



Six Years of MoSGrid

Jens Krüger

High Performance and Cloud Computing Group
Zentrum für Datenverarbeitung
Eberhard Karls Universität Tübingen

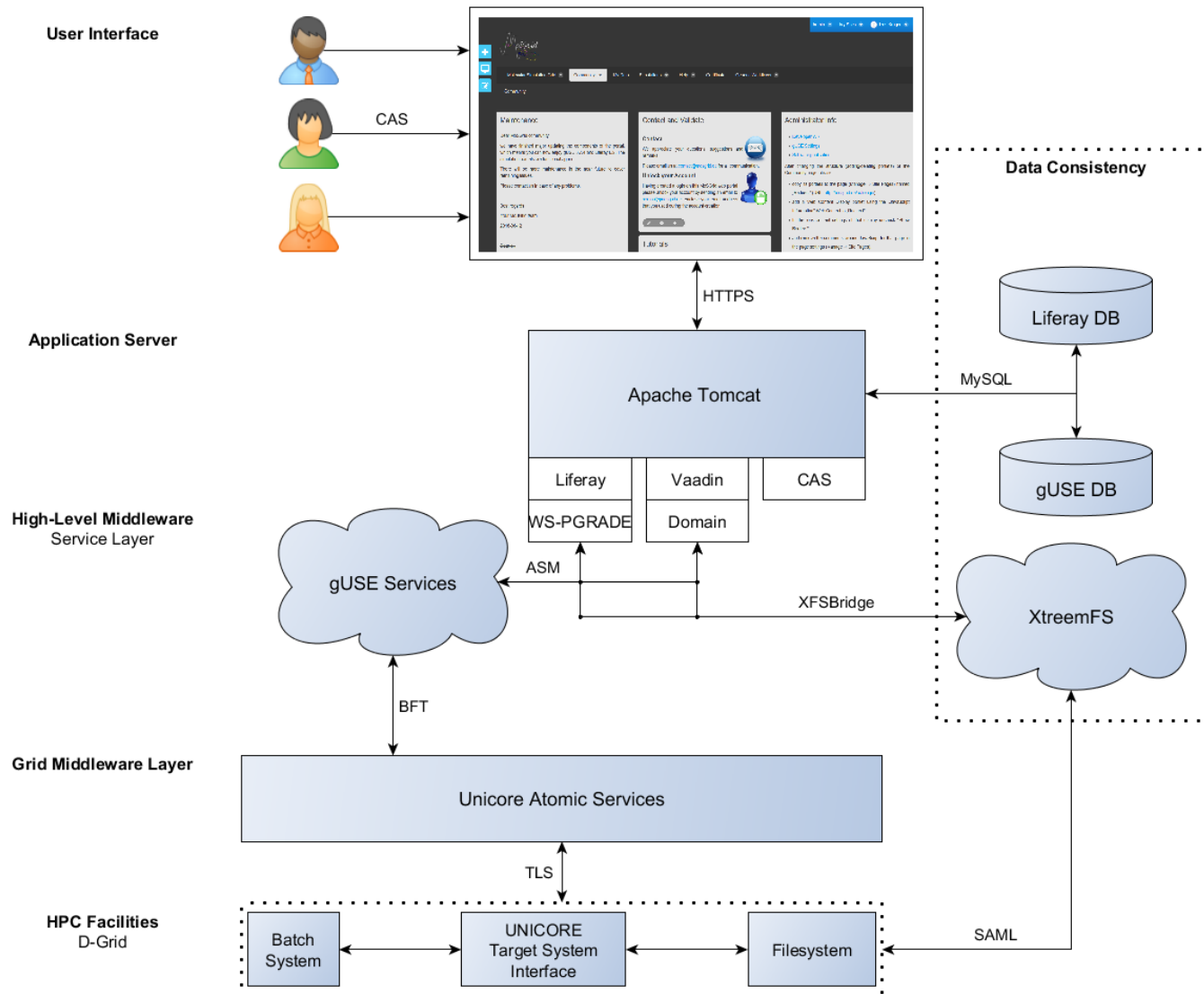
04.04.17
Granada

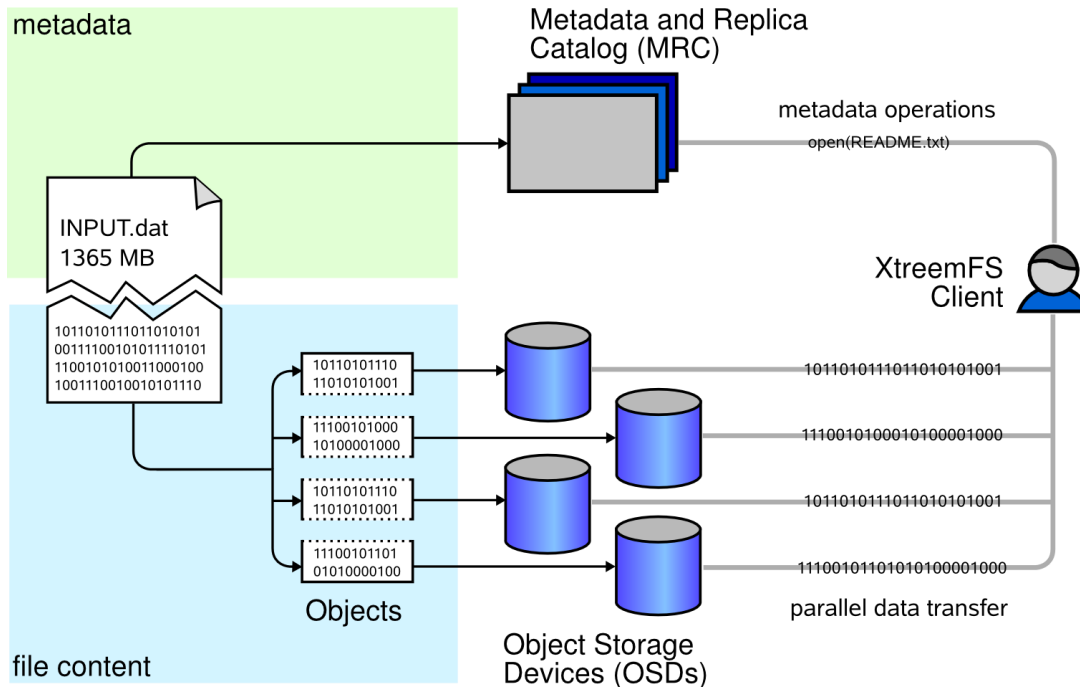


- Motivation and Background
- Infrastructure
 - Storage
 - UNICORE
 - Authentication and Security
 - Portal Instance
- Portal
 - Liferay
 - gUSE/WS-PGRADE
 - Portlet API
 - Simulation Portlets and Workflows
- Stability and Resilience
- Lessons Learned
- Future Work

- MoSGrid was initially a BMBF funded project (since 2009 for 3.5 years)
- Intended for computational chemists and scientists from related fields
- Partners from academia and industry
- The portal is online since 2010
- FP7-EU projects ER-flow and SCI-BUS followed (till end of 2014)
- Currently 493 users and more than 120 different workflows







XTreamFS

Distributed object based file system

DIR directory service (1x)

MRC stores metadata (N x)

OSDs store pure file (N+ x) content

Client provides file system interface



- UNICORE is a middleware facilitating the connection between portal and compute resources
- It offers a set of high-level services for job and workflow handling as well as for data movement
- For MoSGrid UNICORE offers a uniform layer easing operation and maintenance

Challenges

- Renewal of service certificates
- Different UNICORE versions on different HPC resources
- Curation of user list and user credentials
- Synchronization of information over all instances and sites



- MoSGrid relies on personnel user certificates
- The user delegates the right to submit jobs and workflows to the portal instance using a SAML assertion
- Same applies for the secure connection to the storage containing simulation data and metadata information

