

**Science Gateways –  
Leveraging Modeling and  
Simulations in HPC Infrastructures  
via Increased Usability**

Sandra Gesing, Jens Krüger, Rion Dooley, Richard Grunzke,  
Marlon Pierce, Sonja Herres-Pawlis and Alexander Hoffmann  
[sandra.gesing@nd.edu](mailto:sandra.gesing@nd.edu)

HPCS 2015  
20 July 2015

UNIVERSITY OF NOTRE DAME

**CRC**  
CENTER FOR RESEARCH COMPUTING

**Science Gateways**

**CRC**  
CENTER FOR RESEARCH COMPUTING

"A Science Gateway is a community-developed set of tools, applications, and data that is integrated via a portal or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a specific community."

TeraGrid/XSEDE

Sandra Gesing Science Gateways 2

**Science Gateways**

**CRC**  
CENTER FOR RESEARCH COMPUTING

Sandra Gesing Science Gateways 3

**Science Gateways**

**CRC**  
CENTER FOR RESEARCH COMPUTING

Sandra Gesing Science Gateways 4

**Goals of Science Gateways**

**Usability**

"After all, usability really just means that making sure that something works well: that a person ... can use the thing - whether it's a Web site, a fighter jet, or a revolving door - for its intended purpose without getting hopelessly frustrated."

(Steve Krug in "Don't make me think! A Common Sense Approach to Web Usability", 2005)

Sandra Gesing Science Gateways 5

**Goals of Science Gateways**

**Reusability**

"The key to productivity is reusability. The easiest way to produce code is obviously to have it already!"

(John R. Bourne in "Object-oriented Engineering: Building Engineering Systems Using Smalltalk-80", 1992)

Sandra Gesing Science Gateways 6

**Usability - History**

1946 - Invention of a trackball by Ralph Benjamin (patented 1947 but kept as military secret)

1962 - Prototype of a mouse by Douglas Engelbart's team



AFOSR-3222  
Summary Report

AUGMENTING HUMAN INTELLECT: A CONCEPTUAL FRAMEWORK

Prepared for:  
DIRECTOR OF INFORMATION SCIENCES  
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH  
WASHINGTON 25, D.C.

By: D. C. Engelbart

CONTRACT AF 49(38)1024

Sandra Gesing Science Gateways 7

**Usability - History**

A. GENERAL

By "augmenting human intellect" we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems. Increased capability in this respect is taken to mean a mixture of the following: more-rapid comprehension, better comprehension, the possibility of gaining a useful degree of comprehension in a situation that previously was too complex, speedier solutions, better solutions, and the possibility of finding solutions to problems that before seemed insoluble. And by "complex situations" we include the professional problems of diplomats, executives, social scientists, life scientists, physical scientists, attorneys, designers--whether the problem situation exists for twenty minutes or twenty years. We do not speak of isolated clever tricks that help in particular situations. We refer to a way of life in an integrated domain where hunches, cut-and-try, intangibles, and the human "feel for a situation" usefully co-exist with powerful concepts, streamlined terminology and notation, sophisticated methods, and high-powered electronic aids.

Sandra Gesing Science Gateways 8

**Usability - History**

1969 - Alan Kay's Dynabook

1970 - Developments by Apple



Sandra Gesing Science Gateways 9

**Usability - History**

1969 - Alan Kay's Dynabook

1970 - Developments by Apple



Sandra Gesing Science Gateways 10

**Usability - History**

1980s - First science gateways

National Center for Biotechnology Information (NCBI) BLAST server, search results sent by email

1989 - World Wide Web developed by CERN

1990 - First web browser by Sir Tim Berners-Lee

WorldWideWeb, renamed to Nexus

MOSAIC evolved 1992

1992 - ISO 9241-2:1992 Ergonomic requirements for office work with visual display terminals (VDTs)

2004 - Announcement of TeraGrid's Science Gateway Program

Sandra Gesing Science Gateways 11

**TeraGrid – Scientific Profile**

- Start 2001
- Operated 2004 – 2011
- roughly 4,000 users at over 200 universities
- XSEDE is the successor since 2011

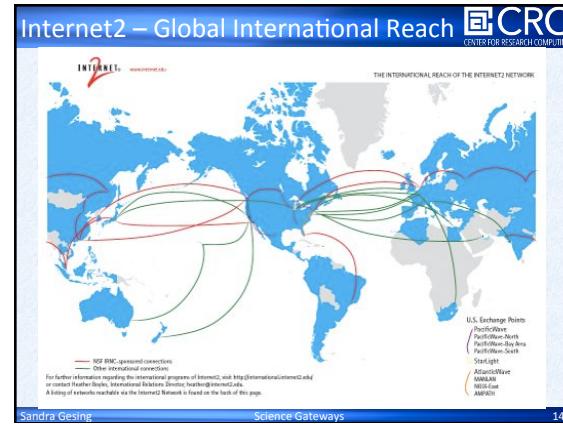
Allocated (%)	Used (%)	Scientific Discipline
19	23	Molecular Biosciences
17	23	Physics
14	10	Astronomical Sciences
12	21	Chemistry
10	4	Materials Research
8	6	Chemical, Thermal Systems
7	7	Atmospheric Sciences
3	2	Advanced Scientific Computing
2	0.5	Earth Sciences
2	0.5	Biological and Critical Systems
1	0.5	Ocean Sciences
1	0.5	Cross-Disciplinary Activities
1	0.5	Computer and Computation Research
0.5	0.25	Integrative Biology and Neuroscience
0.5	0.25	Mechanical and Structural Systems
0.5	0.25	Mathematical Sciences
0.5	0.25	Electrical and Communication Systems
0.5	0.25	Design and Manufacturing Systems
0.5	0.25	Environmental Biology

Sandra Gesing Science Gateways 12

**User Profile**

Generally not IT specialists...

