

# Classification of Different Approaches for e-Science Applications in Next Generation Computing Infrastructures

Morris Riedel, Achim Streit, Felix Wolf, Thomas Lippert  
Jülich Supercomputing Centre, Forschungszentrum Jülich  
D-52425, Jülich, Germany  
m.riedel@fz-juelich.de

Dieter Kranzlmüller  
Ludwig-Maximilians-Universität München, Department of Computer Science  
D-80538, Munich, Germany

## Abstract

*Simulation and thus scientific computing is the third pillar alongside theory and experiment in today's science and engineering. The term e-science evolved as a new research field that focuses on collaboration in key areas of science using next generation infrastructures to extend the powers of scientific computing. This paper contributes to the field of e-science as a study of how scientists actually work within currently existing Grid and e-science infrastructures. Alongside numerous different scientific applications, we identified several common approaches with similar characteristics in different domains. These approaches are described together with a classification on how to perform e-science in next generation infrastructures. The paper is thus a survey paper which provides an overview of the e-science research domain.*

## 1. Introduction

Many scientific applications take advantage of the next generation computing infrastructures that evolved over the last couple of years to production environments. At the time of writing, such next generation infrastructures are represented by Grids with a certain overlap to clouds emerging in the commercial space. Today, such well known scientific Grid infrastructures are for example provided by Enabling Grids for e-Science (EGEE), Distributed European Infrastructure for Supercomputing Applications (DEISA), Open Science Grid (OSG), TeraGrid, and many others.

Although these projects claim production quality and are actually used by scientists world-wide on a daily basis, there are still many challenges in using them. For example, end-

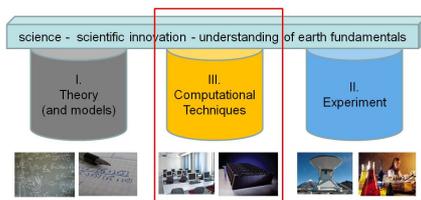
users often do not even know how to get access to these infrastructures and the procedures in place to provide such an access are still cumbersome and tedious. While in EGEE an end-user just joins one Virtual Organization (VO), scientists of DEISA have to pass a selection process based on scientific proposals that are evaluated by a committee. Also, many scientists have no precise understanding of which infrastructure provides which kinds of computational resources and functionality, and how this functionality relates to their particular needs. Simplified computational jobs are distinguished between high throughput computing (HTC) and high performance computing (HPC) and thus they typically need to be executed on different computational resources provided in different Grid infrastructures. Additionally it is sometimes unclear who provides the resources, whether the scientists have to contribute their own computational resources to get involved, or if they just use the infrastructure without any kind of fee. Finally, once end-users get access, often they have to cope with problems such as getting the right valid certificates for using the infrastructure, and their certificates are incompatible with other infrastructures.

Today, people have learned to use the infrastructures by mastering these challenges and different technologies lead in practice to many different approaches to e-science. By talking to scientists via interviews and analyzing the way they work, we have been able to identify common approaches based on deployed Grid technologies. In this paper, we provide a classification of these different approaches to perform e-science on next generation infrastructures. We discuss established techniques in the context of e-science and give insights drawn from lessons learned while working with e-scientists in order to provide a precise understanding how e-science is performed today. Our proposed classification is aligned with a survey of use case applications using the different approaches.

This paper is structured as follows. After reviewing challenges for e-scientists using Grid infrastructures, Section 2 sets the scene by discussing the e-science paradigm in the context of well-known established techniques. Section 3 describes the classification of different approaches to e-science in general alongside numerous use case applications in particular. Finally, after surveying related work in Section 4, we present our conclusions in Section 5.

## 2. Computing Infrastructures & Fourth Pillar

Today, scientists regard computational techniques as the third pillar alongside experiment and theory as shown in Figure 1. In this illustration, the first pillar stands for a certain theory or a specific model in a research field. One specific example of this pillar are scientists that use complex mathematical models to predict the diffusion of harmful materials in soil. The second pillar points to experiments performed in laboratories or probes of material, for instance, of harmful materials in soil. The computational techniques allows for computer-simulations based on efficient numerical methods and known physical laws. In our example, scientists can calculate the flow of water underground and simulate the way in which various harmful substances react with potentially damaging consequences.