Challenges

- Yearly renewal of user certificates
- Maintenance of the Certificate Portlet which allows the creation of SAML assertions (Java Applet → plain Vaadin portlet)
- Synchronization between all instances



- Running as VM on a shared host
- Scientific Linux 6.7
- Java Virtual Machine 1.7
- MySQL 5.1.73 for Liferay and gUSE databases
- gUSE 3.7.4 including Liferay 6.2 GA2 and Apache Tomcat 7.0.55

- No foreign software repositories are used
- No automated updates

Challenges

- Changes to the basic infrastructure like network or VM host
- Possibility to continuously upgrade to recent versions without breaking compatibility



Admin | My Sites | Jens Krüger

Molecular Simulation Grid | Community | My Data | Simulations | Help | Certificate | Generic Workflows

Maintenance

Dear MoSGrid community,

we have finished major updating the components of the portal, which means you can now enjoy gUSE 3.7.4 and Liferay 6.2. The simulation portlets are functional again.

There will be more maintenance in the near future to cover remaining issues.

Please contact us in case of any problems.

Best regards
Your MoSGrid team
2016-06-12

Dear-all,
we will update some components of the portal.

Contact and Validate

Contact

We appreciate your questions, suggestions and remarks!

Please email us at contact@mosgrid.de for all communication.

Unlock your Account

Having created a login on this MoSGrid web portal please unlock your account by sending an email to contact@mosgrid.de. Enclose your email address that you used during the account creation.

Welcome and Further Necessa...

Dear User,

welcome to MoSGrid. In order to actively carry out simulations the following steps have to be sequentially accomplished (Clicking the icons will lead you through the respective processes):

- 1.
- 2.
- 3.

Unfortunately, due to policy reasons related to the underlying German HPC systems a user needs to be connected to a German research institute in order to get a German Grid Certificate and thus be able to run simulations.

To give you the necessary insight into this procedure and provide you with additional information, a detailed tutorial is provided: [Grid Certificates and Security Assertions](#).

We appreciate every input, no matter if it's technical,

Challenges

- Incompatibilities between different versions of Liferay (6.x→7.x), gUSE/WS-PGRADE, Vaadin, ...

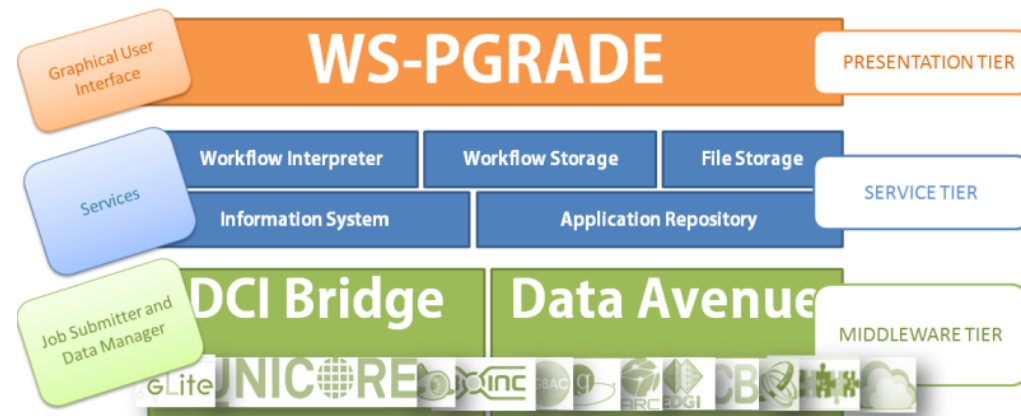


Generic-purpose gateway framework

- Based on Liferay
- Workflow-oriented gateway framework
- Supports the development and execution of workflow-based applications
- Supports the fast development of domain-specific gateways by a customization technology
- Most important design aspects are flexibility and robustness

