

## A Science Gateway Getting Ready for Serving the International Molecular Simulation Community

---

### **Sandra Gesing**

*Center for Bioinformatics & Department of Computer Science, University of Tübingen  
Sand 14, 72076 Tübingen, Germany  
E-mail: [sandra.gesing@uni-tuebingen.de](mailto:sandra.gesing@uni-tuebingen.de)*

### **Sonja Herres-Pawlis<sup>1</sup>**

*Department of Chemistry, Ludwig-Maximilians-University Munich  
Butenandtstr. 5-13, 81377 München, Germany  
E-mail: [sonja.herres-pawlis@cup.uni-muenchen.de](mailto:sonja.herres-pawlis@cup.uni-muenchen.de)*

### **Georg Birkenheuer**

*Paderborn Center for Parallel Computing, University of Paderborn  
Warburger Str. 100, 33089 Paderborn, Germany  
E-mail: [birke@uni-paderborn.de](mailto:birke@uni-paderborn.de)*

### **André Brinkmann**

*Johannes Gutenberg-University Mainz  
55099 Mainz, Germany  
E-mail: [brinkman@uni-mainz.de](mailto:brinkman@uni-mainz.de)*

### **Richard Grunzke**

*Center for Information Services and High Performance Computing, Technische Universität Dresden  
Zellescher Weg 12-14, Germany  
E-mail: [richard.grunzke@tu-dresden.de](mailto:richard.grunzke@tu-dresden.de)*

### **Peter Kacsuk**

*Laboratory of Parallel and Distributed Systems, MTA SZTAKI  
Kende Street 13-17, 1111 Budapest, Hungary  
E-mail: [kacsuk@sztaki.hu](mailto:kacsuk@sztaki.hu)*

### **Oliver Kohlbacher**

*Center for Bioinformatics & Department of Computer Science, University of Tübingen  
Sand 14, 72076 Tübingen, Germany  
E-mail: [oliver.kohlbacher@uni-tuebingen.de](mailto:oliver.kohlbacher@uni-tuebingen.de)*

---

<sup>1</sup> Speaker

**Miklos Kozlovsky**

*Laboratory of Parallel and Distributed Systems, MTA SZTAKI*

*Kende Street 13-17, 1111 Budapest, Hungary*

*E-mail: m.kozlovsky@sztaki.hu*

**Jens Krüger**

*Center for Bioinformatics & Department of Computer Science, University of Tübingen*

*Sand 14, 72076 Tübingen, Germany*

*E-mail: krueger@informatik.uni-tuebingen.de*

**Ralph Müller-Pfefferkorn**

*Center for Information Services and High Performance Computing, Technische Universität Dresden*

*Zellescher Weg 12-14, Germany*

*E-mail: ralph.mueller-pfefferkorn@tu-dresden.de*

**Patrick Schäfer**

*Zuse Institute Berlin*

*Takustraße 7, 14195 Berlin, Germany*

*E-mail: patrick.schaefer@zib.de*

**Thomas Steinke**

*Zuse Institute Berlin*

*Takustraße 7, 14195 Berlin, Germany*

*E-mail: steinke@zib.de*

The project MoSGrid (Molecular Simulation Grid) has been developing a web-based science gateway supporting the community with various services for quantum chemistry, molecular modeling, and docking. Users gain access to distributed computing infrastructures (DCIs) via intuitive user interfaces for sophisticated tools, specialized workflows, and distributed repositories. Currently, the MoSGrid community consists of about 120 users from a number of fields related to chemistry and bioinformatics located in Germany. However, the underlying security infrastructure is generally applicable and can be deployed in arbitrary projects.

MoSGrid intends to address the international community by participating in the EU-projects SCI-BUS (Scientific gateway Based User Support) and ER-flow (Building an European Research Community through Interoperable Workflows and Data), and collaborating with the EU-project EDGI (European Desktop Grid Initiative).

*EGI Community Forum 2012 / EMI Second Technical Conference,  
Munich, Germany  
26-30 March, 2012*

## 1. Introduction

Molecular simulations provide novel insights in the structure of molecules and their function. They have numerous applications in areas like structural biology, material science, and drug design. A wide range of computational tools has been developed to aid researchers and to process data on a large scale via DCIs. Drug design, for example, is a very time-consuming and expensive process. Researchers in this field can be aided by tools to filter lead compounds via simulations and for predicting the most promising ones. Only these are synthesized and analyzed in wet laboratories. However, the usability of the tools and of the underlying complex infrastructure is often limited and hampers the broad acceptance by researchers. The implemented methods are very complex reflecting the underlying complex theory. Especially, the lack of graphical user interfaces distracts novice users from utilizing the tools on a large scale. They have to become acquainted with the domain-related features, of course, but employing them via command lines increases the barrier for a broad acceptance. Additionally, underlying DCIs raise the complexity, especially for users who lack computer science background. Altogether, these reasons lead to the need of self-explanatory and intuitive user interfaces. An approach to offer easy and intuitive access to the tools is the use of science gateways.

The MoSGrid project [1] has been developing a web-based science gateway [2] for the molecular simulation community. The features include specialized workflows and their management for applications in quantum chemistry (e.g., for Gaussian [3]), molecular dynamics (e.g., for Gromacs [4]), and docking (e.g., for CADDSuite [5]). Users are enabled to create, change, invoke, and monitor workflows and share them via repositories. Currently, the science gateway is available on the German D-Grid infrastructure but its concepts are transferrable to other projects on the international level.

The task to support the international community is divided into two main parts; the security-related part and the domain-related part. The first one sets the stage for international users so that they have access to the services of the MoSGrid science gateway. In the second part, MoSGrid forges links to international user communities and further develops domain-related services.