**Figure 1. Computational techniques are the third pillar alongside experiment and theory.**

In the context of the defined three pillars of science, the term enhanced science (e-science), sometimes called electronic science evolved in the last couple of years. One fundamental definition of e-science [8] by John Taylor is as follows: *e-science is about global collaboration in key areas of science and the next generation infrastructure that will enable it.* This definition has been extended in several ways to point to a particular focus or a dedicated technology. Nevertheless the definition is still valid and we keep this mature definition as a base for our discussions.

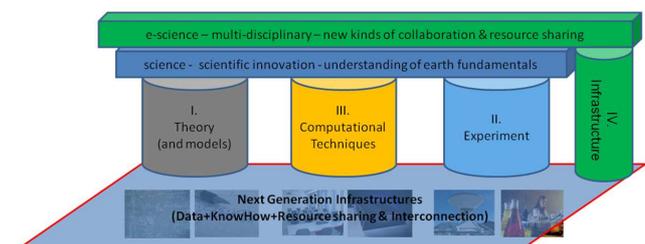
Today, such next generation infrastructures are mostly represented by Grids. There are various types of Grid infrastructures today, for instance, while EGEE and OSG are rather HTC oriented Grids, DEISA and TeraGrid are rather HPC-driven. In addition, clouds can be seen as infrastructures, which provide certain Grid capabilities. The boundaries and scope of these Grids are fundamentally different

leading to a well-known set of world-wide *Grid islands* majorly funded through public sources today. To provide an example, EGEE and DEISA are projects funded by the European Commission, while TeraGrid, OSG, D-Grid or NGS are national infrastructures based on different funding sources.

In addition, these different types of Grid infrastructures evolve differently. In Europe, we observe the evolution of EGEE-like grids in the so-called European Grid Initiative (EGI), whose members are National Grid Initiatives (NGIs). In contrast, the sustainability of DEISA and the evolution towards peta-scale HPC infrastructures is handled via different phases of the Partnership for Advanced Computing in Europe (PRACE) project [6]. Additionally, the different types of infrastructures and their different evolution phases lead to a different set of technologies deployed on them.

A fundamental goal of any technology deployed on the Grid infrastructure is to ease the usage of Grids and to facilitate the routine interaction of scientists, thereby enabling new collaborations in key areas of science. Often grand challenge problems require multi-disciplinary research and thus raise the demand for collaboration in key areas of science to tackle major problems of science and society today. Examples are protein folding, global weather prediction, or the virtual physiological human. Such problems can not be solved in a reasonable amount of time with computers that are broadly available today and require an increase in computing power by orders of magnitude.

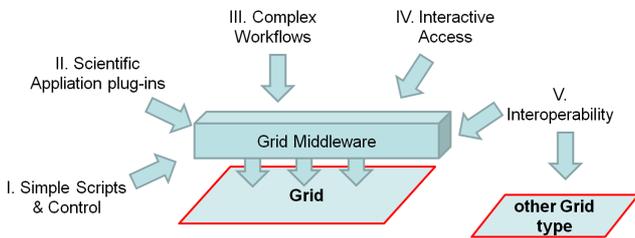
The solutions of these grand challenges can have a significant economic and social impact and the next generation infrastructures provide the computing power and collaboration mechanisms to support their research, hopefully and expectedly leading to new scientific breakthroughs. As shown in Figure 2, we consider e-science on top of science majorly supported by a fourth pillar, which is represented by emerging next generation infrastructures. The figure also illustrates that these infrastructures are not limited to computational resources, but also integrate large data facilities and thus enable know-how and resource sharing in an unprecedented way through high-speed network interconnections.



**Figure 2. e-Science is multi-disciplinary collaboration supported by Grid infrastructures that represent the fourth pillar.**

### 3. Classification of Different Approaches

In this section, we present the five identified different approaches to perform e-science. Each approach is underpinned by use case applications and discussed in the context of the four pillars of e-science as described in the previous chapter. Some approaches seem similar, and more precisely, there is sometimes no precise boundary between them allowing a smooth transition from one approach to another. Some approaches basically re-use concepts of other approaches to achieve a higher goal. Nevertheless, the distinction provided here is underlined by interviews and discussions with scientists of different research fields who use computing infrastructures very differently. While biologists work with Web-based portals, physicists in general seem to like command-line interfaces avoiding to use Graphical User Interfaces (GUIs) as much as possible.