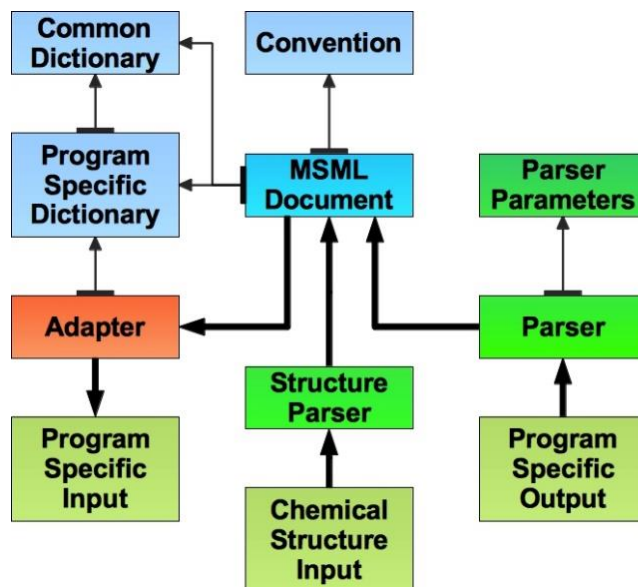
Challenges

- Academic project
- Rather steep learning curve





- Molecular Simulation Markup Language (MSML)
- CML compliant
- Template for each and every workflow
- Molecular input
- Domain specific tools
- Job configuration
- Optimized structures, trajectories, energies, ...
- Semantic search



```

<?xml version="1.0" encoding="UTF-8"?>
<cml convention="convention:compchem">
  <module dictRef="compchem:jobList">
    <cmlx:parserConfiguration/>
    <module dictRef="compchem:job" id="Job1">
      <module dictRef="compchem:environment">
        <propertyList/>
      </module>
      <module dictRef="compchem:initialization">
        <parameterList/>
        <cmlx:adapterConfiguration/>
        <cmlx:parserConfiguration/>
      </module>
      <module dictRef="compchem:finalization">
        <propertyList/>
      </module>
    </module>
    <module dictRef="compchem:job" id="Job2"/>
    ...
  </module>
</cml>
  
```



```
<module dictRef="compchem:initialization">
  <parameterList>
    <parameter dictRef="g09:loglevel">
      <scalar dataType="xsd:string" units="si:none">p</scalar>
    </parameter>
    <parameter dictRef="g09:jobtype">
      <scalar dataType="xsd:string" units="si:none">opt</scalar>
    </parameter>
    <parameter dictRef="g09:hf.theory" cmlx:editable="true">
      <scalar dataType="xsd:string" units="si:none">hf</scalar>
    </parameter>
    <parameter dictRef="g09:basisset" cmlx:editable="true">
      <scalar dataType="xsd:string" units="si:none">6-31G</scalar>
    </parameter>
    <parameter dictRef="g09:formal.charge" cmlx:editable="true">
      <scalar dataType="xsd:integer" units="si:none">0</scalar>
    </parameter>
    <parameter dictRef="g09:spin" cmlx:editable="true">
      <scalar dataType="xsd:integer" units="si:none">1</scalar>
    </parameter>
    <parameter dictRef="g09:checkpointfile">
      <scalar dataType="xsd:string" units="si:none">job.chk</scalar>
    </parameter>
  </parameterList>
  <cmlx:adapterConfiguration adapterID="g09adap"
    fileExtension="com" portName="job.com"/>
</module>
```



Quantum Chemistry Portlet

Import
Submission
Monitoring

Welcome

Welcome to the Quantum Chemistry portlet.

Import a workflow

Toolsuite

Workflow *

- Submission of prepared job-files. (Yields formatted checkpoint file)
- Optimization with DFT methods
- Optimization with HF methods
- Optimization + frequency calculations with DFT methods
- Submission of prepared job-files. (No postprocessing)
- Generate cube files for the visualization of HOMO and LUMO
- Submission of prepared job-files. (No postprocessing)



Molecular Simulation Grid Community Simulations Generic Workflows My Data Certificate Help Coming Soon

Molecular Simulation Grid > Simulations > Docking

MD Portlet

Import Submission

Welcome

This is the M

Import a workfl

Toolsuite

Gromacs 4.

