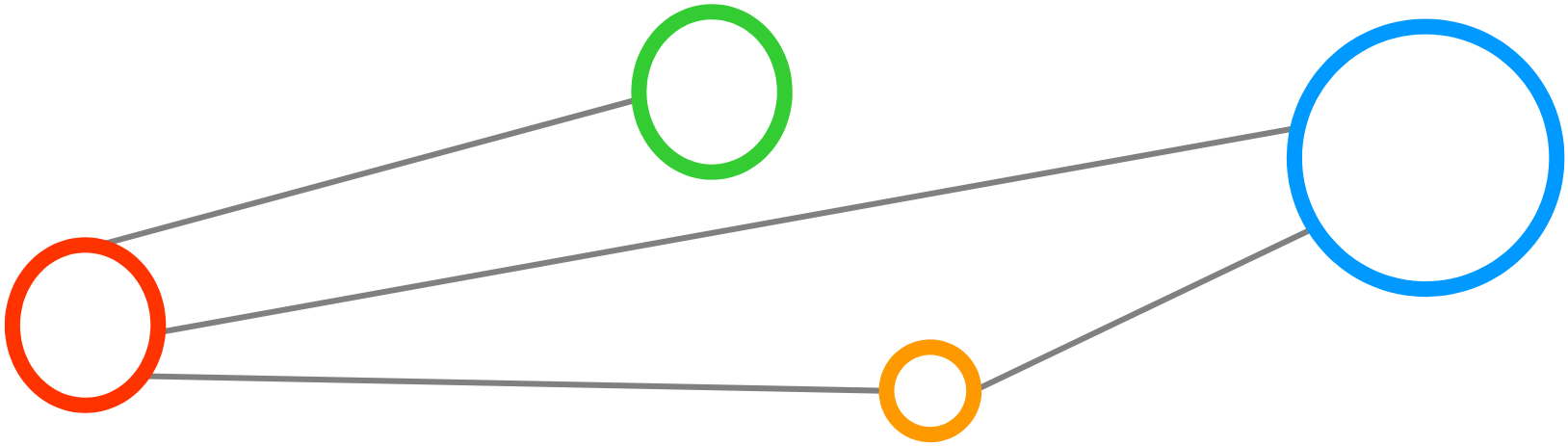
Transformation Logic

- Time consuming & error-prone
- Difficult to maintain (n*n versions)
- Protocol B‘ implies result is often only a subset of protocol B