Sandra Gesing Science Gateways 13



**Distributed Infrastructures**

- Definition Grid (Ian Foster, 1998)  
*„A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.“*
- Definition Cloud (Sam Johnston, 2008)  
*„The Cloud is what The Grid could have been.“*
- Virtualization
- Services
  - Infrastructure as a Service (IaaS)
  - Platform as a Service (PaaS)
  - Software as a Service (SaaS)

Sandra Gesing Science Gateways 15

**Distributed Infrastructures**

- National and international infrastructures (EGI, PRACE, XSEDE)
- UNICORE, HTCondor, Globus Toolkit, gLite
- Docker, OpenStack, Amazon EC2, Windows Azure
- Mixed computing paradigms
  - Grid of federated clusters (NGI-DE)
  - Grid of federated clouds (EGI, CERN-openlab)
  - Grid over cloud (StratusLab)
  - Cloud over grid (WNoDeS (Worker Nodes on Demand Services))

Sandra Gesing Science Gateways 16

**Workflows**

A sequence of connected steps in a defined order based on their control and data dependencies

Slide copied from: Stuart Owen, „Workflows with Taverna“

Sandra Gesing Science Gateways 17

**Workflow Systems**

- Different workflow concepts
- Different workflow languages
- Different workflow constructs

OGSA-DAI GUSE TRIANA  
Kepler Galaxy  
UNICORE Wf4Ever Taverna KNIME

Sandra Gesing Science Gateways 18

**Workflow Editors**

- Different technologies (workbenches, web-based)
- Different look-and-feel

Sandra Gesing Science Gateways 19

**Distributed Data Management**

- Amazon Simple Storage Service (Amazon S3)
- Google File System (GFS)
- XtreemFS
- dCache
- iRODS

Figure copied from „iRODS Overview“

Sandra Gesing Science Gateways 20

**State-of-the-Art**

- Data-intensive and compute-intensive problems
- High-speed networks available
- Sophisticated tools and methods available
- DCIs (Distributed Computing Infrastructures) available
- Workflow systems available
- Distributed data management available

**How do researchers use the tools and distributed environments on a large scale?**

Sandra Gesing Science Gateways 21

**Science Gateways**

**Why Science Gateways?**

- Modeling of data and simulations on a large scale with DCIs
- Usability of computational tools, applications, and DCIs is limited
- Scientists are generally not IT specialists

→ Need for intuitive and self-explanatory GUIs

Sandra Gesing Science Gateways 22

**Solution Science Gateways**

**Characteristics of science gateway**

- Convenient access to data and tools
- Single point of entry for multiple organizations
- No firewall issues
- Diverse frameworks and APIs

Sandra Gesing Science Gateways 23

**Frameworks and APIs**

**Re-inventing is not always necessary...**

Sandra Gesing Science Gateways 24

**Frameworks and APIs**

... but the model should fit to the demands of the community



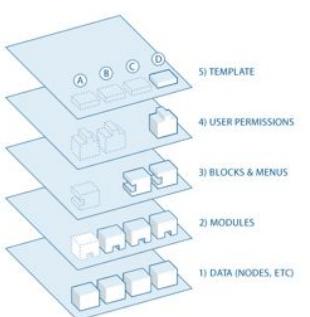
Sandra Gesing Science Gateways 25

**Science Gateway Technologies**

- Content management systems (Drupal)
- Libraries for implementation (Django)
- Portal frameworks (Liferay)
- Science gateway frameworks (WS-PGRADE, Galaxy)
  - Static layout
  - Layout extendable
  - Workflow-enabled
- APIs for implementation (Apache Airavata, Agave)