The MoSGrid science gateway employs WS-PGRADE [6] that is the highly flexible user interface of the grid User Support Environment (gUSE) [7]. gUSE provides a large set of DCI services by which interoperation among classical services and desktop grids, clouds and clusters, unique web services, and user communities can be achieved in a scalable way. The integration of the grid middleware UNICORE [8] into gUSE has been developed in the MoSGrid project. Furthermore, WS-PGRADE, gUSE, and the cloud file system XtremFS [9] were extended to support SAML (Security Assertion Markup Language) [10]. The science gateway will be offered as open-source software. Thus, user communities working with UNICORE-based DCIs are able to make use of the MoSGrid services without any changes. Minor changes are necessary for Shibboleth-oriented infrastructures [11]. The basic infrastructure also allows authentication via MyProxy [12] certificates and the use of DCIs with foreign middlewares.

The collaboration with SCI-BUS [13], with the recently accepted project ER-flow, and EDGI [14] facilitates the use of various European Grid infrastructures. Furthermore, additional features and workflows will be developed together with the communities and regular international community workshops are planned.

## 2. Background

The enhancement of the support of communities to international scale via a science gateway affects security infrastructures for DCIs, community management activities, and domain-related features.

### 2.1 Security Infrastructures

The security features of a science gateway are based on the security of the underlying DCIs. To ensure confidentiality, integrity, and availability of data, various security infrastructures have been developed for DCIs. They follow the principle of single sign-on, which allows users to authenticate just once and gain access to all connected systems without the need to manually conduct further authentication processes. Trust delegation enables systems to act on behalf of a user and, thus, fosters single sign-on mechanisms.

The security of most grid middlewares (e.g., UNICORE, Globus Toolkit [15], gLite [16]) relies on X.509 [17] certificates. Via the International Grid Trust Federation (IGTF) [18], a structure has been established on common policies and guidelines for managing X.509 certificates across organizations, across countries, and across continents. The coordination between grid projects on the international level ensures unique and secure authentication of users, services, and servers. The European part of this structure is the Policy Management Authority (PMA) EUGridPMA (European Grid Policy Management Authority) [19] and each country fosters the structure via national PMAs.

Nowadays, there are two kinds of certificates to support the users: user certificates, which are each correlated with a single user and robot certificates, which are correlated to communities, to applications, or to science gateways. Robot certificates are not supported in all DCIs since their security processes rely on the distinguished names in the certificates. These are not unique for each user within a robot certificate.

SAML, a standard framework for single sign-on procedures, carries out trust delegation by using SAML assertions. These contain information about the issuer, the subject, the duration of validity, and a maximum number of systems that are involved in the authentication process. A second widespread concept of trust delegation in grid environments is formed by the GSI (Grid Security Infrastructure) [20] proxy certificate. Proxy certificates are short-lived credentials generated by a proxy server from a user certificate. The certificate has to be transferred to the server and the proxy certificate grants access to all systems the initiating user has access to. This authentication procedure lessens the security compared to SAML assertions. The user certificate is not only transferred via the network with its private key, but also stored in the proxy server. Proxy servers are typically well secured but still less secure than storing an assertion file on the users' computers only. Therefore, MoSGrid chose to employ SAML assertions in the infrastructure of its science gateway.

SAML assertions are used in UNICORE infrastructures, GridShib [21] allows applying SAML assertions in Globus Toolkit infrastructures, and SAML is supported by various security infrastructures like the open-source software Shibboleth. The latter provides federated and scalable access to DCIs and is widely used (e.g., in the Swiss grid infrastructure [22], in a UK grid infrastructure [23]). Currently, no standard has evolved for cloud infrastructures like Amazon EC2 [24]. The security relies on login/password mechanisms and encryption and depends on the vendors and institutes offering access to the DCIs [25].

A survey by EGI.eu (European Grid Infrastructure, the coordinating body of the EU EGI-InSPIRE project) [26] about requirements for authentication and authorization infrastructures for DCIs has led to the result that the major technologies currently used include SAML and X.509 certificates. One of the goals for the future is to bridge the different security domains (e.g., via Shibboleth). Since the MoSGrid science gateway already uses SAML, its security infrastructure can be easily adapted to rely on Shibboleth for user authentication instead of certificates.

## 2.2 Community Management

Community management includes collecting workgroups interested in molecular simulation into suited subgroups. Chemists are not a homogeneous user group but can be subdivided into several domains depending on the simulation method. The domain “quantum chemistry (QM)” is based on approximated solutions of the Schrödinger equation and among the most prominent codes are Gaussian, NWChem [27], and Turbomole [28]. They possess average scalability and are mostly used for parallel running jobs. The domain “molecular dynamics (MD)” is based on forcefields describing molecular interactions and exhibits good scalability within the codes Gromacs, NWChem, and Amber [29]. The domain “docking” is based on simplified forcefields with excellent scalability as implemented for instance in CADDSuite and FlexX [30]. Combing these three domains within one science gateway and managing them within one community demands to consider the very different requirements of the three domains: i) the input and output complexity varies strongly between a rather simple QM input file of some lines and the more complex group of small files for submitting a MD simulation; ii) analogue, a MD simulation yields 100 MB up to several GB whereas a QM calculation gives only MB amounts; iii) within the simulation codes, large differences exist between the needed storage time; iv) some QM calculations such as optimisations allow to be restarted, some frequency calculations are long-runners which need weeks of time in one bunch; v) finally, the desired workflows are highly depending on the domains and simulation codes. All these detailed needs and requirements have to be taken into account when managing such a combined community.

In order to design a user-friendly portal for molecular simulations, direct feedback from active users is urgently needed. Therefore, MoSGrid organizes at the moment 120 users from the three domains. In the early project phase requirement analyses have been performed. With MoSGrid now being in the beta phase, 20 users test the MoSGrid portlets regularly. The community management canalizes the feedback and disseminates actual developments back to

the community. The dissemination is accomplished via regular newsletters, the webpage, conference activity within the specific communities and tutorial workshops (planned in 2012).

### 2.3 Domain-related Features

WS-PGRADE and other workflow-enabled grid portals (e.g., Genius [31], OGCE [32], Virtual Toolkit [33]) offer features for managing the whole life cycle of generic workflows; from creating and changing, over invoking and monitoring, to aborting or deleting workflows on diverse DCIs. They allow intuitive parameterization of the included jobs for fitting DCIs and executables. Furthermore, existing workflows can be shared via a repository. The MoSGrid science gateway has been extended with the unique feature to use the incarnation database (IDB) of UNICORE, which fosters to select available tools on selected UNICORE resources. Users are relieved from uploading binaries or executable scripts and obtain up-to-date information about the currently available tools.

