Enhancing Research Data Collaboration Using Science Gateways

Sandra Gesing
Center for Research Computing
sandra.gesing@nd.edu

25 August 2015
State of the Art

Data and compute-intensive problems

Web-based agile frameworks

Distributed data and computing infrastructures

Users generally not IT specialists

Tools and workflow engines

High-speed networks
Challenge for Developers

- Data and compute-intensive problems
- Web-based agile frameworks
- Distributed data and computing infrastructures
- Tools and workflow engines
- Users generally not IT specialists
- High-speed networks
Reusability

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

Frameworks and APIs

Re-inventing is not always necessary...
Frameworks and APIs

... but the model should fit to the demands of the community
Science Gateway Technologies

- Agile web frameworks (AngularJS, Semantic UI)
- Content management systems (Drupal)
- Libraries for implementation (Django)
- Science gateway frameworks (Galaxy, WS-PGRADE, Catania Science Gateway Framework, HubZero)
  - Static layout
  - Layout extendable
  - Workflow-enabled
- APIs for implementation (Apache Airavata, Agave, Vine Toolkit)
Science Gateway Technologies

Sandra Gesing

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
New Science Gateways - Checklist

Organizational Aspects

Technical Aspects

Domain-Specific Aspects

Developers

Domain Experts
New Science Gateways - Checklist

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
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
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
Knowledge Hubs

- Centralized Teams
- Projects
- Science Gateway Institute
- IEEE Technical Area
CRC – Centralized Team

- Software development and profiling
- Cyberinfrastructure/science gateway development
- Geographical Information Systems
- Visualization Support
- Computational Scientist support
- Collaborative research/grant development
- System administration/design and acquisition
- ~40 researchers, research programmers, HPC specialists
Resources for Researchers

**Chalmers**
- Supercomputing Networking
- Human Metabolic Atlas
- Personalized GEMs for liver CA

**Notre Dame**
- HPC Bioinformatics
- Federated Hybrid Clouds
- eScience Portal Development

**PSNC**
- High-Performance Computing
- Pioneer NREN, optical R&D
- Applications – Oncosimulation

**CDAC**
- Computational Genomics
- Biomolecular Simulations
- Anvaya Genomics Platform
Connecting Clouds

CRC - ICTBioMed

Sandra Gesing

Science Gateways
EU COST Action cHiPSet (IC1406)

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
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

WG & industry coordination (interchanging & storage formats...)
- Test-bed model refinement

T+24
- BigData
  - Frameworks for Big Data provision and management
- BigData
  - Novel approaches to data intensive MS Life, Socio-Economic, Physical Sciences

T+36
- Finalisation of integrated frameworks for HPC-enabled MS

T+48
- BigData

COST - ICT Open Call Hearing, 25.09.2014, Larnaca, Cyprus
Projects declared interest for collaboration

- **NESUS (Network for Sustainable Ultrascale Computing)**
  http://www.nesus.eu/

- **KEYSTONE (Semantic keyword-based search on structured data sources)**
  http://www.keystone-cost.eu/

- **AAPELE (Algorithms, Architectures and Platforms for Enhanced Living Environment)**
  http://aapele.eu/

And maybe YOU?
Enabling Global Infrastructures

Data Providers

Research Communities

Community-specific tools

Service Providers

Services
(Storage, HTC, Cloud)

Technology Providers

Technology
(storage, data management, job scheduling and execution, workflow management, Auth and Authz, gateways ..)

Data
(discovery, data management, repositories)

Knowledge
(training, education, technical support)

All
EGI-Engage

- Distributed, federated storage and compute facilities
- Compute platforms (HTC, Cloud)
- Virtual Research Environments
- > 200 user research projects

Performance:
- 350 resource centres in 40 countries
- 400,000 logical CPU cores
- 190 PB disk, 180 PB tape

Community SLAs, Provider OLAs
Indigo DataCloud

Project type: Research & Innovation Action (RIA)

• Duration: 30 months
• Start date: 1st April 2015
• 26 partners from 11 countries

www.indigo-datacloud.eu
Indigo DataCloud

WP6
- APIs and Toolkits
- Science Gateways and UI
- Big data driven workflows for e-Science

WP5
- APIs
  - Application/Service Deployment
  - Storage Access Layer
  - QoS/SLA
  - Capacity Management
  - Elasticity/Scalability
  - High Level app deployment
  - Unified Data access