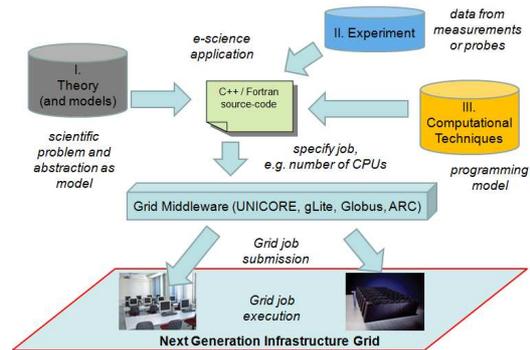
**Figure 3. Classification of different approaches to perform e-science today.**

Figure 3 illustrates the different approaches that all use Grid middleware as the access method to the Grid infrastructures. At the time of writing, there is quite a wide variety of Grid middleware available such as UNICORE, gLite, Globus Toolkits, ARC, GRIA, just to list but a few. Many of the use case applications below are using UNICORE as the access method to Grids, but the identified approaches can be also performed with other Grid middleware. In principle, the approaches are combinations of computational techniques (i.e. pillar III) and use of technologies available in today's next generation infrastructures (i.e. pillar IV). Needless to say, that theoretic models (i.e. pillar I) are typically implied since they are the foundation of the science performed. In addition, some scientific applications take also outcome of experiments (i.e. pillar II) into account: for instance, one project uses real values for burned materials in simulated computations to improve the design of trains in the field of disaster control research.

Each of the identified approaches use certain unique features of the Grid middleware deployed on the respective e-science infrastructure. While *approach I* (i.e. simple scripts & control functionalities) provides a very easy access with certain control functionalities via the submission of simple

scripts, *approach II* (i.e. application plug-ins) simplifies usage even more via predefined GUIs and a wide variety of job configuration options. *Approach III* (i.e. complex workflow) provides unique features in order to define dependencies between tasks leading to much flexibility for job submissions, while *approach IV* (i.e. interactive access) enables the most flexibility, for instance via SSH connections secured using Grid credentials, but the scientist utilizes less support from the Grid middleware itself in achieving tasks. Finally, *approach V* (i.e. interoperability) is unique in using different Grid infrastructures to perform different kinds of computations for one common scientific goal. One often used paradigm is to leverage the combined access to a HTC-driven Grid and an HPC-driven infrastructure. In the following sections we describe use cases of these approaches and thus provide end-user perspectives.

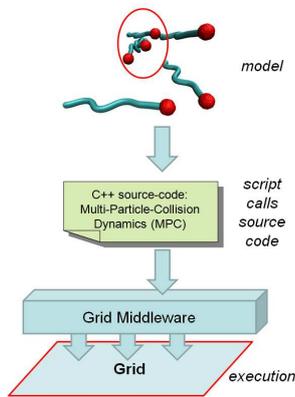
We identify typically three reasons why these different approaches emerged over time. First, the application itself required a certain approach, for instance, a simple application code that has to be called several times raises a demand for control functionalities such as loops, the same applies to workflows, which require mechanisms to handle different workflow steps. Second, the combination of technologies available on the infrastructures is different leading to different usage models and approaches. Finally, the third one is rather easy, because we learned there are certain approaches of scientists that evolved even over years via computational techniques (i.e. pillar III) in science and are thus adopted for the research supported within e-science infrastructures (i.e. pillar IV). For instance, the interactive access using Grid credentials is still an often used feature since SSH access has been used in scientific computing for years in most of the scientific communities. Figure 4 illustrates the dependency of the described pillars and Grid middleware and thus represents one basic usage model of Grids today.



**Figure 4. The basic usage model of Grids means a source code/application (often in C++/Fortran) is executed on the Grid infrastructure using Grid middleware.**

### 3.1. Simple Scripts & Control Functionality

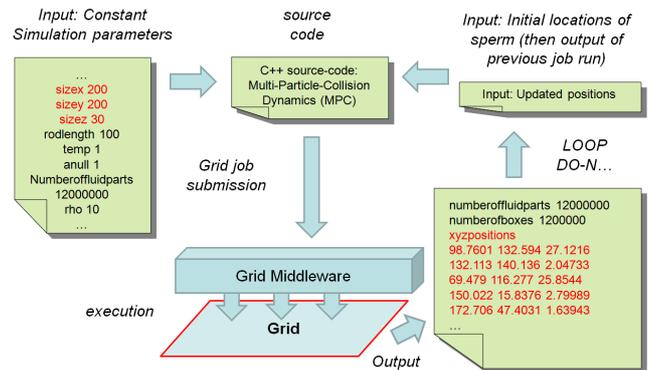
In the basic usage model that underlies any of the above mentioned approaches, we typically see C/C++ or Fortran90/77 codes that are used on computational Grid resources. These codes are typically derived from a particular model (i.e. pillar I) in a research field and are mostly submitted via simple UNIX-based scripts calling a parallel executable (i.e. pillar III) of the code in the Grid infrastructure (i.e. pillar IV). Submissions often use command-line interfaces, or when using simple control functionalities, also GUI-based Grid clients such as the UNICORE Rich Client or g-Eclipse. Control functionalities are constructs such as Do-N, Do Repeat, Hold Job, or If-Then-Else that all influence and control the execution of the executables in a limited manner.