Additionally, three domain-specific portlets have been developed to meet the requirements in the targeted domains; quantum chemistry, molecular dynamics, and docking. Their user interfaces are tailored especially to the parameters for a workflow in each domain and they allow visualizing results via Jmol [34]. The quantum chemistry portlet, for example, supports the users by automatically extracting relevant information from Gaussian output files.

Portals like WeNMR [35] are also adapted for the use of tools like Gromacs or Haddock [36] for molecular simulations but they lack features for selecting resources or managing workflows. However, they provide file conversion from different formats to fitting formats for a selected tool. In the field of molecular simulations, there are some standard formats available like PDB (Protein Data Bank) [37], which are unfortunately often slightly not adhered to stringently. MoSGrid has been developing MSML (Molecular Simulation Markup Language) to overcome some of these difficulties. MSML is designed for specifying structural information of small and large molecules and macromolecules as well as the input parameters and results of various molecular simulations. Parsers and adapters are being developed to allow a consistent data representation independent of the applied tools and to enable workflows with automatic conversion between different jobs.

## 3. MoSGrid Security Infrastructure

The implemented security concept in the MoSGrid science gateway [38] covers all layers of the involved infrastructure (see Fig. 1).

Visible to the user, Liferay [39] including WS-PGRADE serves as intuitive user interface to the science gateway on the top layer. The following high-level middleware service layer consists of the gUSE services and XtreamFS. They enable access to compute resources and the underlying storage. The grid middleware layer includes the UNICORE middleware. It enables access to geographically distributed computing resources. High-performance computing facilities are depicted as the base layer. They represent the actual computing resources accessible via UNICORE.

In MoSGrid, Liferay is deployed inside the Apache Tomcat 6 [40] application server, which is based on the Servlet 2.5 [41] and the JavaServer Pages 2.1 [42] specifications and security models.

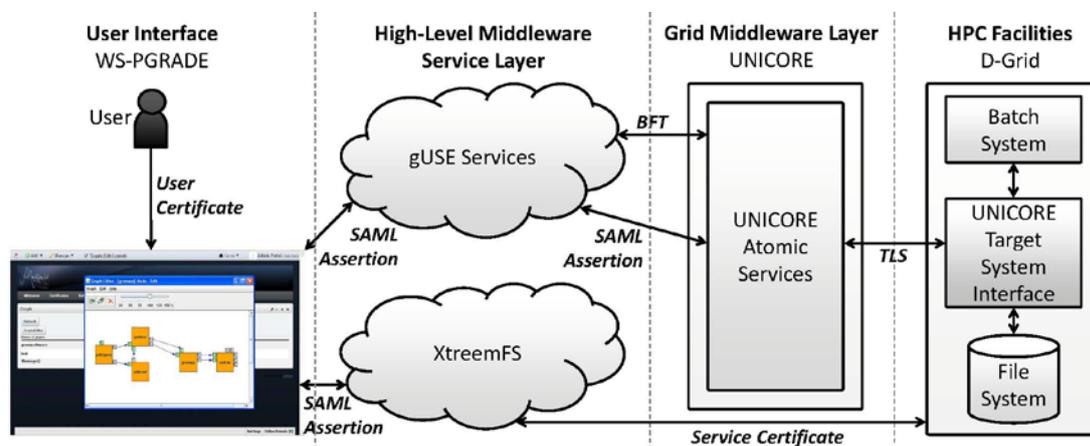


Figure 1: The MoSGrid Security Infrastructure

The access control inside the science gateway concerning resources is handled by Apache Tomcat, which also offers role-based authorization and login procedures via name and password. Liferay extends these by adding role attributes like organization, community, and group. MoSGrid uses these attributes to implement six hierarchical roles: guests, novice users, advanced users, consortium users, developers, and administrators. Depending on the role, the user has further rights, from simple viewing of public content to managing the whole science gateway. Additional to the depicted user management, a SAML assertion is needed to access the underlying DCI infrastructure. The token is generated by a certificate portlet of WS-PGRADE. For that to happen, the user needs an X.509 user certificate. Since the certificate is security sensitive, the creation process is designed for the certificate to not leave the user's computer by using an applet running in the user's browser.

WS-PGRADE provides graphical ways to create, modify, re-use, submit, and monitor computational workflows. In general, computational workflows can be described as a sequence of connected steps in a defined order based on their control and data dependencies. They can be shared between users via a local repository. The user is enabled to easily interact with the UNICORE grid middleware through configurable registries. The gUSE services provide the underlying services for managing workflows, data in workflows, and the submitters, which let gUSE interact with grid middlewares. The UNICORE submitter, developed in MoSGrid, uses the UCC (UNICORE commandline client) [43] libraries and includes the use of SAML assertions for authentication. The user needs to generate a SAML assertion via the certificate portlet for the UNICORE submitter to be able to interact with UNICORE. The submitter defines the job characteristics via the job description language JSDL (Job Submission Description Language) [44] before submitting it. This includes for example imports, exports, walltime, and number of nodes. When submitting a job, its working directory (USpace) is automatically generated by UNICORE, the input files are imported by the submitter, and the application

defined in the IDB is executed on the cluster like initiated by the submitter. The IDB maps generic application names to site-specific paths of applications. This way only the application name needs to be known. To integrate gUSE functionality into domain specific portlets, gUSE provides the Application Specific Module (ASM). It provides a programming interface, which allows portlets to easily interact with gUSE, its services, and makes use of the gUSE security mechanisms. This makes it much easier for developers to provide well-tested, easily usable, and secure portlets for the users.