WP4
- APIs
  - Orchestration
  - Storage
  - Network
  - IaaS Orchestration
  - Cross Protocol Access
  - Network Virtualiz.
  - Container Repository
  - Monitoring & Discovery
  - Advanced Scheduling
  - Template Repository

IaaS
- AAI Management
- App Deployment
- Storage Access
- Network Virtualization
- Authentication

SaaS
- Orchestration of diverse architectures
- Lifecycle Engine
- ESB
- Accounting
- Service Discovery
- End User Monitoring

PaaS
- Elasticity/Scalability Mngt

www.indigo-datacloud.eu
CloudSME

- EU project with 16 partners
- Develops a cloud-based, one-stop-shop solution
- Scalable platform for small or larger scale simulations
- Wider take-up of simulation technologies in manufacturing and engineering SME’s

http://cloudsme.eu/
CloudSME

Simulation and Application Layer

Front end cloud

GUI & frontend tools

Eurobios
Saker Solution
Podoactiva
2MORO

TransAT plug-in
Simul8 plug-in
3D Scan plug in
BFly plug-in

commercial client
commercial components
commercial client
commercial client

PaaS for simulations (based on WS-PGRADE)

DCI Bridge

Java API

REST API

Application patterns & deployments
TransAT
Simul8
3D Scan
BFly
...

Generic cloud access platform (CloudBroker Platform)

Cloud adapters

Private clouds

SME clouds (CloudSigma,...)

Commercial clouds (Amazon,IBM,...)

Academic clouds (UoW, SZTAKI, BIFI,...)

Cloud Platform Layer

Resource Layer

Cloud Sigma

Amazon

IBM

Eucalyptus

Open Stack

Open Nebula

http://cloudsme.eu/
Information on Science Gateways

• Science Gateway Workshops
  in conjunction with eResearch Australia
  Europe: IWSG - http://iwsg.info
  USA: GCE - http://sciencegateways.org

• Science Gateway Institute
  http://sciencegateways.org

• IEEE Technical Area on Science Gateways
  http://ieeesciencegateways.org

• XSEDE Science Gateways
  https://www.xsede.org/gateways-overview

• CRC Science Gateways
  https://crc.nd.edu/index.php/research/gateways

19 October 2015
Brisbane
<table>
<thead>
<tr>
<th>Speakers/Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keynote: Nancy Wilkins-Diehr</td>
<td>Science Gateways: The Importance of Building Community</td>
</tr>
<tr>
<td>Nigel Ward, Glenn Moloney</td>
<td>Reflections on the NeCTAR Virtual Laboratory Program</td>
</tr>
<tr>
<td>Richard O. Sinnott, et al.</td>
<td>The Collaborative Urban Research Environment for Australia</td>
</tr>
<tr>
<td>Wojtek James Goscinski</td>
<td>The Characterisation Virtual Laboratory</td>
</tr>
<tr>
<td>Richard O. Sinnott, Jemie Effendy, Stephan Gloeckner, Anthony Stell</td>
<td>Beyond a Disease Registry: An Integrated Virtual Environment for Adrenal Cancer Research</td>
</tr>
<tr>
<td>Michelle Barker</td>
<td>A Science Gateway for Malaria: Successes and Challenges</td>
</tr>
<tr>
<td>Uwe Rosebrock, et al.</td>
<td>The Marine Virtual Laboratory – ocean modeling made easy.</td>
</tr>
<tr>
<td>Aurel F. Moise, Tim Pugh, Martin Dix, Bertrand Timbal</td>
<td>The Australian Climate and Weather Science Virtual Laboratory (CWSLab)</td>
</tr>
<tr>
<td>David Abramson</td>
<td>Workflow driven Science Gateways</td>
</tr>
<tr>
<td>Sandra Gesing</td>
<td>Developing Science Gateways: Current Solutions and Future Challenges</td>
</tr>
<tr>
<td>Panel Discussion: Chair Rhys Francis</td>
<td>The Future of Science Gateways</td>
</tr>
</tbody>
</table>
THANK YOU

sandra.gesing@nd.edu
Information on Science Gateways

• Science Gateway Workshops
  in conjunction with eResearch Australia
  Europe: IWSG - http://iwsg.info
  USA: GCE - http://sciencegateways.org

• Science Gateway Institute
  http://sciencegateways.org

• IEEE Technical Area on Science Gateways
  http://ieeesciencegateways.org

• XSEDE Science Gateways
  https://www.xsede.org/gateways-overview

• CRC Science Gateways
  https://crc.nd.edu/index.php/research/gateways

19 October 2015
Brisbane