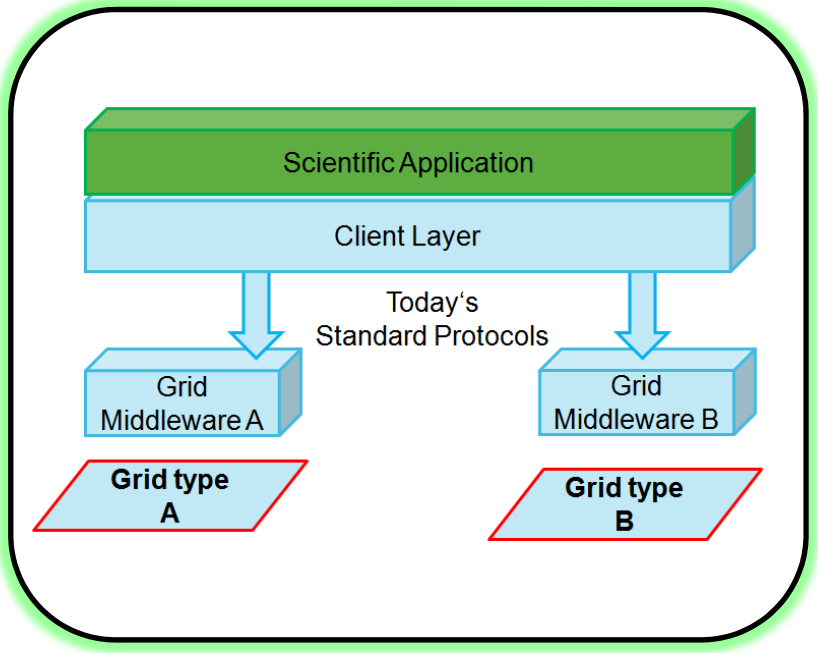
protocol B‘ or schema B‘

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

Emerging Open Standards



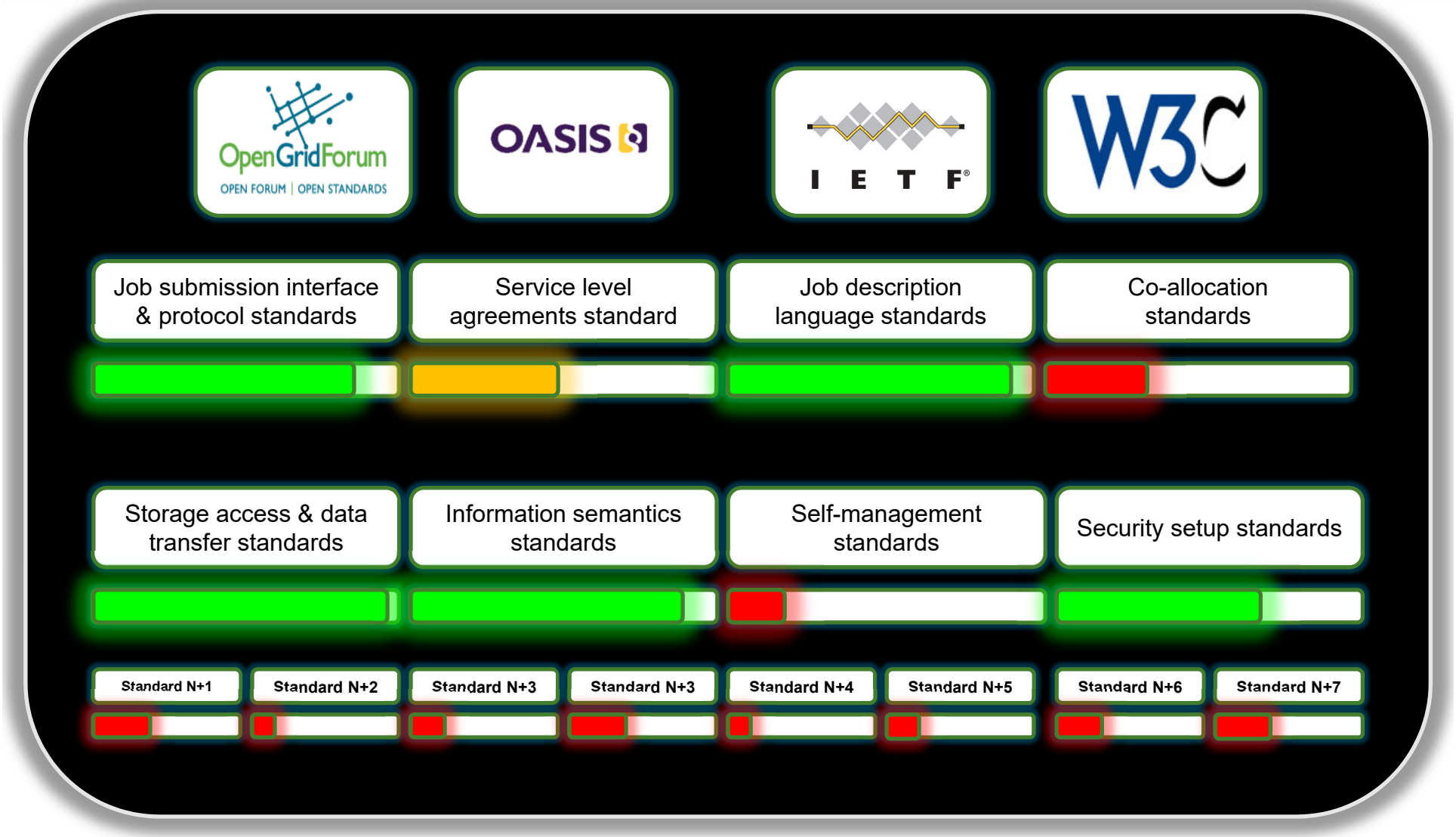
Open Standards Approach



Open Standards Approach

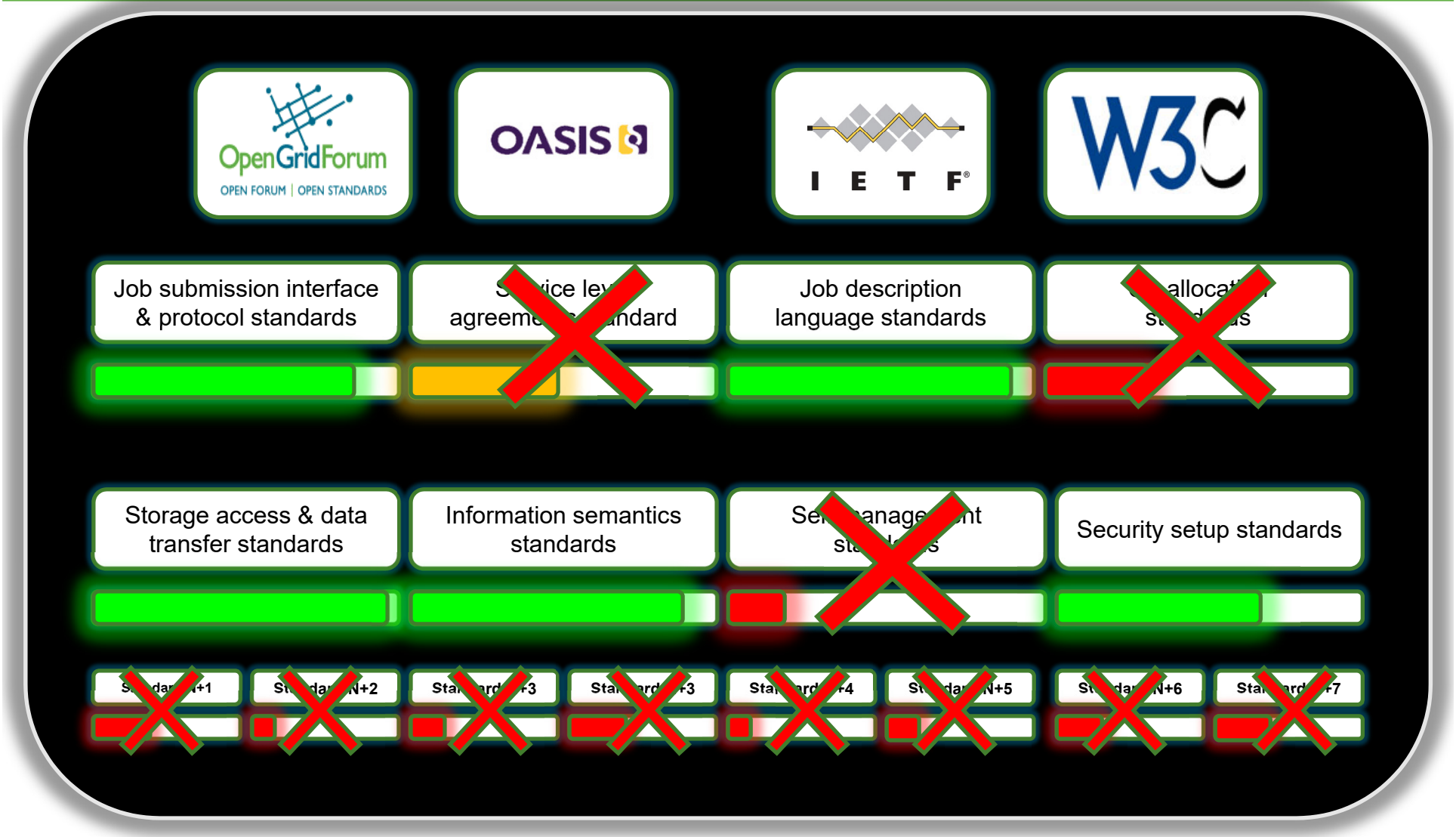
- No transformation logic required
- Requires substantial effort to reach an agreement between middlewares that adopt them
- Should not only be based on (rather theoretical) use cases
- Instead they should also take lessons learned from real production usage into account

OGSA Standards & Adoption



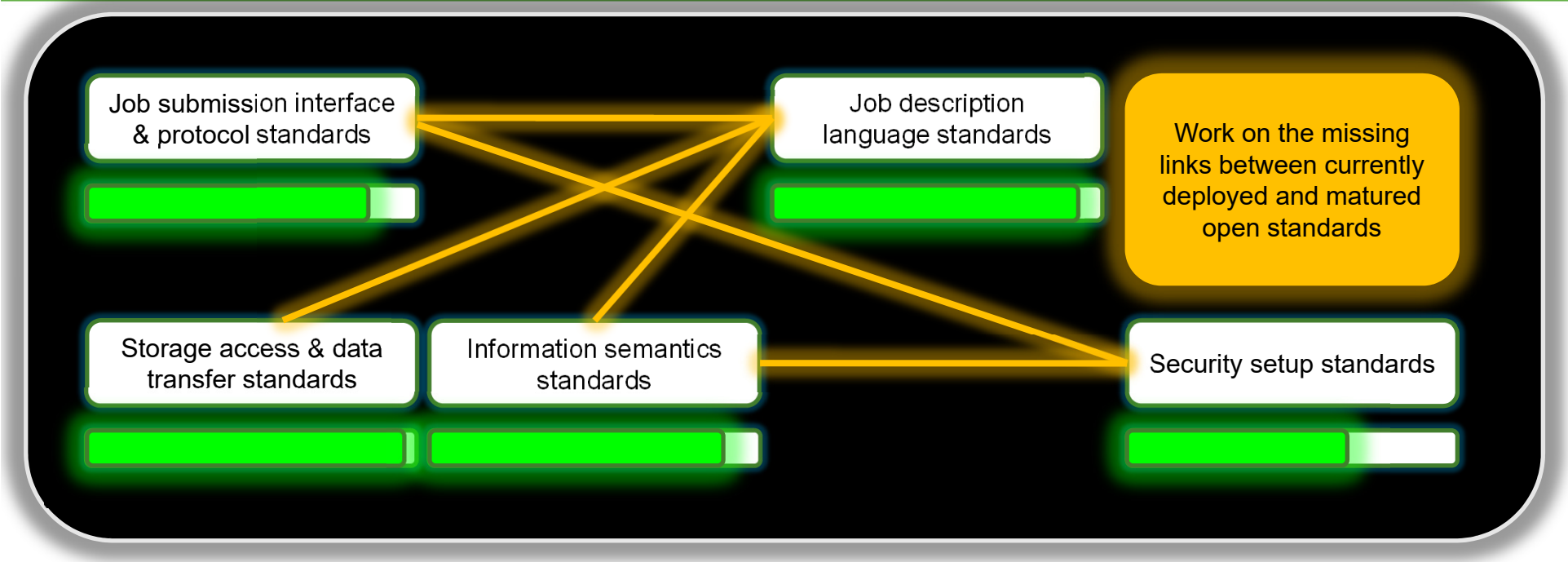
[8] Foster et al., *'The Open Grid Services Architecture'*

GIN Production Experience



[9] Riedel et al., *'Interoperation of World-Wide Production e-Science Infrastructures '*

PGI Approach (1)