The MoSGrid science gateway uses XtreamFS for the basis of its distributed data management. XtreamFS is an efficient, object-oriented file management system. It provides a global namespace, good scalability, replication features, and it efficiently handles extensive file transfers. It also allows being easily integrated into other services. In MoSGrid, XtreamFS was extended to support SAML assertions. XtreamFS can be accessed via a dedicated portlet, a Java API, or with a new URL schema directly via UNICORE. With the URL schema files in XtreamFS can be directly referenced in workflows. This relieves the science gateway from directly handling distributed files.

#### **4. Addressing the International Molecular Simulation Community**

Several communities from a number of application fields participate in SCI-BUS. COMPCHEM [45], Bio-science Grid [46], the PROSIM protein docking community [47], and MoSGrid represent the life sciences. Partners of SCI-BUS and ER-flow are also members of the virtual organization (VO) Life Science Grid Community [48] that is one of the first established virtual user communities. The close collaboration between these communities allows the exchange of experiences, workflows, results, and molecular structures. Thus, they can benefit from each other without spending time on calculating the same molecular simulations or repeating unsuccessful simulations. The data is sensitive and the analysis of molecular data on a large scale is time-consuming and expensive. The output of a survey in the MoSGrid community shows that 70% would share their results and molecular structures in a repository after they have published them or own a patent. We assume that researchers of the other communities would agree to share data in a repository at a similar rate.

The willingness to share tools and workflows is even higher (nearly 90%). Thus, the developers of services for the communities can also profit from the synergies. They can coordinate domain-related work as well as general developments for the science gateways. The participation in and collaboration with the three projects SCI-BUS, ER-flow, and EDGI allows developers of domain-related tools and workflows to concentrate on the specific demands, whereas, general developments (e.g., security) can be completed by partners who are working on the infrastructure of science gateways. This leads to provision of advanced services better tailored to the users' needs and supporting innovation and efficiency in the scientific discovery process.

##### **4.1 Contributions via SCI-BUS**

SCI-BUS focuses on the coordinated further development of science gateways and on offering portlet repositories. It intends to support a toolset for a generic-purpose science

gateway technology, which can be used for the development of domain-specific science gateways for various types of user communities (see Fig. 2). The targeted technology relies, like the MoSGrid science gateway, on the Liferay-version of WS-PGRADE. Domain-specific evolutions will be published and disseminated as best-practice case studies.

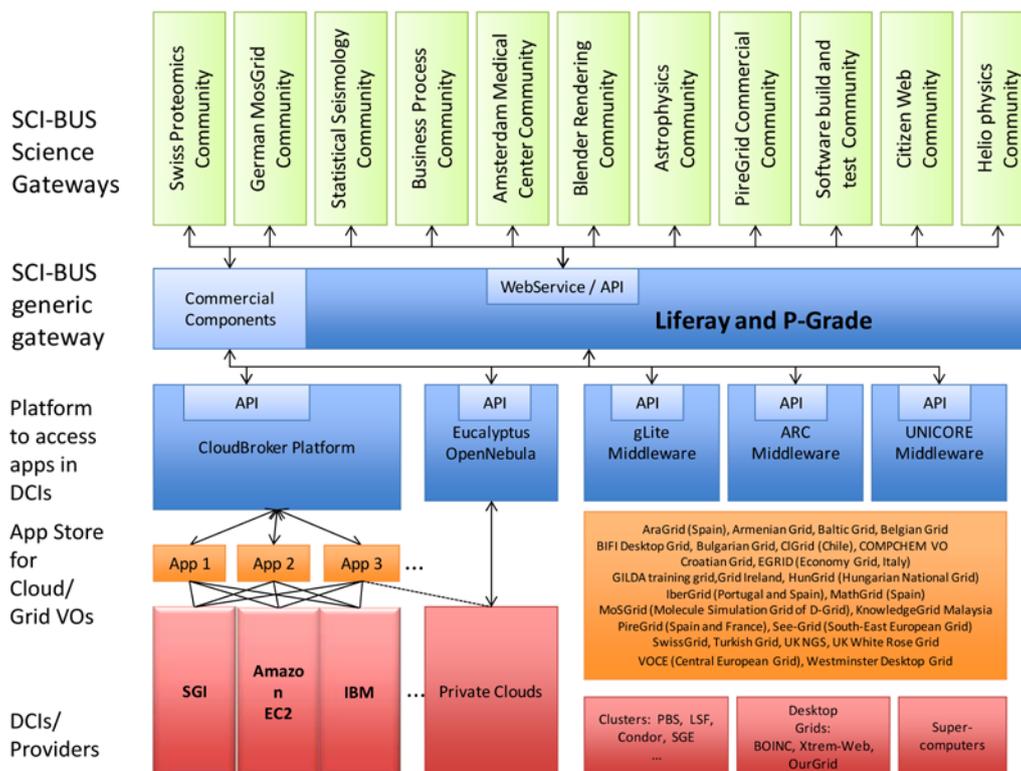


Figure 2: The SCI-BUS Infrastructure

Via the participation in SCI-BUS, MoSGrid intends to enhance services applicable for all three domains quantum chemistry, molecular dynamics, and docking supported in the science gateway. A semantic search engine will extend the science gateway that allows multiple XML schemata and supports the users to find appropriate workflows and results for their specific research. Existing semantic search engines (e.g., XSearch [49], XXL Search Engine[50]) will be evaluated and a fitting one will be supplied to support semantic searches for various ontologies using ontology mapping [51]. For concreteness, the first integrated ontology will rely on MSML. Furthermore, the search engine will support external data resources (e.g., PDB), which can be exported in XML format. Local copies of selected external data sources will be made available in the science gateway and to the workflows. This concept ensures that it can be adapted to other science gateways with different external data resources and various XML schemata.

A further enhancement meets the requirement for visualization and interactive editing of molecular simulation results. A 3D molecule editor will support the users to add information to a molecule. MoSGrid will use WebGL [52], which is a 3D standard that derived from OpenGL [53] and is developed for dynamic 3D web applications. It can be integrated plugin-free in browsers and is already supported by Firefox, Chrome, Internet Explorer, Opera, and Safari.