Workflow *

Energy Mini

Name *

EM_2014-0

Import

Docking Portlet

Import Submission Mo

results.sdf

vinareresult.pdbqt

BALL 1.4.0 (Jan 4 201

23	24	0999	0	0	0	0	0
3.7570	64.2200	64.16					
3.7340	65.4120	63.42					
4.4150	65.5210	62.20					
4.3740	66.6930	61.50					
3.6550	67.7840	61.99					
2.9790	67.7120	63.17					
3.0010	66.5200	63.91					
2.3580	65.3210	65.78					
3.0580	64.1850	65.36					
4.9720	64.6820	61.81					
4.9010	66.7770	60.57					
3.6340	68.7020	61.42					
2.4260	68.5660	63.53					
1.8130	65.2850	66.71					
2.3040	66.4860	65.10					
1.7770	67.2960	65.45					
4.5210	63.0150	63.68					
3.8840	62.1330	63.74					
4.8240	63.1720	62.64					
5.4060	62.8700	64.30					
3.0520	63.0240	66.13					
3.7820	62.3890	66.06					

protein.pdb

Protein

Use Cartoon Style

Show Backbone

Hide Ribbons

Atoms

Representation Mode

Ball and Stick

Use Stars

Show Water

Hide Non Polymers

Coloring

Background

Black

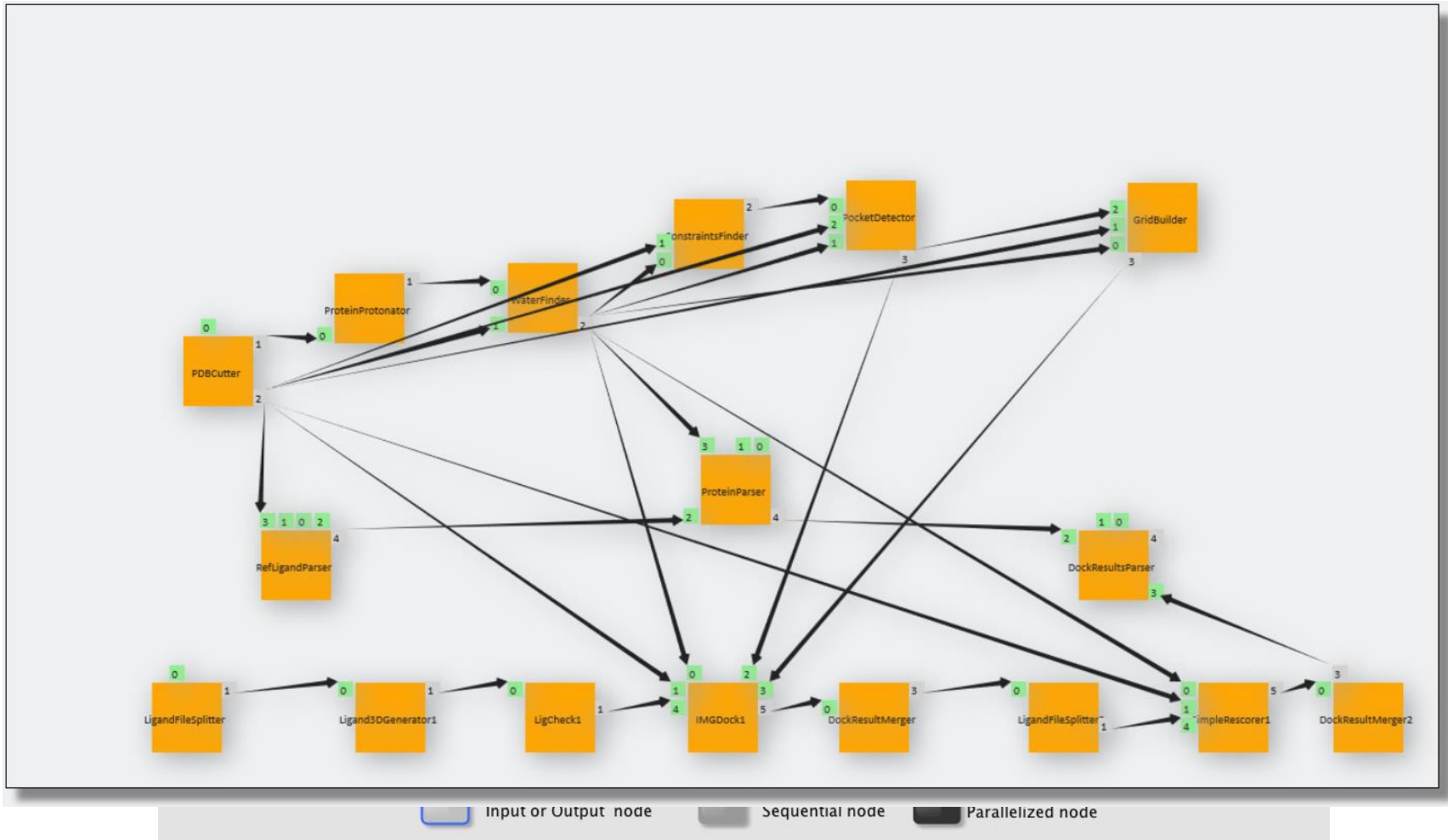
Color By Chain

Color By Residue

Size

Canvas Size

(500x500)





- MoSGrid relies on the support of various university compute centers across Germany
- Multiple loosely coupled layers

Challenges

- Communication about maintenances
- Correct detection of errors and problems
- Changing policies
- Decommission of resources
- Availability of dedicated personnel in an academic environment



95 % of all users are smart and hard working people
Same applies for administrators, developers, ...

... but everybody may have a weak day.

Be prepared for every imaginable stupidity.



Challenges

- MoSGrid relied on the availability of D-GRID services
- Migration of a whole science gateway
- Consistency over multiple levels and multiple sites

Solution

- Constant communication with the compute centers
- Careful planning and plenty of manpower