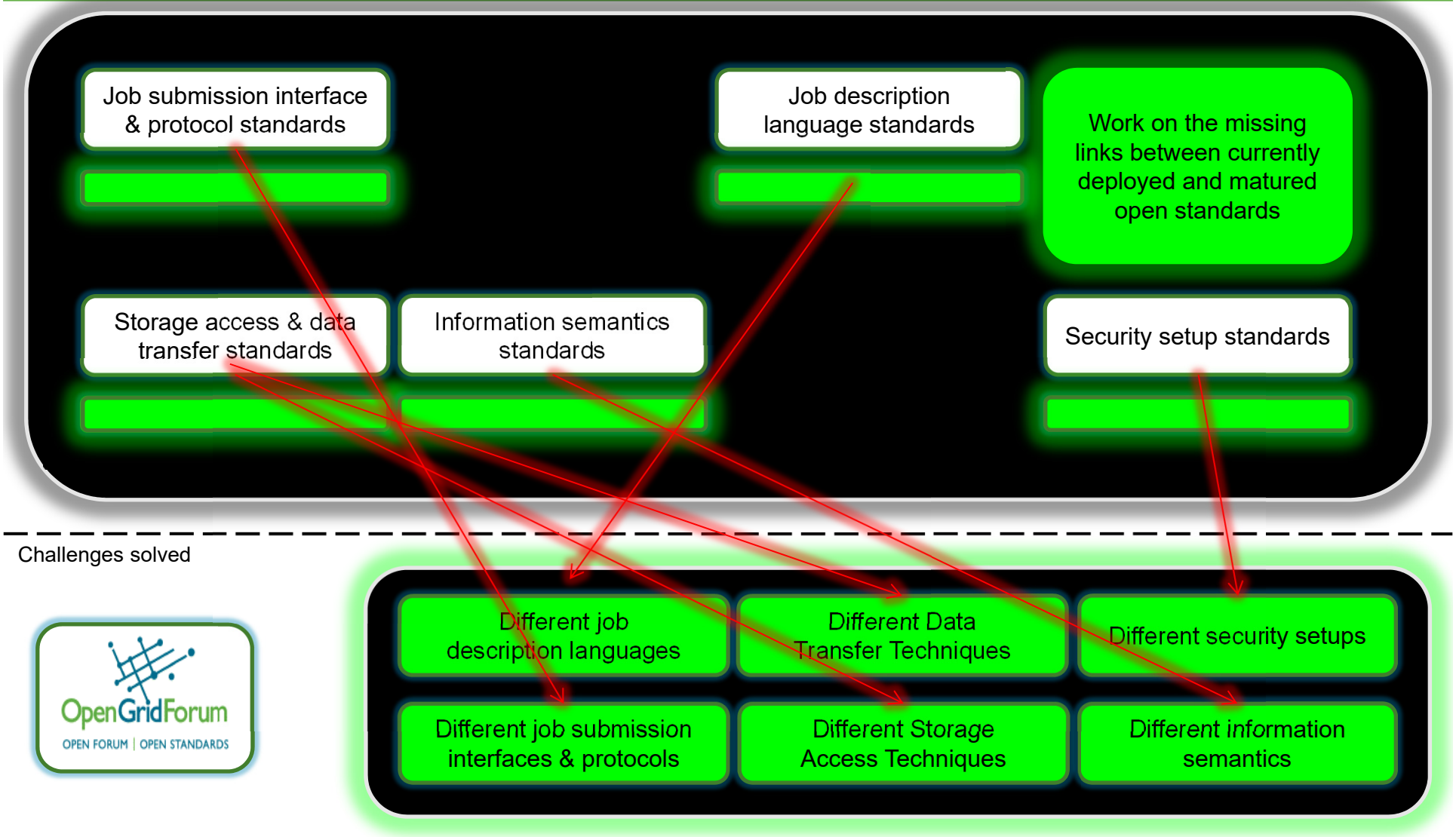


Challenges

- Different job description languages
- Different Data Transfer Techniques
- Different security setups
- Different job submission interfaces & protocols
- Different Storage Access Techniques
- Different information semantics

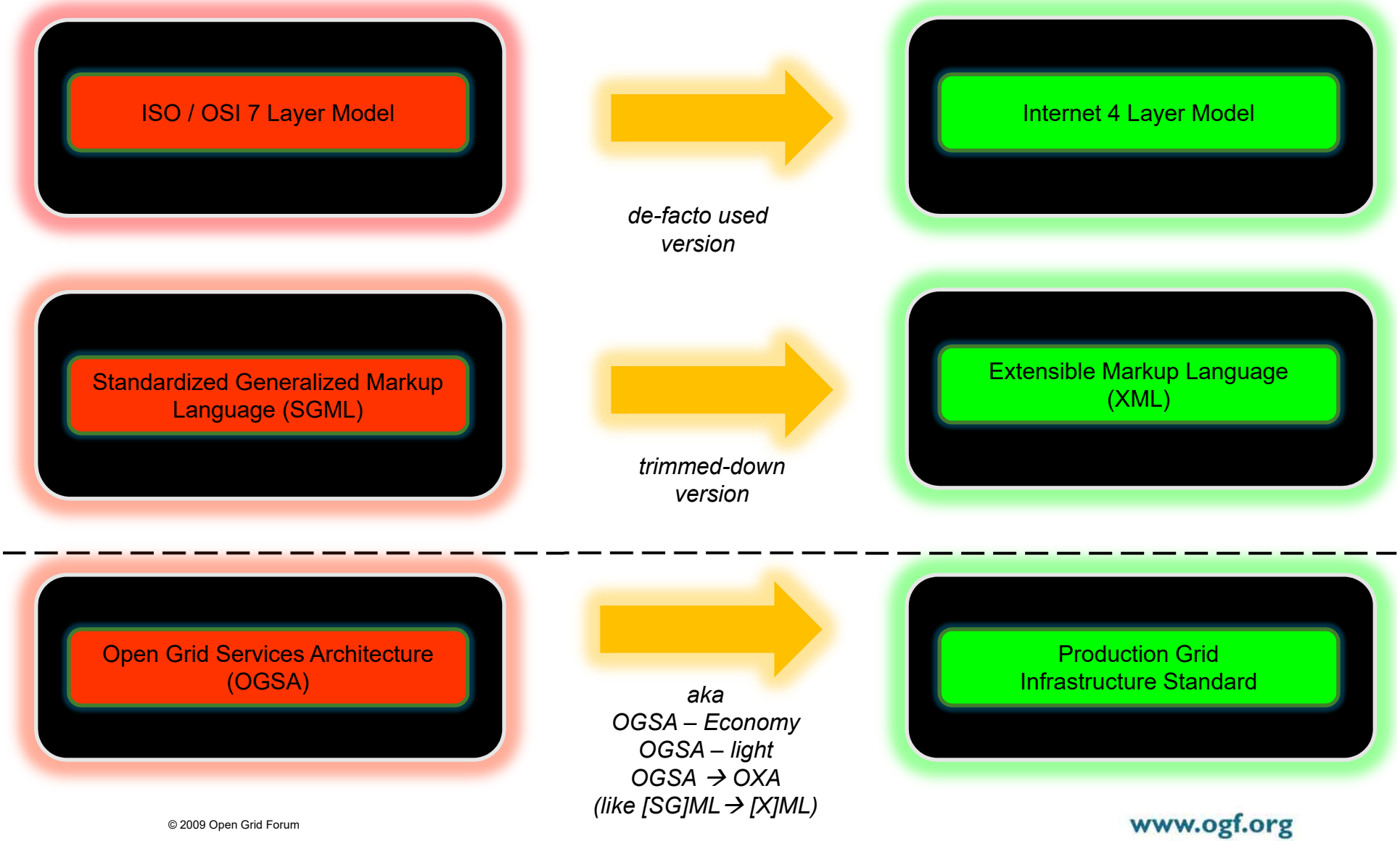
[5] Riedel et al., *'Experiences and Requirements for Interoperability between HTC- and HPC-driven e-Science Infrastructures'*