Sandra Gesing Science Gateways 26

**Drupal**



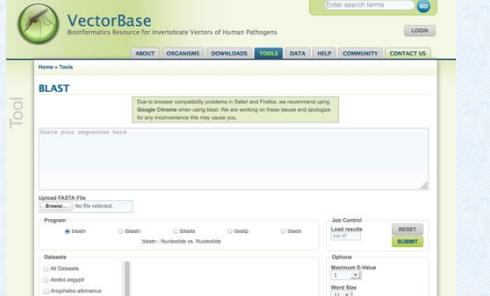
Sandra Gesing Science Gateways 27

**VectorBase - Example for Drupal**



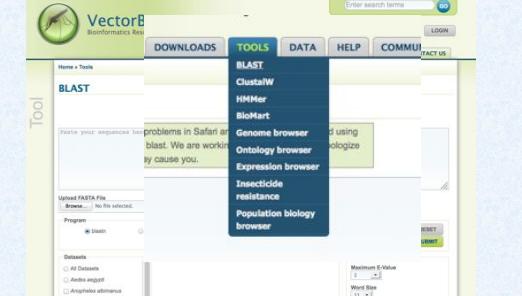
Sandra Gesing Science Gateways 28

**VectorBase**

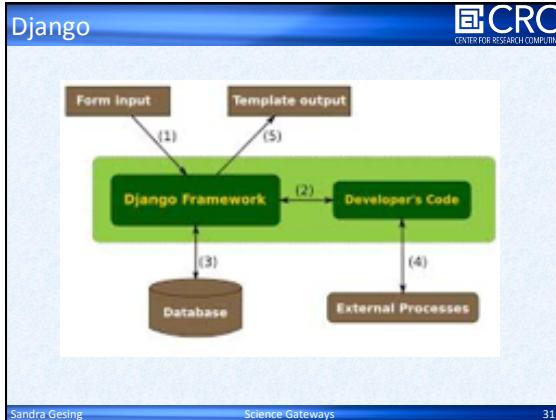


Sandra Gesing Science Gateways 29

**VectorBase**



Sandra Gesing Science Gateways 30



**VecNet – Example for Django**

Sandra Gesing      Science Gateways      32

**VecNet**

Sandra Gesing      Science Gateways      33

**VecNet**

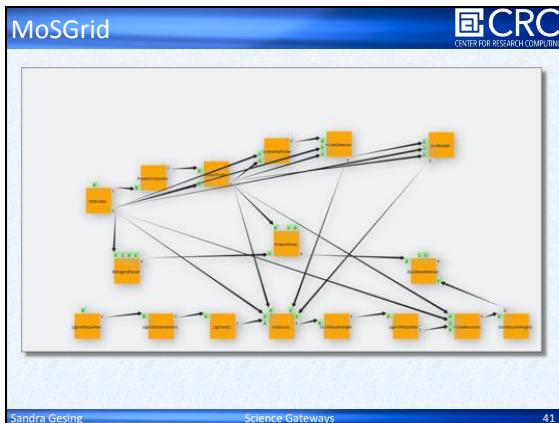
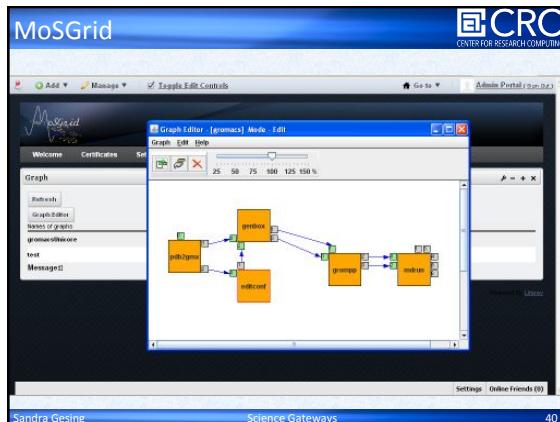
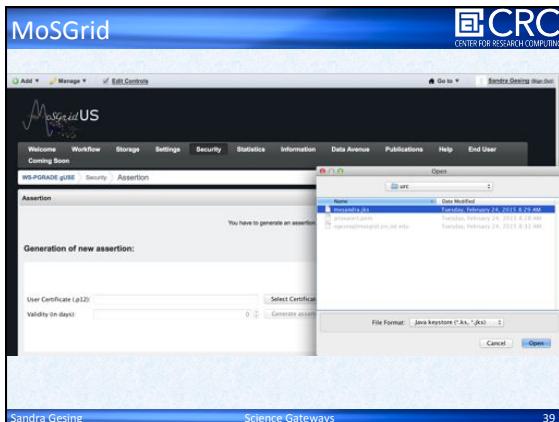
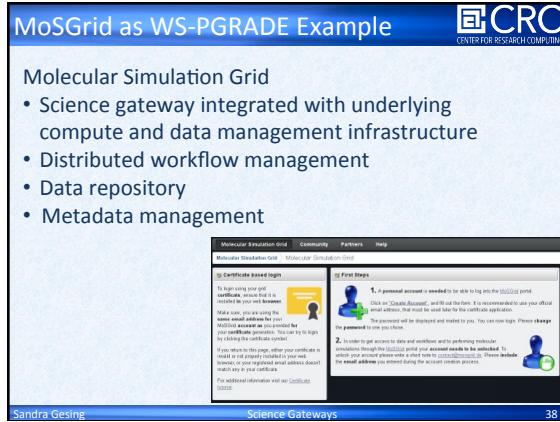
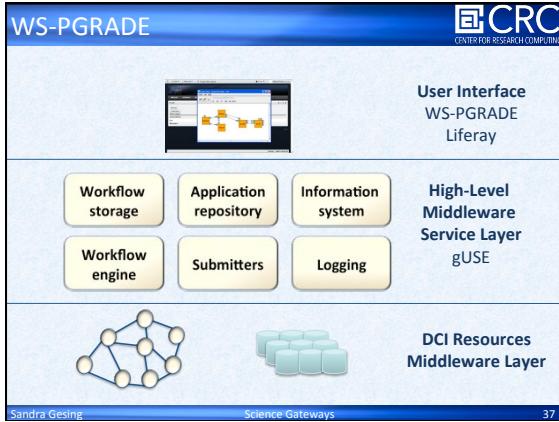
Sandra Gesing      Science Gateways      34

**VecNet**

Sandra Gesing      Science Gateways      35

**Liferay**

Sandra Gesing      Science Gateways      36



This screenshot shows a job configuration window titled "Job's name: PartonFitter". It includes sections for "Optional note:", "Job Executable", "URI: VO", "DURSGU", and "History". A "Put Number" section contains fields for "Put Number ID", "Put Name", "service\_id", and "Description of Job". Below this is a "Source of input directed to this port" section with a file selection dialog showing "ltreents.litesigenparser.jar". A "Parametric input details" section follows. At the bottom are "Put Number ID: Put Name: status" and "Description of Put".

Sandra Gesing Science Gateways 43

This screenshot shows a list of workflows under "WS-PGRADE gUSE > Workflow > Concrete". The table has columns for "Workflow name", "Status" (Running, Finished, Error, Suspended), and "Actions". The workflows listed are: NWChem-specWF\_2015-03-05-092900 (2014-4-9), NWChem\_Mull-part\_2015-03-05-092900 (2014-4-9), NWChem\_TD-part\_2015-03-05-092900 (2014-4-9), NWChem\_basic-part\_2015-03-05-092900 (2014-4-9), NWChem\_free-part\_2015-03-05-092900 (2014-4-9), NWChem\_solv-part\_2015-03-05-092900 (2014-4-9), and hello\_2015-02-07-114739 (2015-2-7).

Sandra Gesing Science Gateways 44

This screenshot shows a list of workflows under "MoSGrid Portal > Workflows > Concrete". The table has columns for "Workflow", "Status", "Instances", and "Actions". The workflows listed are: 2011-1-3 14:53, 2011-1-17 12:00, 2011-2-8 9:34, 2011-1-18 8:40, 2011-2-1 14:44, 2011-2-7 18:55, 2011-1-19 9:21, 2011-1-14 10:36, 2011-1-18 10:13, and 2011-2-10 12:56. A "Selected WF instance" dropdown is set to "2011-1-10 12:56". Below the table is a search bar and a link to "View logs".