MoSGrid's investigation of available 3D molecule editors applying WebGL has led to ChemDoodle Web components [54], which already provide basic features for editing molecules. Modules for the visualization of molecules based on ChemDoodle Web components have been integrated into a portlet in the MoSGrid science gateway. The application of a 3D molecule editor in the science gateway allows not only graphical interaction by the users and to process the whole lifecycle of data analysis but also simplifies the process of creating figures capable for publications.

The expected growth of the community will be met with the setup of several gateway instances to foster good response times even if a very large number of users simultaneously uses the gateway. Firstly, three MoSGrid science gateways will be set up where two of them serve as backup. In such cases the load balancing among the gateways should be solved concerning both the user interface and the backend. In case of the user interface the main goal is to reduce response time even if very large numbers of users simultaneously use the gateway. In case of the backend, the goal is to evenly distribute the management of jobs when one of the gateways is overloaded by several large parameter sweep applications that require the management of extremely large numbers of jobs. A backup gateway can take over such applications managed from the primary gateway.

#### **4.2 Contributions via ER-flow**

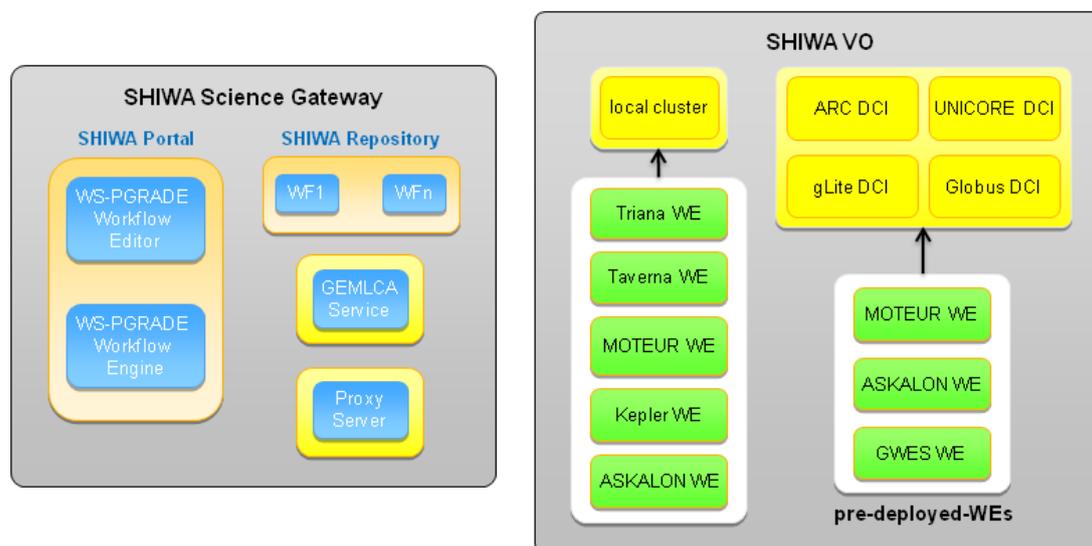
ER-flow concentrates on workflow interoperability and data sharing. It aims to create a community to promote workflow sharing and to investigate data interoperability supporting the SHIWA [55] simulation platform (see Fig. 3) and to promote its achievements.

In the area of workflow sharing it targets different research communities that use workflows to run their experiments. The participating communities are Astrophysics (IVOA), Computational Chemistry (MoSGrid), Heliophysics (HELIO), and Life Sciences (Life Sciences Grid Community). These communities are going to be supported in porting their applications to the SHIWA simulation platform, which will enable convenient usage of the applications. The involved partners of MoSGrid will analyze the three major computational chemistry domains within the MoSGrid community and identify which applications supported by the MoSGrid science gateway are required most. They will organize two MoSGrid specific workshops to disseminate and discuss the ported applications. As the contact point with the MoSGrid computational chemistry community, they will work on enlargement into the European context.

The project is tightly connected with the national grid infrastructures through EGI.eu, which will help to involve further major research communities. The research communities will select pilot workflows to demonstrate how to develop, use, and share workflows. It will port these pilot workflows to the simulation platform and publish them in a workflow repository. This will show how to use workflows from the simulation platform in experiments and how to modify them to create own workflows. The pilot workflows will create a critical mass of workflows to promote and enable workflow sharing inside and between communities.

It is also planned to conduct a study on data interoperability regarding workflows. Requirements of the MoSGrid community as well as of the other supported communities on data interoperability will be gathered and protocols and standards that support interoperability

will be researched. The study will make recommendations on how to best achieve data interoperability for the involved research areas, e.g., molecular simulations.



**Figure 3: The SHIWA Platform**

### 4.3 Contributions via EDGI

EDGI aims to foster Desktop Grids (DGs) in Europe. To reach this goal, EDGI develops necessary desktop grid software components [56]. In addition, EDGI supports the user communities in EGI.eu and coordinates the activities of desktop grids in Europe by establishing the International Desktop Grid Federation (IDGF) [57]. This federation allows a sustainable user-support even after EDGI's project end. A DG aggregates and utilizes otherwise unused idle resources from desktop computers. Due to the distributed and loosely coupled architecture of DGs and the high number of available resources, they are perfectly suited to serve batch jobs. These can typically be coarse-grained parallel tasks like parameter studies. Use cases of molecular simulations include extensive parameter studies and DGs are an option for the molecular simulation community to process their data on a large scale.

DG systems are beneficial for two different use cases. The first use case is to implement a desktop grid on already available computers in a company or institute. In this case, the aim is to get cheap resources within the secure environment and use them as resources for computations. The second use case is to contribute own, private resources for popular projects. A well-known example in this scope is the seti@home project [58]. EDGI supports both use cases. Thus in practice, the supported desktop grids are built on institutional as well as volunteered resources.