PGI Approach (2)

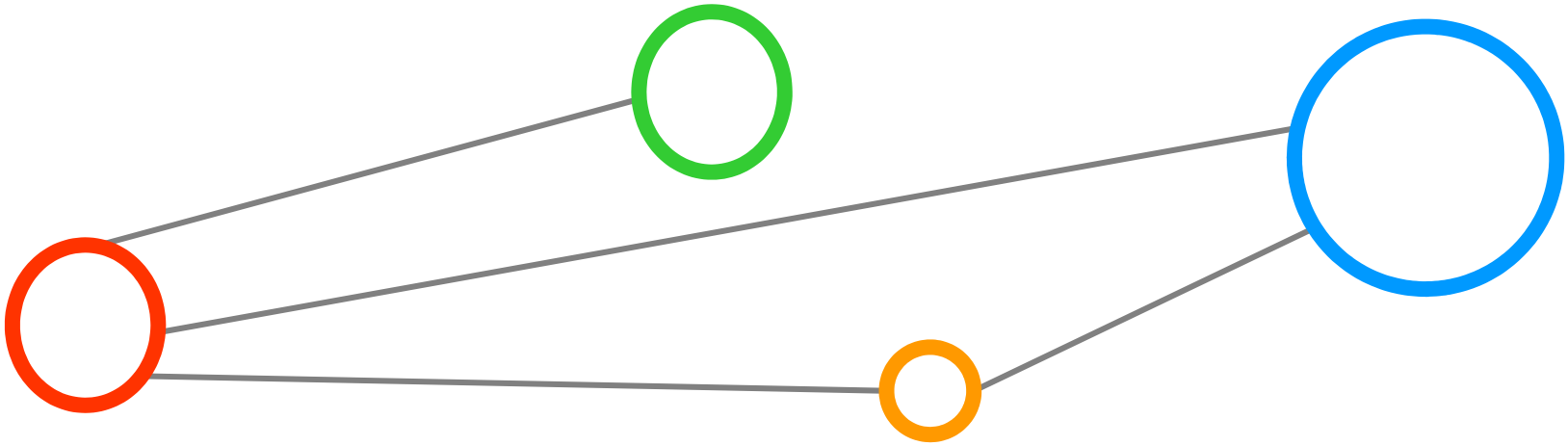


[5] Riedel et al., *'Experiences and Requirements for Interoperability between HTC- and HPC-driven e-Science Infrastructures'*

Compare History of Computer Science



Interoperability Reference Model



Often Used Functional Interfaces

GIN Interoperation demonstrations from numerous world-wide projects

Work with emerging open standards on real production Grid applications

International Grid Interoperability & Interoperation Workshops 2007, 2008 & Grid Computing Journal Special Issue Interoperability 2009

GridFTP
OGF Specification GFD

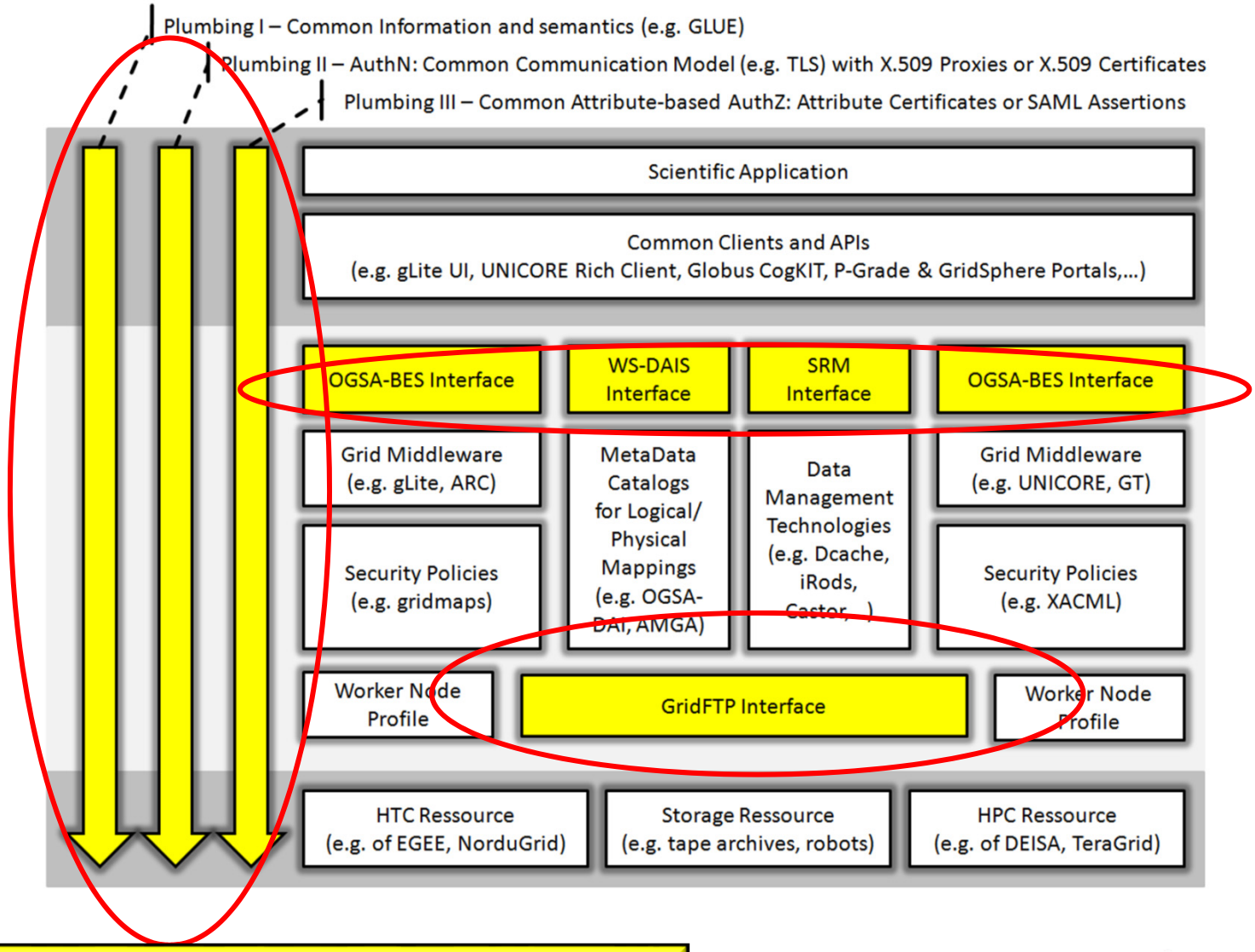
Storage Ressource Manager (SRM)
OGF Specification GFD