Sandra Gesing Science Gateways 45

This screenshot shows a section titled "Molecular Dynamics" with the following text: "• Study and simulation of molecular motion" and "Quantum Chemistry". It also lists: "• Study and simulation of molecular electronic behavior relative to their chemical reactivity" and "Docking". Below this is a list: "• Main focus on evaluation of ligand-receptor interactions (e.g., for drug design)".

Sandra Gesing Science Gateways 46

This screenshot shows a list of features under "MoSGrid - Metadata":

- 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 (Apache Lucene)

Sandra Gesing Science Gateways 47

This screenshot shows a diagram illustrating the MSML processing flow:

```

graph TD
    PD[Program Specific Dictionary] --> Adapter[Adapter]
    CD[Common Dictionary] --> Convention[Convention]
    Adapter --> StructureParser[Structure Parser]
    Convention --> MSMLDocument[MSML Document]
    MSMLDocument --> Parser[Parser]
    Parser --> PP[Parser Parameters]
    PP --> PSO[Program Specific Output]
    StructureParser --> CSInput[Chemical Structure Input]
    CSInput --> PSO
  
```

Below the diagram is an XML snippet:

```

<?xml version="1.0" encoding="UTF-8"?>
<ml convention="convention:comphen">
  <module dictRef="comphen:$objList">
    <module dictRef="comphen:$obj">
      <module dictRef="comphen:$job" id="obj">
        <propertyList/>
        <module dictRef="comphen:initialization">
          <parameterList/>
          <module dictRef="comphen:configuration"/>
          <module dictRef="comphen:finalization">
            <propertyList/>
          </module>
        <module dictRef="comphen:$job" id="obj2">
          ...
        </module>
      </module>
    </module>
  </module>
</ml>
  
```

Sandra Gesing Science Gateways 48

**MoSGrid - Metadata**

The screenshot shows the MoSGrid Metadata interface. On the left, there is a code editor displaying XML configuration for a workflow. On the right, there is a form titled "Import a workflow" with dropdown menus for "ToolSuite" (Gaussian 09) and "Workflow" (Optimization with DFT methods). Below the form, there is a list of submission options.

```

<module distid="g09:qc:initialization">
  <parameters>
    <parameter distid="g09:qc:initialization">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">opt</scalar>
    </parameter>
    <parameter distid="g09:qc:initialization">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">opt+disp</scalar>
    </parameter>
    <parameter distid="g09:qc:theory">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">hf</scalar>
    </parameter>
    <parameter distid="g09:qc:theory">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">hf+disp</scalar>
    </parameter>
    <parameter distid="g09:basisset">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">6-31G</scalar>
    </parameter>
    <parameter distid="g09:basisset">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">6-31G</scalar>
    </parameter>
    <parameter distid="g09:formal_charge">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">true</scalar>
    </parameter>
    <parameter distid="g09:integer">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">true</scalar>
    </parameter>
    <parameter distid="g09:rep3">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">true</scalar>
    </parameter>
    <parameter distid="g09:checkpointfile">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">true</scalar>
    </parameter>
    <parameter distid="g09:checkpointfile">
      <scalar datatype="grid:string" unit="a|nm|eV|rci">true</scalar>
    </parameter>
  </parameters>
  <comgridadapterConfiguration adapterID="g09:adap" clientExecution="true" port="g09_qm7">
    </comgridadapterConfiguration>
</module>

```

Sandra Gesing Science Gateways 49

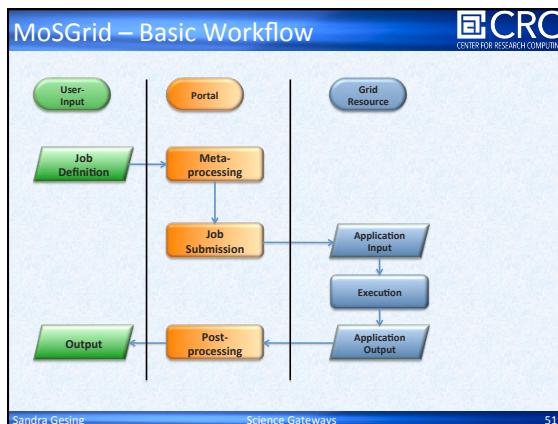
**MoSGrid – Visualization**

The screenshot shows the MoSGrid Visualization interface. It features two 3D molecular models: a DNA molecule on the left and a protein molecule on the right, both visualized in a 3D space with various colored atoms and bonds.

Testing of ChemDoodle and MolCAD

web.chemdoodle.com molcad.de

Sandra Gesing Science Gateways 50



**MoSGrid – QC Portlet**

The screenshot shows the MoSGrid QC Portlet. It has a form for importing a workflow, with dropdowns for tool suite (Gaussian 09), workflow (GOMM/initial), and name (T1Seq\_26). There is also an "Import" button.

- Specialised interface for quantum chemistry software (Gaussian, NWChem, ORCA)
- Basic workflows
- Easy Generation or Uploading of Input Files
- Parsing of result files

Sandra Gesing Science Gateways 52

**MoSGrid - Metaworkflows**

The diagram shows a metaworkflow structure. At the top, a "Job Creator" leads to an "Opt WF" (Opt Input, QM code, Output file). This workflow then branches into two parallel "Frequency Input" paths: "Frequency Input OK" and "Frequency Input 400K". Each path has its own "QM code" and "Frequency Output" steps, resulting in "Freq-OK WF" and "Freq-400K WF" respectively.

- Fundamental step: optimisation
- Frequency calculation at 0 K and at 400 K (polymerisation temperature) for the same molecule
- Small workflow with 3 atomic workflows (opt and freq)

Sandra Gesing Science Gateways

**MoSGrid - Metaworkflows**

The diagram shows a more complex metaworkflow. It consists of four parallel "Job Creator" blocks, each leading to a different "WF" input (WF1, WF2, WF3, WF4). Each "WF" input leads to a "QM code" and "Output file" step. The outputs then lead to four parallel "Opt plus 2freq" blocks (WF1, WF2, WF3, WF4), each with its own "QM code" and "Output file" steps. Finally, all four "Opt plus 2freq" blocks converge into a "Parser" block, which generates a "Table with energies".