The EDGI architecture [59] allows the combination of different multi-national desktop grids to a new extremely large desktop grid suitable even for very heavy users. To reach this goal, the different multinational infrastructures with their own technologies and software stacks have to be integrated. EDGI has developed software bridges to allow interoperability to some of the most important service grids including UNICORE, ARC [60], and gLite (see Fig. 4). The

Generic Grid-Grid Bridge (3G-Bridge) [61] is an abstraction layer to arbitrary grid resources for job submissions. If the available desktop grid resources run short, it is additionally possible to integrate cloud resources that are managed by OpenNebula [62], Eucalyptus [63], or OpenStack [64]. In addition to the grid adoption, EDGI has a strong focus on the integration of promising programs into a DG application repository. Here, partners of MoSGrid and EDGI are integrating Gromacs via UNICORE in DGs. Further integrations of software packages for the molecular simulation community are planned.

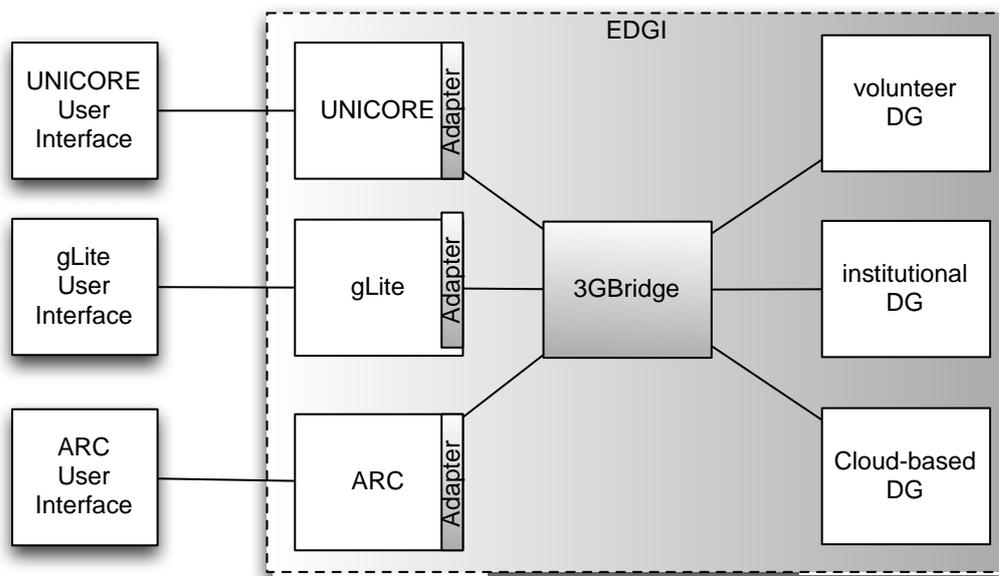


Figure 4: The EDGI Software Infrastructure

## 5. Summary and Outlook

Currently, the MoSGrid community consists of about 120 German users. However, the MoSGrid science gateways and its infrastructure with the integrated tools and workflows as well as generated results are valuable for the international community. The design of the science gateway is open for international use and the software will be provided as open-source software.

SCI-BUS will advance a generic-purpose science gateway technology that will provide access to DCIs and their services in Europe. Additionally, MoSGrid and other life-sciences-related communities contribute to the project and partners of the communities will further develop these gateways. Several user community workshops will bring together users, developers, and providers of the science gateways. The soon starting project ER-flow and the collaboration with EDGI forges links to more related communities and opens up the possibility for the molecular simulation community to influence workflow interoperability, data interoperability, and desktop grids on the European level.

**Acknowledgements** This work is supported by the German Ministry of Education and Research under project grant #01IG09006 (MoSGrid) and by the European Commission's 7th Framework

Programme under grant agreement #RI-261556 (EDGI), #RI-261323 (EGI-InSPIRE), #261585 (SHIWA), and #RI-283481 (SCI-BUS).

## References

- [1] S. Gesing, P. Kacsuk, M. Kozlowszky, G. Birkenheuer, D. Blunk, S. Breuers, A. Brinkmann, G. Fels, R. Grunzke, S. Herres-Pawlis, J. Krüger, L. Packschies, R. Müller-Pfefferkorn, P. Schäfer, T. Steinke, A. Szikszay Fabri, K. Warzecha, M. Wewior, and O. Kohlbacher, *A Science Gateway for Molecular Simulations*, in proceedings of *EGI User Forum 2011*, Book of Abstracts, pp. 94-95, ISBN 978 90 816927 1 7 (2011).
- [2] *MoSGrid science gateway*, <http://www.mosgrid.de>
- [3] M.J. Frisch, et al., *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford CT (2004).
- [4] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, *Journal of Chemical Theory and Computation* 4(3), 435-447 (2008).
- [5] *CADDSuite (Computer-aided Drug Design Suite)*, <http://www.ball-project.org/caddsuite>.
- [6] P. Kacsuk, *P-GRADE portal family for grid infrastructure*, *Concurrency and Computation: Practice and Experience*, 23(3):235-245 (2011).
- [7] Z. Farkas and P. Kacsuk. *P-GRADE Portal: a generic workflow system to support user communities*, *Future Generation Computer Systems*, 27(5):454-465 (2011).
- [8] A. Streit, P. Bala, A. Beck-Ratzka, K. Benedyczak, S. Bergmann, R. Brey, J. M. Daivandy, B. Demuth, A. Eifer, A. Giesler, B. Hagemeyer, V. Huber, S. Holl, N. Lamla, D. Mallmann, A. S. Memon, M. S. Memon, M. Rambadt, M. Riedel, M. Romberg, B. Schuller, T. Schlauch, A. Schreiber, T. Soddemann, and W. Ziegler, *Unicore 6 - Recent and Future Advancements*, *JUEL-4319*, ISSN 0944-2952 (2010).
- [9] F. Hupfeld, T. Cortes, B. Kolbeck, J. Stender, E. Focht, M. Hess, J. Malo, J. Marti, and E. Cesario, *The XtreamFS Architecture - A Case for Object-based File Systems in Grids*, *Concurrency and Computation: Practice and Experience*, 20(17):2049-2060, ISSN 1532-0634 (2008).
- [10] *Security Assertion Markup Language (SAML) V2.0*, <http://docs.oasis-open.org/security/saml/v2.0/saml-2.0-os.zip> (2002).
- [11] R. L. Morgan, S. Cantor, S. Carmody, W. Hoehn, and K. Klingenstein, *Federated Security: The Shibboleth Approach*, *EDUCAUSE Quarterly*, 27(4):12-17, ISSN 1528-5324 (2004).
- [12] S. Tuecke, V. Welch, and J. Novotny, *An Online Credential Repository for the Grid: MyProxy*, in proceedings of the *Tenth International Symposium on High Performance Distributed Computing (HPDC-10)*, pages 104-111, IEEE Press (2001).
- [13] *SCI-BUS (SCientific gateway Based User Support)*, <http://www.sci-bus.eu/>.
- [14] *EDGI (European Desktop Grid Initiative)*, <http://edgi-project.eu/>.
- [15] I. Foster, *Globus Toolkit Version 4: Software for Service-Oriented Systems*, in proceedings *IFIP International Conference on Network and Parallel Computing*, Springer-Verlag (LNCS 3779):2-13 (2006).
- [16] *gLite 3.2 User Guide*, <https://edms.cern.ch/file/722398/1.4/gLite-3-UserGuide.pdf>.