OGSA – Basic Execution Service (BES)
OGF Specification GFD

Job Submission & Description Language (JSDL)
OGF Specification GFD

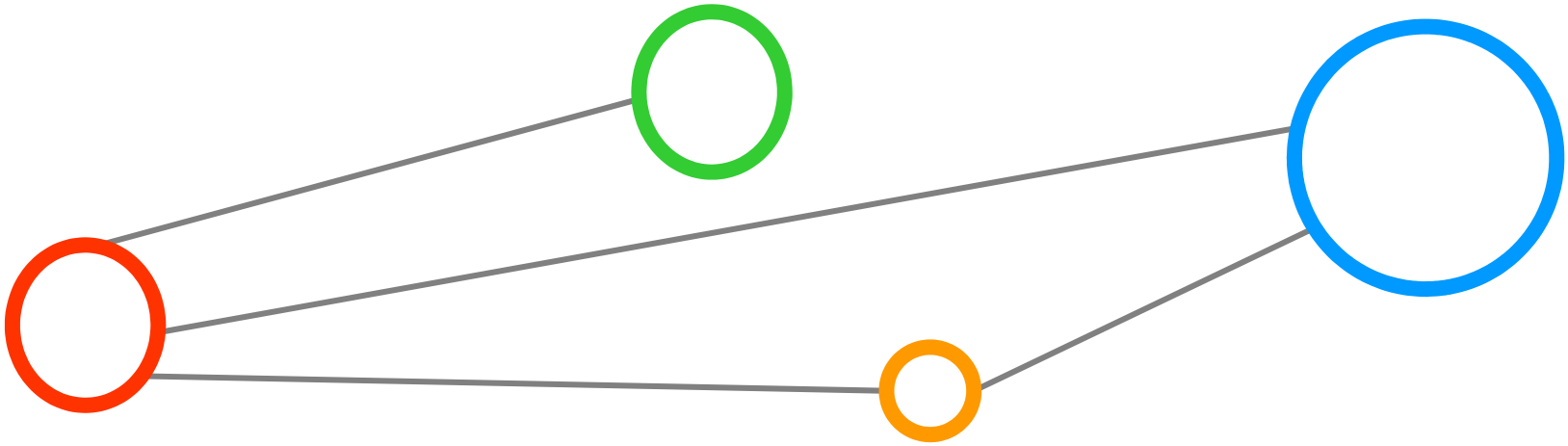
WS-Data Access&Integration Service (DAIS)
OGF Specification GFD

Reference Model Overview



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

Computing Refinement Concepts



Emerging Standards in Context

- OGSA – Basic Execution Service (BES)
  - OGF Specification GFD108, out since 2007-08-07
  - Provides a functional interface to manage computational jobs
  - Implies the use of JSDL as jobs description language
  - Defines a job state model that is simple – but extensible
  - Since 2007 in use in many different use cases and some middleware
- Job Submission and Description Language (JSDL)
  - OGF Specification GFD56, out since 2005 / 2006
  - Some standardized extensions since then: Single Process Multiple Data (SPMD) – 2007, HPC-Profile – 2007, Parameter Sweep – 2009
  - Since 2005 in use in many different use cases and many middleware
- OGSA-BES and JSDL already a good starting point
  - No need to start from scratch and a good base for refinements
  - Lessons learned: Over the years many additional required concepts have been identified mostly driven by the needs of e-scientists

Refinement Concepts Overview

Concepts	OGSA-BES / JSDL	Improvements
Simple job submission	Yes	Yes
Cancellation of submitted jobs	Yes	Yes
Getting submitted job states	Yes	Yes
Remote management operations	Yes	No
Client initiated data-staging	No	Yes
Immediate job working directory access	No	Yes
Predefined hold points	No	Yes
Manual manipulation of job states	No	Yes
Data-staging in state model	No	Yes
Wipe-out of submitted jobs	No	Yes
Standardized information model	No	Yes
Recent HPC resource support	No	Yes
Pre-/post processing	No	Yes
Data-transfer delegation	No	Yes
Multiple computing share support	No	Yes

Fundamental Concepts Ok

Concepts	OGSA-BES / JSDL	Improvements
Simple job submission	Yes	Yes
Cancellation of submitted jobs	Yes	Yes
Getting submitted job states	Yes	Yes

- Simple job submission
  - Refers to run one executable on a remote machine with limited resource requirements (CPUs) and automatic data-staging
  - OGSA-BES & JSDL (with extensions) supports this already via the ‚application' elements in JSDL
- Cancellation of submitted jobs
  - Refers to once submitted jobs can be cancelled
  - OGSA-BES / JSDL supports this already via *TerminateActivities()* operation and the ‚cancelled' job state
- Getting submitted job states
  - Refers to the ability to request the up-to-date state of the job
  - OGSA-BES / JSDL supports this via *GetActivityStatuses()* operation

Remote Management

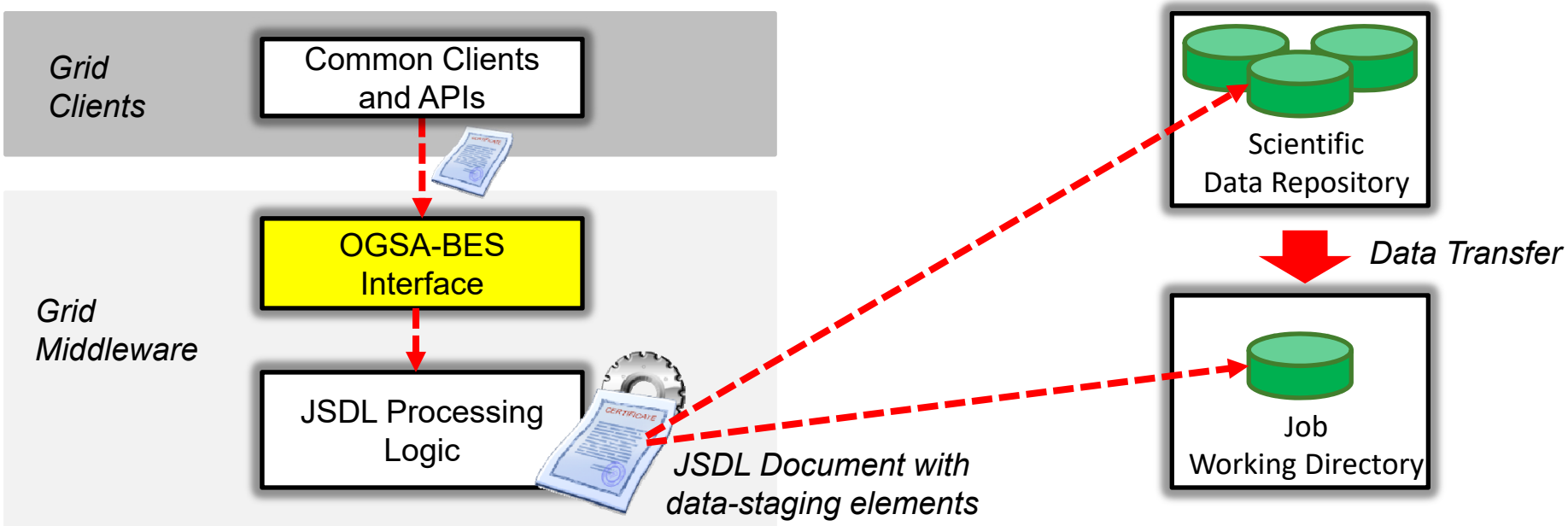
Concepts	OGSA-BES / JSDL	Improvements
Remote management operations	Yes	No

- OGSA-BES / JSDL define funtionality for remote management in terms of ‚accepting new activities'
  - OGSA-BES provides a BES-Management portType with two operations
  - StartAcceptingNewActivities() / StopAcceptingNewActivities()
  - IsAcceptingNewActivities as boolean for BES Factory attributes that describe the fundamental properties of one computing site
- Improvements (here reduction)
  - The BES-Management concept is marked as ‚deprecated'
  - Major reason is that production use reveals that this concept is rather rarely remotely used in production Grids
  - Site property is preferred configured locally by site administrators

Client initiated data-staging (1)

Concepts	OGSA-BES / JSDL	Improvements
Client initiated data-staging	No	Yes

- OGSA-BES / JSDL define functionality for staging data automatically performed via the middleware
  - Works via data-staging-in and data-staging-out JSDL elements
  - Can be considered as a kind of ‚data-pull' concept