**Figure 5. Experiments (pillar II) revealed an un-clarified swarm behavior of sperm (pillar I) that is a field of research using computational techniques (pillar III) such as OpenMP and the Message Passing Interface (MPI) on Grid infrastructures (pillar IV).**

Figure 5 illustrates a use case of this approach in the field of hydrodynamics, the study of liquids in motion. In this use case, a fluid dynamics code named as multi-particle collision dynamics (MPC) [12] is applied to simulate active biological system models (i.e. pillar I) named as sperm. Experiments (i.e. pillar II) have revealed an interesting swarm behavior of sperm, when the sperm concentration is high [18]. But the mechanism behind the experimental phenomenon is still not clear. Thus, the fundamental goal of this e-science application is to study the cluster size dependence for 2D and 3D systems in terms of studying the hydrodynamics interaction between sperm and explain its importance to the cooperation behavior. Simulations in 3D are very time consuming such that systematic study raises the demand for powerful computing infrastructures (i.e. pillar

IV), in our case DEISA. Figure 6 illustrates how the simple script mechanisms work well together with control functionalities such as the Do-N loop provided by the Grid middleware UNICORE deployed on DEISA. Many other Grid systems (i.e. Globus, gLite, etc.) also support such submission techniques and can be used in a similar way for this kind of e-science applications, but in terms of control activities, many other Grid middleware systems use a dedicated workflow system to perform such simple submissions.



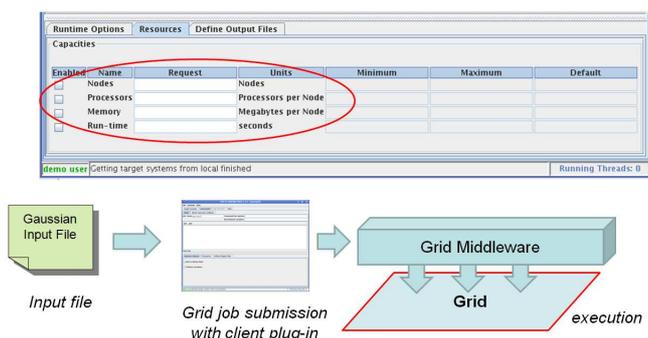
**Figure 6. This e-science application intensively uses the Do-N control functionality where the output of the previous job run is given as an input to the subsequent job.**

We also identified that this approach is independent from the underlying machine type. The above mentioned hydrodynamics e-science application is using the approach with the JUMP supercomputer with 41 SMP nodes where each node has 32 processors. This 1312 Power4+ 1.7 GHz CPU machine (8.9 Teraflops peak performance) is deployed in DEISA and accessible via UNICORE. Another e-science application [23] in the field of theoretical fluid mechanics is using the same approach with a very similar setup but with the JUGENE supercomputer, which consists of 65536 processors of the type 32-bit PowerPC 450 core 850 MHz (223 Teraflops peak performance). Both applications use UNICORE for job submissions of their scripts, but also use the Do-N control functionality on these rather different Grid resources (i.e. machines). The reason for end-users to use such loop functionality in particular are twofold. First, once the job is defined, new jobs are submitted to the Grid resource for each iteration without manual interaction, which is especially helpful during weekends. Second, many Grid resources have a limited job run time (e.g. 12 hours), and the Do-N control functionality provides a way to partition long jobs (e.g. over 60 hours) into smaller portions that not exceeding the allowed maximum run-time.

### 3.2. Scientific Application Plug-ins

The previous approach described above uses simple UNIX scripts for Grid job submissions. This implies that end-users (e-scientists) have to create the scripts that call executables by themselves. As a consequence scientific domain scientists have to know the potentials and drawbacks of UNIX scripts or scripting languages such as Python and Perl which are used for job submissions in Grid middleware systems. Hence, the scientists that are naturally experts in their research field and thus fully understand the theoretical model of their research (i.e. pillar I), also have to become experts in computational techniques (i.e. pillar III).