- [17] International Telecommunication Union (ITU), *ITU-T Recommendation X.509*, <http://www.itu.int/rec/T-REC-X.509/en> (1988).
- [18] *IGTF (The International Grid Trust Federation)*, <http://www.igtf.net/>.
- [19] *EUGridPMA (European Grid Policy Management Authority)*, <http://www.eugridpma.org/>.
- [20] I. Foster, C. Kesselman, G. Tsudik, and S. Tuecke, *A Security Infrastructure for Computational Grids*, in proceedings *CCS '98 Proceedings of the 5th ACM conference on Computer and communications security*, ISBN 1-58113-007-4 (1998).
- [21] T. Barton, J. Basney, T. Freeman, T. Scavo, F. Siebenlist, Von Welch, R. Ananthkrishnan, B. Baker, M. Goode, and K. Keahey, *Identity Federation and Attribute-based Authorization through the Globus Toolkit, Shibboleth, GridShib, and MyProxy*, in proceedings *5th Annual PKI R&D Workshop* (2006).
- [22] R. Murri, P. Z. Kunszt, S. Maffioletti, and V. Tschopp, *GridCertLib: A Single Sign-on Solution for Grid Web Applications and Portals*, *Journal of Grid Computing*, 9(4):441-453 (2011).
- [23] X. D. Wang, M. Jones, J. Jensen, A. Richards, D. Wallom, T. Ma, R. Frank, D. Spence, S. Young, C. Devereux, and N.I. Geddes, *Shibboleth Access for Resources on the National Grid Service (SARoNGS)*, in proceedings *Fifth International Conference on Information Assurance and Security*, volume 2, pages 338-341 (2009).
- [24] Amazon Web Services, *Amazon Elastic Compute Cloud (Amazon EC2)*, <http://aws.amazon.com/ec2/>.
- [25] US Department of Energy (DOE), *The Magellan Report on Cloud Computing for Science*, <http://www.nersc.gov/assets/StaffPublications/2012/MagellanFinalReport.pdf> (2011).
- [26] *EGI (European Grid Infrastructure)*, <http://www.egi.eu/>.
- [27] M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. van Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, and W.A. de Jong, *NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations*, *Computer Physics Communications* 181 (9): 1477-1489 (2010).
- [28] University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007, *TURBOMOLE*, <http://www.turbomole.com> (2011).
- [29] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossvai, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman, *AMBER 12*, University of California, San Francisco (2012).
- [30] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, *A fast flexible docking method using an incremental construction algorithm*, *J. Mol. Biol.* 261: 470-489 (1996).
- [31] R. Barbera, G. Andronico, G. Donvito, A. Falzone, J.J. Keijsers, G. La Rocca, L. Milanese, G. P. Maggi, and S. Vicario, *A Grid Portal with Robot Certificates for Bioinformatics Phylogenetic Analyses*, *Concurrency and Computation: Practice and Experience*, 23(3):246-255 (2011).
- [32] C. Zhang, I. Kelley, and G. Allen. *Grid portal solutions: a comparison of GridPortlets and OGCE*, *Concurrency and Computation: Practice & Experience - Workshop on Grid Computing Portals (GCE 2005)*, 19(12) (2007).

