S2I2: Science Gateways Community Institute

Sandra Gesing
sandra.gesing@nd.edu

http://sciencegateways.org/
7 December 2016
What are science gateways? Why are they important?

• Increased complexity of
  – today’s research questions
  – hardware and software
  – skills required

• Greater need for openness and reproducibility
  – Science increasingly driving policy questions

• Opportunity to integrate research with teaching
  – Better workforce preparation

We need end-to-end solutions that provide broad access to advanced resources and allow all to tackle today’s challenging science questions.
What are science gateways? Why are they important?

• Explosion in the quantity, variety and complexity of data (e.g., NGS, SKA)
• Questions can be answered impossible to even ask about 10 years ago
• Costs far reduced (e.g., Human Genome project, 15 years, ~$2 billion; today ~3 days, $1000)
Science Gateways – NSF Definition

Science Gateways (also known as portals and hubs) are themselves synergistic focal points where scientists form growing communities; where digital resources, expensive equipment, and collaboration resources are available to those who would otherwise not have access to them; and where the public can participate in the scientific process, spanning science and engineering research and education. Gateways assemble and integrate some of the most complex components of today’s cyberinfrastructure (CI), making them accessible to a wider spectrum of users through easy-to-use interfaces. They provide researchers with unified human and programmable access to facilities:

- computing resources (e.g., supercomputers, clouds),
- instruments (e.g., telescopes, sensor networks),
- data (e.g., data collections, collaborative spaces),
- software (e.g., simulation, modeling, and analysis capabilities, workflow systems),
- and more,

thus increasing the value of these facilities. They make the interdisciplinary collaborations needed to solve the most complex problems more feasible. They support CI abstractions that allow scalable, dynamic use of diverse CI without demanding detailed and complex technical understanding of CI components, and provide scalable solutions for solving classes of problems, eliminating the need for thousands of individual infrastructure installations.
Significant developments in XSEDE as well Gateway users surpass login users in 2013. Increase in 2015 due to automated user-counting.

Source: David Hart
Gateways enable research, but are not research projects themselves.
Science Gateway Survey 2014

- 29,000-person survey
- 4957 responses from across domains
Science Gateway Survey 2014

What services would be helpful?

<table>
<thead>
<tr>
<th>Proposed Service</th>
<th>% Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation, impact analysis, website analytics</td>
<td>72%</td>
</tr>
<tr>
<td>Adapting technologies</td>
<td>67%</td>
</tr>
<tr>
<td>Web/visual/graphic design</td>
<td>67%</td>
</tr>
<tr>
<td>Choosing technologies</td>
<td>66%</td>
</tr>
<tr>
<td>Usability Services</td>
<td>66%</td>
</tr>
<tr>
<td>Visualization</td>
<td>65%</td>
</tr>
<tr>
<td>Developing open-source software</td>
<td>64%</td>
</tr>
<tr>
<td>Support for education</td>
<td>64%</td>
</tr>
<tr>
<td>Community engagement mechanisms</td>
<td>62%</td>
</tr>
<tr>
<td>Keeping your project running</td>
<td>62%</td>
</tr>
<tr>
<td>Legal perspectives</td>
<td>61%</td>
</tr>
<tr>
<td>Managing data</td>
<td>60%</td>
</tr>
<tr>
<td>Computational resources</td>
<td>59%</td>
</tr>
<tr>
<td>Mobile technology</td>
<td>59%</td>
</tr>
<tr>
<td>Database structure, optimization, and query expertise</td>
<td>59%</td>
</tr>
<tr>
<td>Data mining and analysis</td>
<td>58%</td>
</tr>
<tr>
<td>Cybersecurity consultation</td>
<td>57%</td>
</tr>
<tr>
<td>Website construction</td>
<td>57%</td>
</tr>
<tr>
<td>Software engineering process consultation</td>
<td>53%</td>
</tr>
<tr>
<td>Source code review and/or audit</td>
<td>51%</td>
</tr>
<tr>
<td>High-bandwidth networks</td>
<td>45%</td>
</tr>
<tr>
<td>Scientific instruments or data streams</td>
<td>44%</td>
</tr>
<tr>
<td>Management aspects of a project</td>
<td>38%</td>
</tr>
</tbody>
</table>
57% played some role in gateway creation and these gateways were used for a variety of purposes:

- Educational tools: 18%
- Data collections: 15%
- Data analysis tools, including visualization and mining: 16%
- Computational tools: 16%
- Collaboration tools: 8%
- Frameworks or platforms: 6%
- Workflows: 6%
- Citizen science resources: 5%
- Interfaces to scientific instruments: 4%
- Interfaces to sensor data: 4%
- Other: 2%

n of application types=7,805, by 2,756 creators (out of 2,819); mean=2.8 application types per application creator.
Well-designed gateways require a variety of expertise

- Usability Consultant: 34% wished we had this, 16% yes, we had this
- Graphic Designer: 36% wished we had this, 30% yes, we had this
- Community Liaison/Evangelist: 20% wished we had this, 18% yes, we had this
- Project Manager: 17% wished we had this, 45% yes, we had this
- Professional Software Developer: 31% wished we had this, 44% yes, we had this
- Security Expert: 26% wished we had this, 14% yes, we had this
- Quality Assurance and Testing Expert: 42% wished we had this, 15% yes, we had this

n=2,756 respondents or 98% of application creators
The Science Gateways Community Institute