We also learned that different scientific communities handle these obstacles differently. Again let us point to the example of biologists who prefer high-level Web portals instead of low-level computational techniques and thus differ to physicists who typically also seek to understand the low-level details of high-end computers that they use for research in physics. In addition, the evolution of multi-core and many-core systems in general and the various options of programming high-end computers in particular lead to more and more complex computational techniques that have to be used in order to gain the maximum performance.



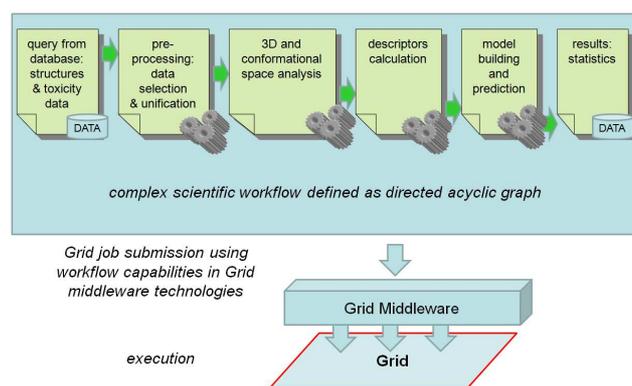
**Figure 7. The scientific application plug-in approach supports e-scientists in widely used applications such as Gaussian.**

In order to support e-scientists, many Grid clients such as the UNICORE Rich Client, g-Eclipse, or the Gridsphere Web portal [4] via the Vine Toolkit allow for client extensions that we call scientific application plug-ins. This plug-in approach basically provides a strong support for computational techniques within the Grid client. Typically, one widely used scientific application is supported in the Grid client via a scientific domain-specific plug-in. To provide an example, Figure 7 shows the Gaussian plug-in for the UNICORE client that enables end-users to easily submit Gaussian03 jobs to Grid infrastructures.

The support for end-users is often provided by offering configuration options of the corresponding scientific application or to allow for job-specific options via convenient GUI controls such as checkboxes or lists. In Figure 7, for instance, the end-user is able to conveniently choose the number of nodes and processors required for the job or is able to define memory requirements without knowing in which format the job is going to be submitted. The plug-in itself is responsible to create the required input scripts in the respective format for the Grid job submission taking the GUI inputs from end-users into account. In other words, this approach helps to focus on pillar I and thus uses pillar III transparently when submitting jobs to a computing infrastructure (i.e. pillar IV).

### 3.3. Complex Workflows

In many e-science applications we see that besides computation and network interconnectivity, data management is essential in that it often raise the demand for multiple processing steps required by the underlying mathematical method (i.e. pillar I). The combined use of computational tasks and data management tasks in conjunction with numerous control functionalities (cp. approach I) is leading to directed acyclic graphs (DAGs) that represent the overall scientific workflow. As a result, most Grid middleware systems support complex scientific workflows either naturally (e.g. UNICORE) or have dedicated workflow systems such as Taverna for the OMII-UK software stack and DAGMan for Condor. All of them have in common that they provide a GUI for the definition of DAGs since defining complex scientific workflows with XML configuration files and command-line interfaces is rather cumbersome.



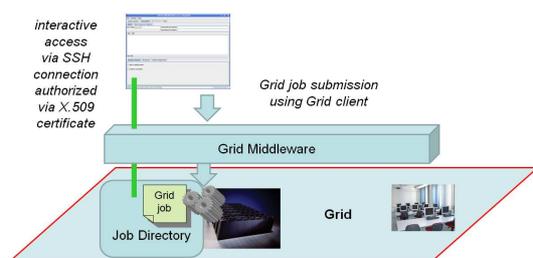
**Figure 8. Using the approach of complex workflows is often used with Grid middleware in e-science, for instance in health care research using QSAR workflows.**

Figure 8 illustrates one use case of this approach in the field of health care research with focus on regulatory purposes. In this context, the current European regulatory framework is named as registration, evaluation, authorisation, and restriction of chemical substances (REACH) [7]. The goal is to improve the protection of the human health/environment through the characterization of intrinsic properties of chemicals (i.e. pillar I) using the quantitative structure-activity relationships (QSAR) [10] computational method (i.e. pillar III).