- Careful planning and good testing procedures

Take home message

- Choose an actively maintained science gateway framework
- Enrich with features which truly add an extra value
- Keep the number of used technologies, interdependencies and overall complexity as low as possible
- Ensure that direct access to remote compute and data instances is possible
- Close communication with the user community



- Migration of key services to a RHEV environment
- Containerization of applications using Docker or Singularity
- Support of mobile devices
- Virtualization possibilities such as microservices enabling scientists to move compute resources to their research data and not vice versa.



- A Science Gateway can ease the access to computational and storage resources for a specific community
- MoSGrid is successfully operated for more than six years
- Several hundred users have used MoSGrid successfully
- Regularly used in teaching higher education classes
- Experience gathered for the computational chemistry community could provide useful insights for other communities
- A lean, robust and open-source technology stack is advised when making design decisions for future science gateways.



Core team

- Sandra Gesing, Sonja Herres-Pawlis and Richard Grunzke

Contributors

- Lars Packschies, Andreas Zink, Lukas Zimmermann, Patrick Schäfer, Zoltan Farkas, Peter Kaczuk, Alexander Hoffmann, Sebastian Breuers, Gregor Fels, Georg Birkenheuer, André Brinkmann, Luis de la Garza, Oliver Kohlbacher, Martin Kruse, Wolfgang Nagel, Ralf Müller Pfefferkorn, Thomas Steinke, Tobias Schlemmer, Klaus Dieter Warzecha, ...

Compute Centers

- ZIH Dresden, ZDV Tübingen, PC² Paderborn, RRZK Cologne, ZIB Berlin, ...

Funding

- MoSGrid (BMBF), SCI-BUS (EU-FP7), ER-flow (EU-FP7) and MASi (DFG)

<https://scholarspace.manoa.hawaii.edu/handle/10125/41918>