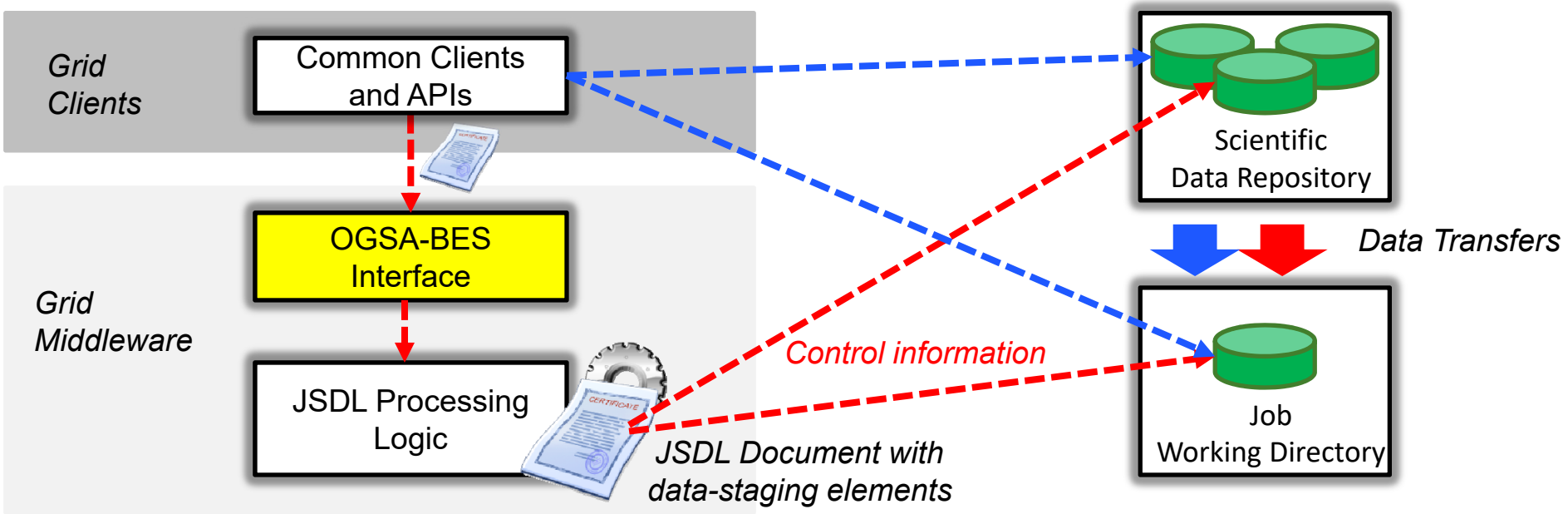


[10] M. Riedel and D. Kranzlmüller et al., 'Concepts and Design of an Interoperability Reference Model for Scientific- and Grid Computing Infrastructures

Client initiated data-staging (2)

Concepts	OGSA-BES / JSDL	Improvements
Client initiated data-staging	No	Yes

- Improved OGSA-BES / JSDL defines funtionality for staging data manually performed via the client
  - Identified via data-staging-in and data-staging-out JSDL elements
  - Can be considered as a kind of ‚data-push' concept
  - Requires other concepts ‚holdpoints' & ‚Working Directory Access'



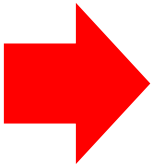
[10] M. Riedel and D. Kranzlmüller et al., 'Concepts and Design of an Interoperability Reference Model for Scientific- and Grid Computing Infrastructures

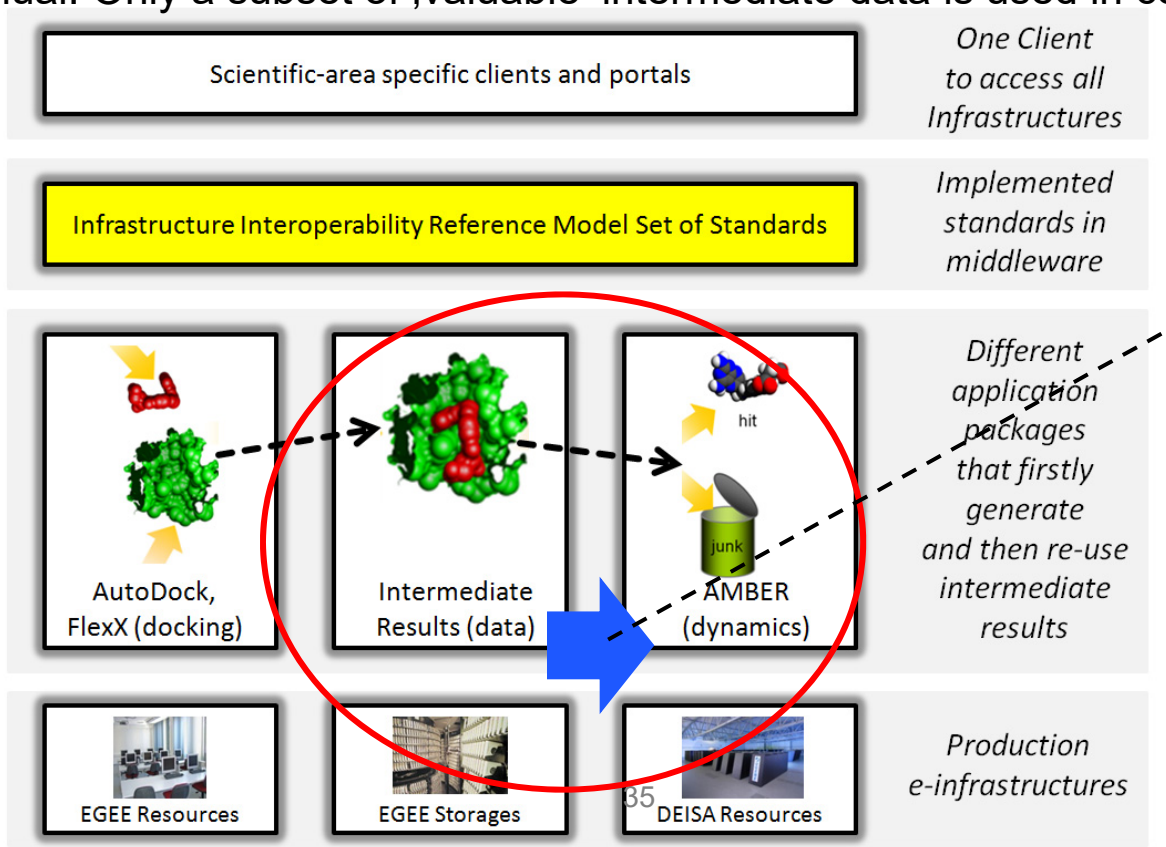
Client initiated data-staging (3)

Concepts	OGSA-BES / JSDL	Improvements
Client initiated data-staging	No	Yes

- Example of this requirements from an e-science perspective
  - Manual: Only a subset of ‚valuable' intermediate data is used in costly HPC computing

‚data-pull': automatic data transfers for ligands, protein structures, etc.





‚data-push': manual data transfers initiated on the client level with strong participation of the e-scientists during/after his/her analysis and evaluation

[10] M. Riedel and D. Kranzlmüller et al., 'Concepts and Design of an Interoperability Reference Model for Scientific- and Grid Computing Infrastructures

Client initiated data-staging (4)