An implementation of this particular approach is performed within the Chemomomentum project [2] wherein UNICORE is used with complex QSAR workflows. Different QSAR applications are combined in one workflow, which also includes access to existing databases with experimental data (i.e. pillar II) of chemicals. Figure 8 shows that the first step is to query from existing databases certain structure and toxicity data and afterwards the different computation steps are performed one after another. Finally, the results of the workflow processing are statistics that provide insights if chemical substances are compliant with REACH.

### 3.4. Interactive Access

In the previous three approaches, we learned that Grid middleware supports e-scientists in the way of how they actually submit jobs to Grid infrastructures (i.e. pillar IV). In this section, we describe an approach that is often used for numerous different purposes that can be summarized as interactive access usage models of Grid middleware. We learned from some e-science applications that we can distinguish between two interactive access paradigms that are SSH-like and bi-directional channels/streams.



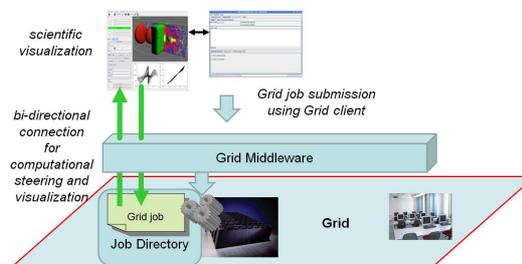
**Figure 9. Using the interactive access approach to check intermediate results.**

Figure 9 illustrates the approach with SSH and shows that the job submission is using the Grid client and thus Grid middleware, but the interactive access is used to get access to the job computation directory to review intermediate results. In this context, the SSH-like connections are typically established by preserving the well-known single sign-on requirements. That means the connections are established by

using the X.509 certificates of scientists as the authorization method in order to avoid tedious password requests for each connection. As a result, Grid middleware supports SSH through GSSSSH for Globus, the SSH Plugin for UNICORE, or glogin [22] for Globus and gLite. One example is the hydrodynamics research in the context of sperm where SSH is used at job runtime to review the positions of sperms in the input/output files. Sometimes certain configuration options create unrealistic positions of sperm (i.e. crossing maximum sizex) and thus taking a sample after some hours and to remove the failed job can save computational time.

Another paradigm of the interactive access approach is the use of bi-directional channels that are used in for scientific visualization and computational steering. In this approach the e-science application is typically submitted via typical Grid middleware techniques. But while the e-science application runs on the resource, scientific data is transferred through this connection to a scientific visualization to observe the current computation status. To influence the simulation during run-time, steering parameters are transferred from the visualization to the simulation.

In this paradigm of the interactive access, the Grid middleware authorizes and creates a bi-directional channel for numerous different use case applications. Different frameworks have been developed in the Grid middleware systems to enable this approach to end-users such as gGRID [17] in the context of gLite, eViz [3] in the context of Globus, and COVS [19] in the context of UNICORE. Examples for this approach are the plasma physics code PEPC [14] and the astro-physics code nbody6++ [24] that are used with the UNICORE-based COVS framework implementation for collaborative visualization and steering sessions.

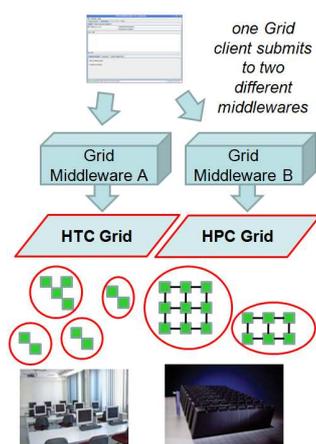


**Figure 10. Bi-directional channel used for scientific visualization and steering.**

### 3.5. Interoperability

Significant international and broader interdisciplinary research is increasingly carried out by global collaborations that use multiple interoperable next generation infrastructures such as Grids. In that sense interoperability represents the final approach to perform e-science. Of course, this im-

plies the use of several of the above mentioned approaches, but as a whole represents an approach useful to numerous e-science topics as well. Today, many e-science applications already take advantage of one single e-science infrastructure to simulate phenomena related to a specific scientific or engineering domain on advanced (parallel) computer architectures. In addition, more and more commercial players adopt the concepts of next generation infrastructures to enable new kinds of economic applications and flexible resource usage models such as clouds.