Sandra Gesing Science Gateways 54

MoSGrid - Metaworkflows

mosgrid.crc.nd.edu 8080/thredds/portal... 6.1.0 [web/guest] /concrete?\_tAuth=37cWuOqAp\_pL\_id=wefConfig\_Wall\_...

Search

WS-PGRADE glueBE Workflow > Concrete

Workflow name:	INNChem-specWF_2015-03-05-092000
Note:	2014-4-9
Workflow Graph:	metaWF_2015-03-05-092000
Workflow Template:	<a href="#">Edit</a> <a href="#">Fit copied workflow to a new graph</a>

```

graph TD
    Inn[Inn] --> key1[key]
    Inn --> key2[key]
    Inn --> key3[key]
    Inn --> key4[key]
    
```

The diagram illustrates the MoGrid – Multiple Layers architecture, showing five layers of workflow abstraction:

- First layer: atomic workflow**
- Second layer: metaworkflow**
- Third layer: meta-metaworkflow**
- Fourth layer: (meta)<sup>3</sup>-workflow**
- Fifth layer: (meta)<sup>4</sup>-workflow**

Each layer interacts with a central **Experimental Structure**. The layers are represented by colored rectangles (blue, green, yellow, orange, red) and contain various chemical structures and reaction schemes, such as  $[Cu(DMEGcu)]^+$  +  $[Cu(TMMe_2Cu)]^{2+}$  and  $CuH_2$ .

The figure shows a screenshot of the MoSGrid - MD Portlet interface. On the left, there is a 3D visualization of a protein-ligand complex. On the right, there is a plot titled "Gromacs Energies" showing the potential energy (kJ/mol) versus Time (ns). The energy starts at approximately -40000 kJ/mol and decreases rapidly towards -45000 kJ/mol over 200 ns. Below the plot, there are checkboxes for "Calculate By Chain" and "Calculate By Residue". At the bottom, there is a "Size" section with a "Canvas Size" slider set to 910px.

# MoSGrid – Docking Portlet

**Docking Portlet**

Import Submission Monitoring About

Select an imported instance

Import  
StandardDockingWorkflow\_2012-03-30-155439\_291

Please fill the input mask to submit your workflow

**POBOuter**

From file (choose)

Upload POB

POB Model:  
Model 0

Chain A  
 Chain B

Select a protein chain from your POB input file to act as receptor (secondary structure) including the binding pocket (orange).

Chain name of ligand:  
A

Name of ligand as stated in pdb file:  
GNT

Protein Chains that are to be defined

Specify a reference ligand (green) by its three letter code including the corresponding chain. It might be necessary to specify the chain ID if there are multiple chains. This is required for the identification of the binding site and the calculation of RMSD values.



Sandra Gesing

Science Gateways

59

The screenshot shows the MoSGrid Docking Portlet interface. At the top, there's a navigation bar with links like Home, Help, Log In, and Log Out. Below it is a search bar with placeholder text "Search for..." and a dropdown menu for "Search by". The main area has tabs for Docking Portlet, Import, Monitor, and Debug. Under the Docking Portlet tab, there's a "Select an i" dropdown with options: RST201101025, RST201101026, RVE, and Docking. A "Results" section shows a list of 15 entries, each with a checkbox and a "Details" link. The first entry is highlighted with a blue box. To the right, a large molecular visualization window displays a protein structure (blue sticks) bound to a ligand (yellow sticks). The ligand is shown in two states: magenta and cyan. A legend at the bottom right identifies these colors. The visualization includes a 3D coordinate system (x, y, z) and numerical values.

The screenshot shows the Galaxy web interface with a workflow titled "WWSMD?". The workflow consists of several steps: "grow modify appendages.", "format test database?", "Test.fasta", "format test database?", "Test.fasta", "format test database?", "Test.fasta", and "format test database?", "Test.fasta". The "Test.fasta" step is highlighted in green, indicating it is the current active step. The interface includes a sidebar with various analysis tools and a main workspace for managing datasets and visualizations.

Sandra Gesing

Science Gateways

61

The screenshot shows the Galaxy web interface with a tool configuration page for "Compute sequence length (version 1.0.0)". The page displays the command "Compute length for these sequences: 2: http://bx.psu.edu/~clmentz/elements/CMOD0203/m.vannelli.sequence.fasta" and a dropdown menu for "How many title characters to keep?". Below the form, there is a section titled "What it does" with a detailed description of the tool's function and an example of FASTA formatted sequences.

Sandra Gesing

Science Gateways

62

The screenshot shows the RNA-Seq Analysis interface on the Galaxy platform. It features a "RNA-Rocket" tool for RNA-seq analysis. The interface includes sections for "Uploads", "Quality Control" (with options like "How good are my base calls?", "Trim low quality sequence", and "Are all my reads mapped?"), and a "RNA-Seq Analysis" section with a "Map Reads & Assess QC Metrics" button. The interface also includes a "Launch Job" button and a "Project View" section.

Sandra Gesing

Science Gateways

63

### Apache Airavata

The diagram illustrates the Apache Airavata architecture. It shows the following components and their interactions:

- External Clients:** Admin Interfaces (Science Gateway Portals and Desktop User Interfaces), XBaya, and Grid Services API.
- Internal Components:** Airavata API Server, Orchestrator, Application Factory, Registry, Workflow Engine, Management System, Real-Time Monitoring, and Job Monitor.
- Computational Resources:** Application Factory (DFaC) and Computational Resources.
- Interactions:**
  - External clients interact with the Airavata API Server.
  - The Airavata API Server interacts with the Orchestrator, Application Factory, Registry, and Management System.
  - The Orchestrator interacts with the Application Factory, Registry, Workflow Engine, and Job Monitor.
  - The Application Factory interacts with the Registry, Workflow Engine, and Application Experiment.
  - The Registry interacts with the Application Experiment and Job Monitor.
  - The Workflow Engine interacts with the Application Experiment.
  - The Job Monitor interacts with the Application Experiment.
  - The Management System interacts with the Registry and the Orchestrator.
  - The Real-Time Monitoring system interacts with the Management System.