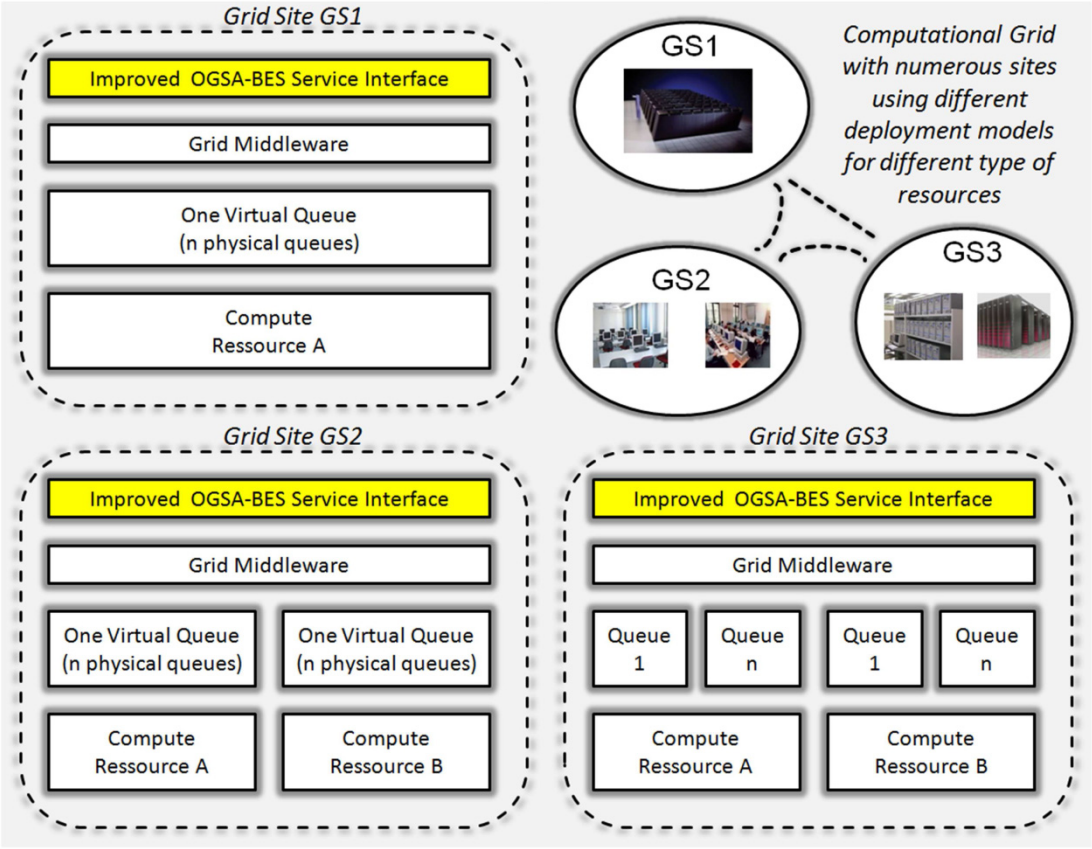
Concepts	OGSA-BES / JSDL	Improvements
Client initiated data-staging	No	Yes
Immediate job working directory access	No	Yes
Predefined hold points	No	Yes
Manual manipulation of job states	No	Yes

- ‚Client initiated data-staging‘ concept requires other concepts
- ‚Immediate job working directory access‘ concept
  - Once job is created the improved OGSA-BES returns the job working directory in order to know where to manually ‚stage-data in&out‘
- ‚Predefined hold points‘ concept
  - Hold points in improved JSDL enables stop of job processing
  - Provides e-scientists with all the time they need to stage-in manually
  - Cp. ‚breakpoints‘, but ‚holdpoints‘ have no direct executable impact
- ‚Manual manipulation of job states‘ concept
  - In order to resume the ‚holded processing‘ a manualy manipulation of states (i.e. continute in hold) is provided via the improved OGSA-BES

Multiple Share Concept

Concepts	OGSA-BES / JSDL	Improvements
Multiple computing share support	No	Yes

More fine-granular URIs are required to specify exactly which ‚computational share' / site:
https://jump.fz-juelich.de:8080/besservice/FZJ/JUMP/c bench
https://jugene.fz-juelich.de:8080/besservice/FZJ/JUGENE/res vph



Other concepts (1)

Concepts	OGSA-BES / JSDL	Improvements
Data-staging in state model	No	Yes
Wipe-out of submitted jobs	No	Yes
Standardized information model	No	Yes

- ‚Data-staging' in state model concept
  - Users have to know all the time what the system does
- ‚Wipe-out of submitted jobs' concept
  - Instead of ‚only cancelled' some jobs should be not tracked by the system anymore
- Standardized information model concept
  - Use of GLUE2 for resource requests in improved JSDL

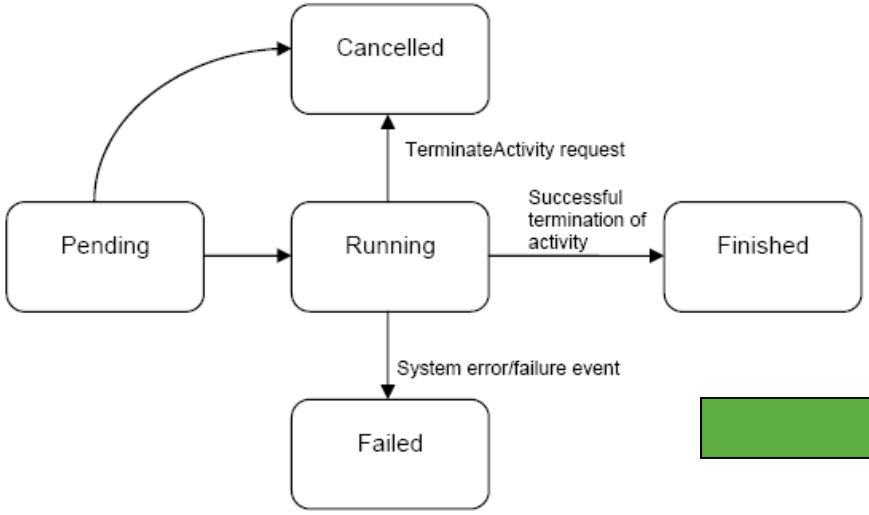
Other concepts (2)

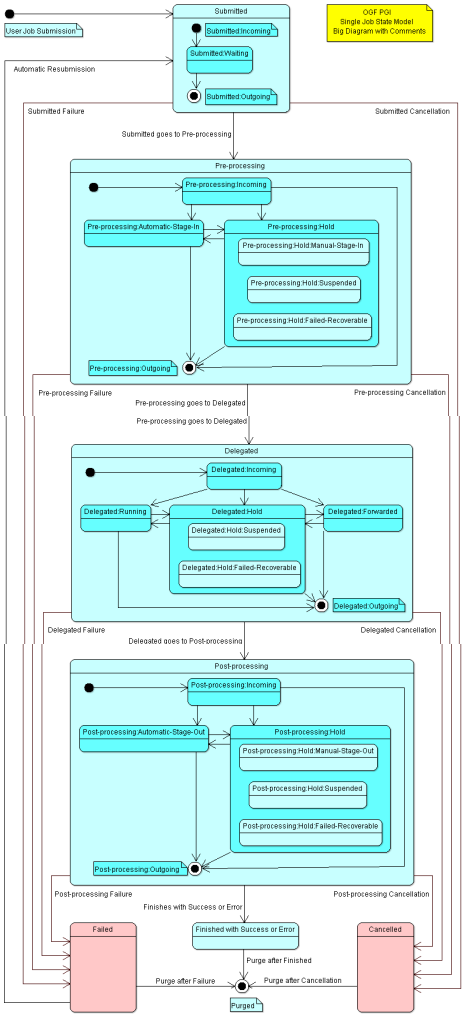
Concepts	OGSA-BES / JSDL	Improvements
Recent HPC resource support	No	Yes
Pre-/post processing	No	Yes
Data-transfer delegation	No	Yes

- ‚Recent HPC resource support' concept
  - Describe state-of-the art HPC resources with Improved JSDL
  - Covers multi-threading, network connectivity (e.g. torus), libraries,…
- ‚Pre-/post processing' concept
  - e-Scientists often require small program (executed non-parallel) before the (parallel) executable starts to run (or after)
- Data-transfer delegation
  - Third-party credentials – how to transfer n different credentials (with different attributes) to a service that performs data-staging on behalf of myself
  - Improved OGSA-BES with portType to create a delegated credential in a two phase operation protocol, enables use of different credentials in data-stagings

Improved State Model

- Improve basic state model of OGSA-BES specification
  - Much more fine granular and refined states
  - More feedback to users (i.e. data-stagings)