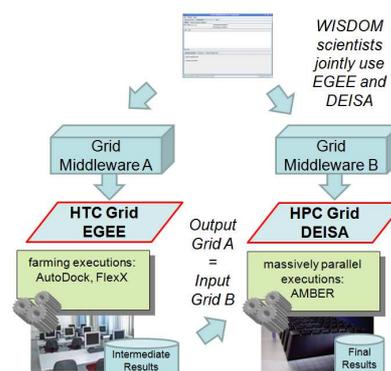
**Figure 11. Interoperability of HTC and HPC infrastructures enables new types of e-science.**

More recently, increasing complexity of e-science application theories (i.e. pillar I) that embrace multiple physical models (i.e. multi-physics) and consider a larger range of scales (i.e. multi-scale) is creating a steadily growing demand to the demand of world-wide interoperable infrastructures (pillar IV) that allow for new innovative types of e-science by jointly using HTC- and HPC-driven infrastructures with known computational techniques (i.e. pillar III) as shown in Figure 11. Thus the only option left to satisfy increasing e-science application demands is to harness a united federation of world-wide Grids, which provides transparent access to different kinds of resources and services.

The interoperability approach itself is a research field with many activities such as those within the Open Grid Forum (OGF) Grid Interoperation Now (GIN) [21] group or the International Grid Interoperation and Interoperability Workshops (IGIWs) [5]. Many projects support interoperability on a pair-wise fashion such as EGEE and OSG, or DEISA and TeraGrid. One use case of the approach via the WISDOM project [9], which recently performs research by jointly using the two major European e-science infrastructures EGEE and DEISA [20]. WISDOM stands for Wide In Silico Docking On Malaria and is developing new drugs for

these neglected diseases with a particular focus on malaria. The goal is to accelerate research and development for diseases via reduced research and development costs [16]. This is achieved by decreasing the amount of in vitro experiments (i.e. pillar II) through using more in silico docking computations (i.e. pillar III) on the infrastructures (i.e. pillar IV), before the drugs go into clinical phases. In silico docking computations represent a computational technique to predict whether one molecule will bind to another.

Figure 12 illustrates the WISDOM project wherein e-scientists submit jobs from one client to both, the EGEE and DEISA. It is important to note that the type of jobs substantially differ from each other. In the first part of the scientific process, WISDOM scientists are using EGEE with farming jobs based on AutoDock [15] and FlexX [11] to perform the in silico docking computations. The result is a list that have a higher potential becoming drugs but not yet constitutes the final selection. Therefore, in the second part of the process, WISDOM scientists use the DEISA infrastructure with massively parallel jobs based on the AMBER molecular dynamics package [1] to identify those candidates with the highest suitability chances as input for clinical phases. Thus by using the interoperability approach, the WISDOM scientists enormously accelerate the drug discovery process.



**Figure 12. WISDOM e-scientists leverage interoperability between EGEE and DEISA.**

#### 4. Related Work

There is not a wide variety of related work in terms of a rather theoretical consideration of ways to perform e-science in next generation infrastructures such as Grids. In many cases, related work focuses on a specific technology or project, or the related work references higher level architectures such as the Open Grid Services Architecture (OGSA) [13]. But OGSA provides no classification of specific scientific approaches and only a general framework of services for performing e-science in infrastructures. In con-

trast to our approach, OGSA is a rather high-level architecture and we identified the mentioned approaches by focusing on lessons learned with real applications on infrastructures in close collaboration with e-scientists.

## 5. Conclusions

We presented a classification of several approaches to perform e-science in the context of the pillars I to IV. The classification has been underpinned with lessons learned from e-science applications that are executed on a daily basis on Grid infrastructures such as DEISA or EGEE today. The identified approaches are not mutual exclusive and thus some approaches re-use other approaches in a specific manner. First, we identified approach I, which is commonly found in the most Grid infrastructures that allows for the submission of simple scripts. But the availability of control functionalities (e.g. loops) can be less often found mostly because loops can be, in principle, also developed within UNIX scripts itself in a slightly more complicated way.

The major difference between approach I and II, which represents the use of scientific application plugins, is the abstraction level. Approach II provides similar functionalities tuned for a specific application to provide a more high-level access mechanism to the Grid conveniently usable within Grid clients or Web portals. Today, there are many scientific application plugins used in infrastructures that are mostly related to well-known software packages (e.g. AMBER, AutoDock, Gaussian, etc.). In comparison with approach I and II, complex workflows (i.e. approach III) raise the demand for GUIs that support scientists with the definition of DAGs. We learned that many complex workflows have not only computational steps but also often data management steps that go beyond typical job data staging. Although often important for e-science applications in the infrastructures, many Grid middleware systems does not support workflows per default and thus imply the installation of related workflow tools (e.g. Taverna, DAGMan, etc.).