Sandra Gesing

Science Gateways

64

This detailed diagram provides a deeper look at the Apache Airavata architecture:

- Core Components:** Admin Interfaces, XBaya, Grid Services API, Application Experiment, Application Factory, Registry, Orchestrator, Workflow Engine, Job Monitor, and Management System.
- Interactions:**
  - Admin Interfaces, XBaya, and Grid Services API interact with the Airavata API Server.
  - The Airavata API Server interacts with the Orchestrator, Application Factory, Registry, and Management System.
  - The Orchestrator interacts with the Application Factory, Registry, Workflow Engine, and Job Monitor.
  - The Application Factory interacts with the Registry, Workflow Engine, and Application Experiment.
  - The Registry interacts with the Application Experiment and Job Monitor.
  - The Workflow Engine interacts with the Application Experiment.
  - The Job Monitor interacts with the Application Experiment.
  - The Management System interacts with the Registry and the Orchestrator.
  - The Real-Time Monitoring system interacts with the Management System.

Sandra Gesing

Science Gateways

65

### Apache Airavata

The diagram illustrates the Apache Airavata architecture. It shows the following components and their interactions:

- External Clients:** Admin Interfaces (Science Gateway Portals and Desktop User Interfaces), XBaya, and Grid Services API.
- Internal Components:** Airavata API Server, Orchestrator, Application Factory, Registry, Workflow Engine, Management System, Real-Time Monitoring, and Job Monitor.
- Computational Resources:** Application Factory (DFaC) and Computational Resources.
- Interactions:**
  - External clients interact with the Airavata API Server (based on Apache Thrift).
  - Internally, components interact through Component Programming Interfaces (thrift-based CPIs).
  - The Airavata API Server interacts with the Orchestrator, Application Factory, Registry, and Management System.
  - The Orchestrator interacts with the Application Factory, Registry, Workflow Engine, and Job Monitor.
  - The Application Factory interacts with the Registry, Workflow Engine, and Application Experiment.
  - The Registry interacts with the Application Experiment and Job Monitor.
  - The Workflow Engine interacts with the Application Experiment.
  - The Job Monitor interacts with the Application Experiment.
  - The Management System interacts with the Registry and the Orchestrator.
  - The Real-Time Monitoring system interacts with the Management System.

Sandra Gesing

Science Gateways

66

**Apache Airavata**

**CRC**  
CENTER FOR RESEARCH COMPUTING

Experiment Definition:

```

1: required string experimentId = DEFAULT_ID;
2: required string projectName = DEFAULT_PROJECT_NAME;
3: optional dde::Timestamp, defaultTimestamp;
4: optional string description;
5: required string name;
6: optional string type;
7: optional string applicationId;
8: optional string workflowDefinition;
9: optional string workflowImplementation;
10: optional UserConfiguration userConfiguration;
11: optional List<Experiment> experiments;
12: optional List<Experiment> experimentDetails;
13: optional List<Experiment> experimentErrors;
14: optional List<Experiment> experimentDetailsList;
15: optional List<Experiment> experimentErrorsList;
16: optional List<Experiment> errors;
    
```

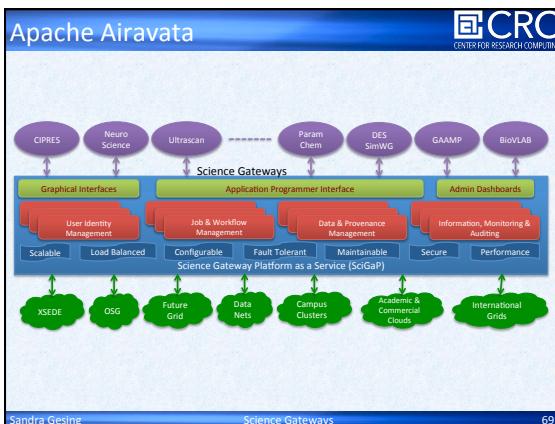
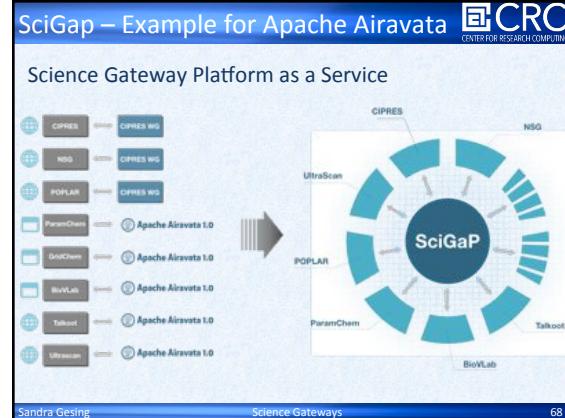
Defining a struct

string createExperiment(); required Experiment(experimentId,Experiment experiment); throw (AiravataErrors,IllegalRequestException, etc);

Defining a service method

**Clean way to define IDLs with richer data structures**

Sandra Gesing Science Gateways 67



**Agave API**

**CRC**  
CENTER FOR RESEARCH COMPUTING

Agave is a *Science-as-a-Service* web API platform

**Run scientific codes**

- your own or community provided codes
- ...on HPC, HTC, or cloud resources
- your own, shared, or commercial systems
- ...and manage your data
- reliable, multi-protocol, async data movement
- ...from the web
- webhooks, rest, json, cors, oauth2
- ...and remember how you did it
- deep provenance, history, and reproducibility built in

Sandra Gesing Science Gateways 71

**Agave API**

**CRC**  
CENTER FOR RESEARCH COMPUTING

<ul style="list-style-type: none"> <li>• Multitenant</li> <li>• Hosted identity management</li> <li>• Supports multiple IdP</li> <li>• OAuth2/OIDC server</li> <li>• API Management</li> <li>• Hosted or on premise</li> </ul>	<ul style="list-style-type: none"> <li>Vertical SSO</li> <li>Analytics and reporting</li> <li>Developer resources</li> <li>Multiple SDK &amp; CLI</li> <li>Reference gateway</li> <li>White labeled</li> <li>100% open source</li> </ul>
--	--

Sandra Gesing Science Gateways 72

**Agave API**

Used to power web & mobile applications

Sandra Gesing Science Gateways 73

**Agave API**

Used to extend existing processes

Sandra Gesing Science Gateways 74

**Agave API**

(Re)Introducing the Micro App Paradigm

AGAVE + DOCKER  
Agave uses Docker container technology to safely and securely run your code on HPC, HTC, Cloud and your local resources.