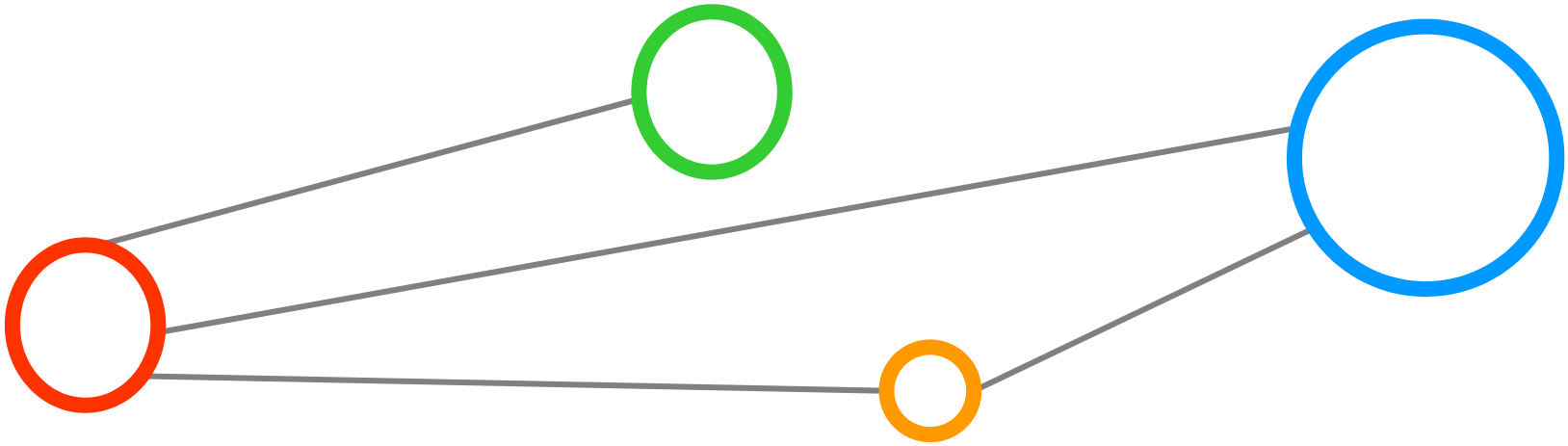


Other Refinement Concepts

- Job Description
  - Concepts to improve JSDL towards production use
- Execution Environment
  - Definition of common execution environments (e.g. environment variables, pre-installed software&libraries)
- Security
  - Common attribute-based authorization based on SAML assertions
  - Common authentication & move away from rather proprietary Grid Security Infrastructure (GSI)
  - Improved delegation model with delegation restrictions
- Data Management and Transfer elements
  - Refinement concepts around WS-DAIS (relational DB access)
  - Concepts to profile Storage Ressource Manager (SRM) interfaces
  - Closer alignment with computation / information / storage

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

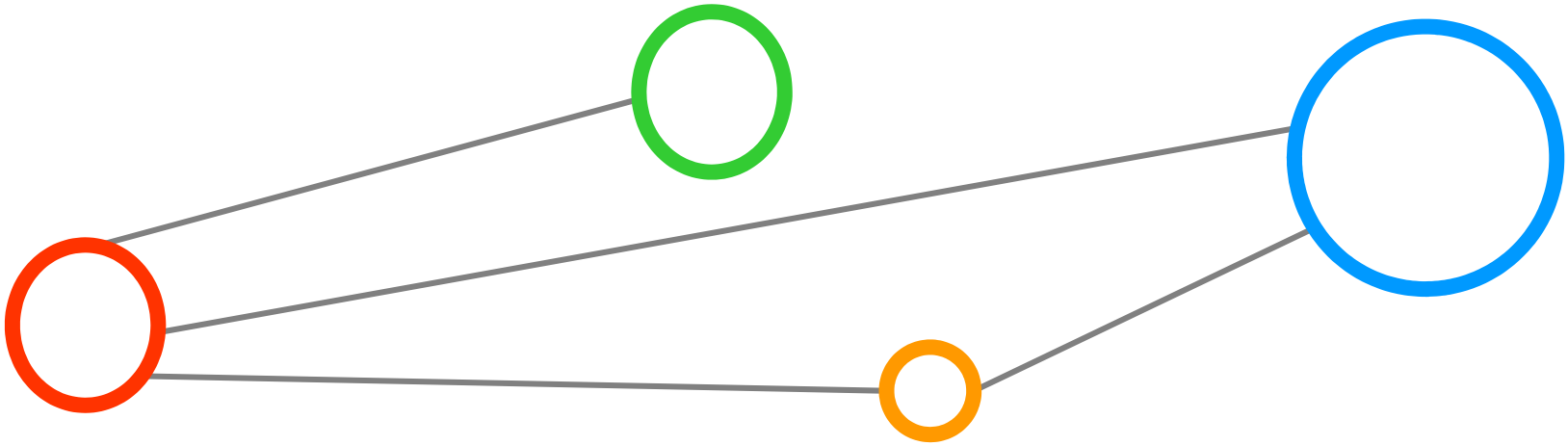
Conclusions



Conclusions

- More and more e-science projects require resources in more than one Grid → Grid interoperability problem
  - Many approaches exist – only production-aware standards help
  - Production Grid Infrastructure (PGI) standardization process
- OGSA exists, but…
  - Hard to maintain, nearly half of all specs defined, missing links,…
- Comparison with history of computer science
  - Cp. XML & SGML, Internet model vs. ISO / OSI model
  - Bottom-up (from production) instead of top-down architecture
- Reference model obtained from real scientific use cases
- Interoperability reference model (or aka profiles) make sense
  - Scientific use cases proof feasibility of initial reference model
  - Might be a milestone towards full OGSA-conformance roadmaps

References



References (1)

[1] Lippert et al., 'IBM delivers Europe's biggest supercomputer',
Online: http://news.zdnet.co.uk/hardware/0,1000000091,39146680,00.htm

[2] John Taylor, 'The Definition of e-Science',
Online: http://www.lesc.ic.ac.uk/admin/escience.html

[3] M. Riedel, A. Streit, F. Wolf, Th. Lippert, D. Kranzlmüller, *Classification of Different Approaches for e-Science Applications in Next Generation Computing Infrastructures* Proceedings of the 4th IEEE Conference on e-Science (e-Science) 2008, Indianapolis, Indiana, USA

[4] M. Riedel, A.S. Memon, M.S. Memon, D. Mallmann, A. Streit, F.Wolf, Th. Lippert, V. Venturi, P. Andreetto, M. Marzolla, A. Ferraro, A. Ghiselli, F. Hedman, Zeeshan A. Shah, J. Salzemann, A. Da Costa, V. Breton, V. Kasam, M. Hofmann-Apitius, D. Snelling, S. van de Berghe, V. Li, S. Brewer, A. Dunlop, N. De Silva, *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

[5] M.Riedel, A.Streit, D.Mallmann, F. Wolf, Th. Lippert, *Experiences and Requirements for Interoperability between HTC- and HPC-driven e-Science Infrastructures, Proceedings of the Korean e-Science All Hands Meeting, Daejeon, Korea, 2008*

References (2)

[6] 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
Accepted for publication in Journal of Cluster Computing,
Special Issue Recent Advances in e-Science

[7] S. Manos & M. Riedel, DEISA Newsletter Contributions, December 2008

[8] I. Foster et al. 'The Open Grid Services Architecture', OGF Grid Final Document #80, Online: http://www.ogf.org/documents/GFD.80.pdf

[9] M.Riedel 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

[10] M.Riedel and D. Kranzlmüller et al., Concepts and Design of an Interoperability Reference Model for Scientific- and Grid Computing Infrastructures, accepted for proceedings of the Applied Computing Conference (ACC) 2009, Athens