Different from the approaches I, II and III is the interactive access approach, which focuses on connections established by Grid middleware. We learned that this approach is mostly used with SSH connections to review output files during job runs. In some cases, such a connection is used for real-time visualization and, if a bi-directional connection is provided, for computational steering as well. In a similar manner like approach IV, approach V is also re-using numerous other approaches, but its uniqueness is the use of more than one infrastructure in parallel to solve a scientific goal. But we also identified that interoperability of infrastructures is still a work in progress that in the most cases is related to the adoption of OGSA conform open standards that emerge slowly raising a demand for a more light-weight interoperability reference model than full OGSA.

## References

- [1] AMBER - Assisted Model Building with Energy Refinement. <http://amber.scripps.edu>.
- [2] Chemomomentum Project. <http://www.chemomomentum.org/>.
- [3] EViz Project. <http://www.eviz.org>.
- [4] GRIDSphere. <http://www.gridisphere.org/>.
- [5] IGIW - International Grid Interoperability and Interoperation Workshop. <http://www.fz-juelich.de/jsc/igiw>.
- [6] PRACE - Partnership for Advanced Computing in Europe. <http://www.prace-project.eu/>.
- [7] REACH - Registration Evaluation Authorisation and Restriction of Chemical substances. <http://ec.europa.eu/environment/chemicals/reach/reachintro.htm>.
- [8] Taylor, e-Science Def. <http://www.e-science.clrc.ac.uk>.
- [9] Wide in silico docking on malaria (wisdom). <http://wisdom.healthgrid.org>.
- [10] Casalegno et al. Abstracts of QSAR-related Publications: Multivariate Analysis, QSAR Combinatorial Science.
- [11] S. S. J. Cross. Improved flexx docking using flexx-determined base fragment placement. In *Journal Of Chemical Information And Modeling* 45(4), pages 993–1001, 2005.
- [12] J. Elgeti et al. Hydrodynamics of active mesoscopic systems. In *NIC Series* 39, 53, 2008.
- [13] I. Foster et al. *Open Grid Services Architecture, Version 1.0*. Open Grid Forum Draft 30, 2005.
- [14] P. Gibbon et al. Performance Analysis and Visualization of the N-body tree code PEPC on massively parallel computers. In *Proc. of the ParCo, Malaga, Spain, 2005*.
- [15] R. Huey, G. Morris, A. Olson, and D. Goodsell. A semiempirical free energy force field with charge-based desolvation. In *J. Computational Chemistry*, 28, pages 1145–1152, 2007.
- [16] N. Jacq et al. World-wide in silico drug discovery against neglected and emerging diseases on grid infrastructures. In *GridPP*, 2007.
- [17] T. Köckerbauer et al. GVid - Video Coding and Encryption for Advanced Grid Visualization. In *Proceedings of the first Austrian Grid Symposium, Linz, 2005*.
- [18] H. D. M. Moore, K. Dvorakova, N. Jenkins, and W. G. Breed. Exceptional sperm cooperation in the wood mouse. In *Nature* 418, 174, 2002.
- [19] M. Riedel et al. Design and Evaluation of a Collaborative Online Visualization and Steering Framework Implementation for Computational Grids. In *Proc. of the 8th IEEE/ACM Int. Conf. on Grid Comp, Austin, USA, 2007*.
- [20] M. Riedel et al. Improving e-Science with Interoperability of the e-Infrastructures EGEE and DEISA. In *MIPRO*, 2007.
- [21] M. Riedel et al. Interoperation of World-Wide Production Infrastructures. In *Concurrency and Computation: Practice and Experience Journal*, 2008. OGF Special Issue.
- [22] H. Rosmanith and D. Kranzlmüller. glogin - A Multifunctional Interactive Tunnel into the Grid. In *Proc. Grid 2004*.
- [23] J. Schuhmacher et al. The fine-scale structure of turbulence. In *NIC Symposium*, 341, 2008.
- [24] R. Spurzem and E. Khalisi. Nbody6, features of the computer-code. 2003. <ftp://ftp.ari.uni-heidelberg.de/pub/staff/spurzem/nb6mpi/nbdoc.tar.gz>.