Sandra Gesing Science Gateways 75

**Agave API**

Agave Delivers Process-as-a-Service

Sandra Gesing Science Gateways 76

**Agave API**

**PRODUCTS**

- Atmosphere
- Discovery Environment
- DNA Subway
- Bisque
- Third Party Consumers
- Data Commons

**FOUNDATIONAL SERVICES**

- Data Store
- Science APIs
- Auth
- ElasticSearch

**LOW-LEVEL SERVICES, SECURITY, ASSETS, etc.**

- iRODS (federated storage)
- CAS/Shibboleth/OAuth2 (single sign-on)
- OpenStack (virtualization)
- Condor (job scheduling)
- XSEDE (national CI)

**HARDWARE RESOURCES**

- Cloud Systems
- High-performance Computers
- Databases
- Storage

Sandra Gesing Science Gateways 77

**iPlant – Example for Agave API**

3. Results:

Sandra Gesing Science Gateways 78

**Agave API - Tutorials**

**Advanced Tutorials**

Dive deeper into the Agave REST APIs with these advanced tutorials on the individual APIs.

- Client Registration**: Learn how to register your client applications and obtain API keys.
- Authorization**: Learn how about authentication and authorization in Agave.
- App Management**: Wrap your existing scientific applications and expose them for execution through the API.
- Agave + Docker (TODO)**: Learn how to use Docker and Agave to conduct portable, reproducible science.

Sandra Gesing Science Gateways 79

**Agave API - Tutorials**

System Management	System Monitoring (TODO)
Learn how to create your own HPC, ITO, Cloud, and Big Data resources with Agave.	Learn how to monitor system uptime and availability with Agave.
Job Management (TODO)	Data Management
Learn how to run applications, monitor jobs, and archive data in Agave.	Learn how to manage, move, and share your data with others in this tutorial.
User Management	Using Posts (TODO)
Learn how to use Agave profiles, manage users in your organization and interact with other users in your organization.	Learn how to create deposits, publications, and files that can be shared with anyone.
Metadata Management	Notifications and Events
Learn how to view, validate, and manage metadata in Agave.	Learn about Agave's event system and how to get real time notifications about any event, any time.

Sandra Gesing Science Gateways 80

**Collaboration on Science Gateways**

**Crucial Topics**

- Close collaboration with user communities
- Knowledge about available technical solutions

**Sounds easy but...**

- Requirements of user communities often not so clear
- Technologies sometimes still under development for certain building blocks
- ➔ Slow uptake of solutions
- ➔ Larger effort for creating science gateways

Sandra Gesing Science Gateways 81

**Science Gateways**

**A new era...**

- Novel developments of web-based agile frameworks
- Infrastructure providers report that science gateways are more used than commandlines

Period	New Registrations	Sessions	New Projects
10/10/2012	1,848	10,800	7,444
10/11/2012	2,031	10,710	11,462
10/12/2012	2,211	10,710	14,294
10/13/2012	3,201	11,100	20,372
10/14/2012	4,107	41,197	-

<http://www.iplantcollaborative.org>

Sandra Gesing Science Gateways 82

**Science Gateways**

**A new era...**

- Novel developments of web-based agile frameworks
- Infrastructure providers report that science gateways are more used than commandlines

**But also always new challenges...**

- Novel infrastructures
- Novel data sources like telescopes such as the Square Kilometre Array (SKA) (will create data rates in exa-scale size)

➔ Support of developers necessary

Sandra Gesing Science Gateways 83

**Science Gateway Institute**

**2012 NSF Software Institute conceptualization award**  
**2015 NSF Software Institute implementation proposal (\$15M)**

**Services**

- Incubator
- Developer support team
- Gateway framework directory
- Workforce development

<http://sciencegateways.org>

Sandra Gesing Science Gateways 84

**Science Gateway Survey 2014**

**CRC**  
CENTER FOR RESEARCH COMPUTING

- 29,000-person survey
- 4957 responses from across domains

Discipline	Percentage
Physical and Mathematical Sciences	30%
Life Sciences	22%
No Response	16%
Computer and Information Sciences	16%
Engineering	16%
Environmental Sciences	14%
Social Sciences	10%
Professional Disciplines	9%
Arts and Humanities	2%

Sandra Gesing Science Gateways 85

**Science Gateway Survey 2014**

**CRC**  
CENTER FOR RESEARCH COMPUTING

- 29,000-person survey
- 4957 responses from across domains

Discipline	Percentage
Physical and Mathematical Sciences	30%
Life Sciences	22%
No Response	16%
Computer and Information Sciences	16%
Engineering	16%
Environmental Sciences	14%
Social Sciences	10%
Professional Disciplines	9%
Arts and Humanities	2%

Sandra Gesing Science Gateways 86

**Science Gateway Survey 2014**

**CRC**  
CENTER FOR RESEARCH COMPUTING

What services would be helpful?