- [33] D. Szejnfeld, P. Dziubecki, P. Kopta, M. Kryszynski, T. Kuczynski, K. Kurowski, B. Ludwiczak, T. Piontek, D. Tarnawczyk, M. Wolniewicz, P. Domagalski, J. Nabrzyski, and K. Witkowski, *Vine Toolkit - Towards Portal Based Production Solutions for Scientific and Engineering Communities with Grid-Enabled Resources Support, Scalable Computing: Practice and Experience*, 11(2):161-172 (2011).
- [34] *Jmol*, <http://jmol.sourceforge.net/> (2011).
- [35] T. A. Wassenaar, M. van Dijk, N. Loureiro-Ferreira, G. van der Schot, S. J. de Vries, C. Schmitz, J. van der Zwan, R. Boelens, A. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H. R. A. Jonker, A. Bagaria, V. Jaravine, P. Güntert, H. Schwalbe, W. F. Vranken, J. F. Doreleijers, G. Vriend, G. W. Vuister, D. Franke, A. Kikhney, D. I. Svergun, R. Fogh, J. Ionides, E. D. Laue, C. Spronk, M. Verlat, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero, A. M.J.J. Bonvin, *WeNMR: Structural Biology on the Grid*, in proceedings *International Workshop on Science Gateways for Life Sciences (IWSG-Life 2011)*, CEUR Workshop Proceedings, <http://ceur-ws.org/Vol-819/paper4.pdf> (2011).
- [36] C. Dominguez, R. Boelens and A. M.J.J. Bonvin, *HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information*, *J. Am. Chem. Soc.* 125, 1731-1737 (2003).
- [37] H. M. Berman, K. Henrick, H. Nakamura, J. Markley, P. E. Bourne, and J. Westbrook, *Realism about PDB*, *Nature biotechnology*, 25(8):845-846 (2007).
- [38] S. Gesing, R. Grunzke, A. Balasko, G. Birkenheuer, D. Blunk, S. Breuers, A. Brinkmann, G. Fels, S. Herres-Pawlis, P. Kacsuk, M. Kozłowski, J. Krüger, L. Packschies, P. Schäfer, B. Schuller, J. Schuster, T. Steinke, A. Szikszay Fabri, M. Wewior, R. Müller-Pfefferkorn, and O. Kohlbacher, *Granular Security for a Science Gateway in Structural Bioinformatics*, in proceedings *International Workshop on Science Gateways for Life Sciences (IWSG-Life 2011)*, CEUR Workshop Proceedings, <http://ceur-ws.org/Vol-819/paper8.pdf> (2011).
- [39] Inc. Liferay, *Liferay*, <http://www.liferay.com>
- [40] The Apache Software Foundation, *Apache Tomcat*, <http://tomcat.apache.org/tomcat-6.0-doc/>
- [41] Java Community Process, *Java Servlet 2.5 Specification*, <http://jcp.org/aboutJava/communityprocess/mrel/jsr154/index.html>
- [42] Java Community Process, *JavaServer Pages 2.1*, <http://jcp.org/aboutJava/communityprocess/final/jsr245/index.html>
- [43] *UNICORE Commandline Client*, [http://www.unicore.eu/unicore/architecture/client-layer.php#anchor\\_ucc](http://www.unicore.eu/unicore/architecture/client-layer.php#anchor_ucc)
- [44] A. Anjomshoaa, F. Brisard, M. Drescher, D. Fellows, A. Ly, S. McGough, D. Pulsipher, and A. Savva, *Job Submission Description Language (JSDL) Specification*, <http://www.ggf.org/documents/GFD.56.pdf> (2005).
- [45] *COMPChem Virtual Organization*, <http://compchem.unipg.it/start.php>.
- [46] *Bio-Science Grid*, <http://www.swissbiogrid.org/>.
- [47] H. Heindl, T. Kiss, G. Terstyanszky, N. Weingarten, B. Afrough, and P. Greenwell, *ProSim: development of a user-friendly molecular modelling package*, in proceedings *International Beilstein Symposium on Glyco-Bioinformatics: Bits 'n' Bytes of Sugars*, pp. 61-85 (2010).
- [48] *Life Science Grid Community*, <http://lsgc.org/en/>

- [49] S. Cohen, J. Mamou, Y. Kanza, and Y. Sagiv, *XSearch: a semantic search engine for XML*, in proceedings *29th international conference on Very large data bases - Volume 29 (VLDB '2003)*, VLDB Endowment 45-56 (2003).
- [50] A. Theobald and G. Weikum, *The Index-Based XXL Search Engine for Querying XML Data with Relevance Ranking*, in proceedings *8th International Conference on Extending Database Technology: Advances in Database Technology (EDBT '02)*, Springer-Verlag, London, UK, 477-495 (2002).
- [51] Y. Kalfoglou and M. Schorlemmer, *Ontology mapping: the state of the art*, *The Knowledge Engineering Review*, 18(1):1-31 (2003).
- [52] Khronos Group, *WebGL Specification*, <https://www.khronos.org/registry/webgl/specs/1.0/> (2011).
- [53] E. Angel, *Interactive Computer Graphics: A Top-Down Approach with OpenGL Primer Package-2nd Edition* (2nd ed.). Prentice-Hall, Inc., Upper Saddle River, NJ, USA (2001).
- [54] iChemLabs, *ChemDoodle Web Components*, <http://web.chemdoodle.com/>.
- [55] *The SHIWA project*, <http://www.shiwa-workflow.eu>
- [56] M. Keller, J. Kovacs, and A. Brinkmann, *Desktop Grids Opening up to UNICORE*. in proceedings *UNICORE Summit 2011*, pp. 67-76, Torun, Polan, Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag, (2011).
- [57] *IDGF*, <http://desktopgridfederation.org/>
- [58] *seti@home*, <http://setiathome.berkeley.edu/>
- [59] P. Kacsuk, J. Kovacs, Z. Farkas, A. Cs. Marosi, and Z. Balaton, *Towards a powerful European DCI based on Desktop Grids*, *Journal of Grid Computing*, Vol 9, No 2, pp. 219-240 (2011).
- [60] O. Appleton, D. Cameron, J. Cernak, P. Dóbbé, M. Ellert, T. Frågåt, M. Grønager, D. Johansson, J. Jönemo, J. Kleist, M. Kočan, A. Konstantinov, B. Kónya, I. Márton, B. Mohn, S. Möller, H. Müller, Z. Nagy, J. Nilsen, F. Ould Saada, K. Pajchel, W. Qiang, A. Read, P. Rosendahl, G. Róczyei, M. Savko, M. Skou Andersen, O. Smirnova, P. Stefán, F. Szalai, A. Taga, S. Toor, A. Wäänänen, X. Zhou, *The next-generation ARC middleware*, *Annals of Telecommunications "Towards market-oriented Clouds"*, vol. 65, numbers 11-12, pp. 771-776 (2010).
- [61] Z. Farkas, P. Kacsuk, Z. Balaton, and G. Gombás, *Interoperability of BOINC and EGEE*, *Future Generation Computer Systems*, Vol. 26, no. 8, pp. 1092-1103 (2010).
- [62] B. Sotomayor, R. S. Montero, I. M. Llorente, and I. Foster, *Virtual Infrastructure Management in Private and Hybrid Clouds*, *IEEE Internet Computing*, vol. 13, no. 5, pp. 14-22 (2009).
- [63] D. Nurmi, R. Wolski, C. Grzegorzczuk, G. Obertelli, S. Soman, L. Youse, and D. Zagorodnov, *The Eucalyptus Open-Source Cloud-Computing System*, in proceedings *9th IEEE/ACM International Symposium on Cluster Computing and the Grid, CCGRID '09*, pages 124-131 (2009).
- [64] *OpenStack*, <http://openstack.org/>