Proposed Service	% Interest
Evaluation, impact analysis, website analytics	72%
Adapting technologies	67%
Web/visual/graphic design	67%
Choosing technologies	66%
Usability Services	66%
Visualization	65%
Developing open-source software	64%
Support for education	64%
Community engagement mechanisms	62%
Keeping your project running	62%
Legal perspectives	61%
Managing data	60%
Computational resources	59%
Mobile technology	59%
Database structure, optimization, and query expertise	59%
Data mining and analysis	58%
Cybersecurity consultation	57%
Website construction	57%
Help with grant writing, process consultation	53%
Source code review and/or audit	51%
High-bandwidth networks	48%
Scientific instruments or data streams	44%
Management aspects of a project	38%

Sandra Gesing Science Gateways 87

**New Science Gateways - Checklist**

**CRC**  
CENTER FOR RESEARCH COMPUTING

**DISCUSSION**

Organizational Aspects  
Technical Aspects  
Domain-Specific Aspects

**Developers** **Domain Experts**

Sandra Gesing Science Gateways 88

**New Science Gateways - Checklist**

**CRC**  
CENTER FOR RESEARCH COMPUTING

Domain-specific aspects:

- Goal, target area and target users
- Visions/demands on the layout
- Priorities of features and options, e.g., a list from must-have to great-to-have options
- Integration of existing applications or development of applications
- Technologies of the applications
- Visualization
- Security demands
- Workflows

Sandra Gesing Science Gateways 89

**New Science Gateways - Checklist**

**CRC**  
CENTER FOR RESEARCH COMPUTING

Organizational aspects:

- Time constraints for the development, agreement on a (maybe even rough) project plan with milestones
- Agreement on alpha- or beta-tester
- Regular meetings

Sandra Gesing Science Gateways 90

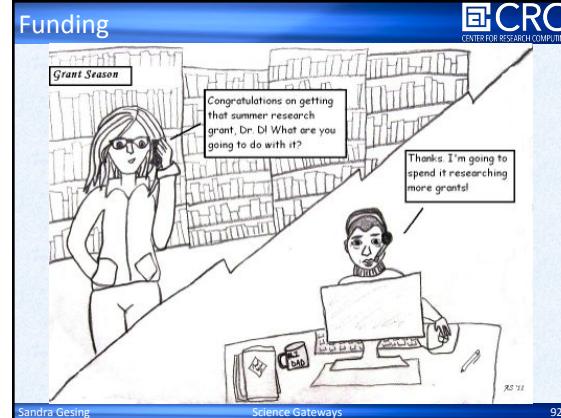
**New Science Gateways - Checklist**

**CRC**  
CENTER FOR RESEARCH COMPUTING

**Technical aspects:**

- Experience with existing frameworks and programming languages
- Available infrastructure including security infrastructure and resources
- Available support of suitable technologies
- Scalability of suitable technologies
- Effort for extending existing technologies compared to novel developments
- Synergy effects with other science gateway projects

Sandra Gesing      Science Gateways      91



**EU COST Action cHiPSet (IC1406)**

**CRC**  
CENTER FOR RESEARCH COMPUTING

**cHiPSet – High Performance Modeling and Simulation for Big Data Applications**

- April 2015 – April 2019
- 15 countries - 12 COST, 3 non-COST (US, China, Australia)
- 37 research organizations/companies (31 COST, 6 non-COST)

[http://www.cost.eu/COST\\_Actions/ict/Actions/IC1406](http://www.cost.eu/COST_Actions/ict/Actions/IC1406)

Sandra Gesing      Science Gateways      93

**EU COST Action cHiPSet**

**CRC**  
CENTER FOR RESEARCH COMPUTING

**Structure & Coordinate Research on HPC-enabled Modelling and Simulation (MS) for Big Data in Europe.**

oc-2014-1-18820      06/14 - ICI Open Call Hearing, 25.09.2014, Larnaca, Cyprus

Sandra Gesing      Science Gateways      94

**EU COST Action cHiPSet**

**CRC**  
CENTER FOR RESEARCH COMPUTING

**Working Groups (WGs)**

WG1: Big Data Infrastructures    WG2: Parallel Programming Models    WG3: MS for Life Sciences    WG4: MS for Socio-Economic, Physical Sciences

Survey: state-of-the-art, industrial solutions, open problems

T+12: Technology requirements and enhancements

T+24: WG & industry coordination (interchanging & storage formats,...)

T+36: Novel approaches to data intensive MS Life, Socio-Economic, Physical Sciences

T+48: Finalisation of integrated frameworks for HPC-enabled MS

BigData

BigData

BigData

BigData

oc-2014-1-18820      COST - ICI Open Call Hearing, 25.09.2014, Larnaca, Cyprus

Sandra Gesing      Science Gateways      95

**cHiPSet - Collaboration**

**CRC**  
CENTER FOR RESEARCH COMPUTING

Projects already declared interest for collaboration

- NESUS (Network for Sustainable Ultrascale Computing)
- KEYSTONE (Semantic keyword-based search on structured data sources)
- AAPEL (Algorithms, Architectures and Platforms for Enhanced Living Environment)

And maybe YOU?

Sandra Gesing      Science Gateways      96

## Acknowledgements



### Co-authors

Rion Dooley  
 Richard Grunzke  
 Sonja Herres-Pawlis  
 Alexander Hoffmann  
 Jens Krüger  
 Marlon Pierce

**WS-PGRADE**  
 Akos Balasko  
 Peter Kacsuk  
 Istvan Marton  
 Gabor Terstyanszki

**XSEDE, especially**  
 Tom Maiden  
 Suresh Marru  
 Marlon Pierce  
 Tabitha Samuel  
 Raminder Singh  
 Nancy Wilkins-Diehr

Sandra Gesing

Science Gateways

97

## Information on Science Gateways



- IEEE Technical Area on Science Gateways  
<http://ieeesciencegateways.org>
- Science Gateway Institute  
<http://sciencegateways.org>
- XSEDE Science Gateways  
<https://www.xsede.org/gateways-overview>
- Science Gateway Workshops (partner workshops)  
 Europe: IWSG - <http://iwsgr.info/>  
 USA: GCE - <http://sciencegateways.org>  
 Australia: IWSG-A - <http://iwsgr.info>

Sandra Gesing

Science Gateways

98



sandra.gesing@nd.edu

Sandra Gesing

Science Gateways

99