- Diverse expertise on demand
- Longer term support engagements
- Software and visibility for gateways
- Information exchange in a community environment
- Student opportunities and more stable career paths
## Incubator Service

**Expertise for the gateway lifecycle**

### A Framework for Decision Making

<table>
<thead>
<tr>
<th>Technology Planning</th>
<th>Business Planning</th>
<th>Client Interaction Planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Choosing technologies</td>
<td>• Business model development</td>
<td>• Usability studies</td>
</tr>
<tr>
<td>• Cybersecurity</td>
<td>• Financial planning</td>
<td>• Web/visual/graphic design</td>
</tr>
<tr>
<td>• Software engineering</td>
<td>• Project management</td>
<td>• Impact measurement</td>
</tr>
<tr>
<td>• Interfaces to compute and data</td>
<td>• Software licensing</td>
<td>• Community engagement</td>
</tr>
<tr>
<td></td>
<td>• Staff and sustainability planning</td>
<td>• Support for education</td>
</tr>
</tbody>
</table>

### Specialized Expertise

**Security**
- • Center for Trustworthy Scientific Cyberinfrastructure

**Sustainability**
- • Nancy Maron, creator of the ITHAKA S+R course on Sustaining Digital Resources

**Evaluation & Impact Measurement**
- • Ann Zimmerman Consulting

**Campus Resource Development**

### Network / Cohort Formation

**Common Experiences**
- • Training sessions
- • Group interactions

**Continuing Engagement**
- • Customized structure, content, goals
- • Mentoring
- • Pay It Forward

### An Ongoing Dispassionate Ear
Extended Developer Support

Dedicated SGCI staff work directly with clients to build and enhance gateways

**Focus**
- Front-end development
- Gateways using all types of CI
- Both sides give 2-month to 1-year commitment
- Well-defined engagements with work plans
- Technology agnostic

**Mission**
- Bring new gateways into existence
- Adapt existing gateways to new resources and technologies
- Provide “burst” support to help gateways with smaller issues

**Benefits**
- Reinforce Incubator lessons
- Develop deep understanding of community needs that feed into other Institute areas
- Capture and document support efforts for scalability
- Hands-on opportunities for student participants

**Data**
- Analysis Tools

**Education**
- Instruments

**Collaboration**
- Workflows

**Mission**
- Compute

**Mission**
- Sensors

**Mission**
- Collaboration

**Mission**
- Galaxy
- HUBzero
- AGAVE
- Jupyter
- And more...
Scientific Software Collaborative

Leveraging existing investments in gateway technologies

Give developers a single destination for gateway software, services and resources to easily build, maintain and manage science gateways.

- Create more gateways to advance scientific discovery, by making them easier to build
- Create more researchers using gateways by increasing awareness and number of gateways
- Enable NSF projects to integrate into the software institute and promote their products

End-to-End Solutions

- Serve a diverse set of scientific domains
- Out-of-the-box gateway solution that can be customized
- Based on Docker – executable images that are the skeleton for a secure and functioning gateway
- Portable and reproducible
- Community-contributed

Gateway Discovery

- Open registry
- Promotes use of existing science gateways
- Community-contributed
- Admin approval
- Automated cleanup

Engage Other Areas of Institute

- Support projects leverage Collaborative components
- Framework evolves as a result of gateway engagements
- Community outreach

Software Marketplace for Science Gateways

“Use-what-you-need”

- API integration
- Variety of services
  - Information
  - Security
  - Execution
  - Data
  - Event
  - Accounting
- Hosting opportunity

Software Integration & Community Contribution

- Docking mechanisms for community-contributed software, including NSF SI2
- Incorporate community standards
Community Engagement and Exchange

Key to a successful institute

Gathering place for scientific web developers across NSF directorates, federal agencies, and international boundaries

- Community members are eager to connect

Website Activities
- Discussion forums
- Gateway showcase with case studies
- Symposium series
- News: media coverage, related happenings, academic publications, job openings, events calendar
- Curated blog with guest authors,
- Professional development: synchronous and asynchronous training
- Capture client/user feedback on web and through other areas

Annual Conference
- Tutorials and workshops
- Paper presentations
- Invited keynotes and panels
- Interactive elements: Open Space, poster session
- Travel support for students and campus IT staff

Builds on 10 years of experience with GCE and IWSG series

Outreach to Complementary NSF Initiatives
- NSF SI2 projects
- Large NSF projects
- Science and Technology Centers
- Engineering Research Centers
- MoISSI software institute collaboration

Campus Gateway Groups
- Task force builds campus-based expertise
- Channel for scaling institute services
Gateways 2016 – a success with about 120 participants

Gateways 2016

Help make Gateways 2017 even better! Answer our conference evaluation now.

Did you attend a tutorial? Please evaluate it!

The 11th Gateway Computing Environments Conference (formerly GCE)
Wednesday and Thursday, November 2 & 3, 2016
Hosted at the San Diego Supercomputer Center in San Diego, California

Science gateways serve as connection points, assembling various components of advanced cyberinfrastructure—data collections, instruments, supercomputers, clouds, collaboration capabilities, and analytical tools—behind streamlined, user-friendly interfaces. They are typically a community-developed web portal or a suite of desktop applications. By providing access to top-tier resources, gateways enable researchers with a common goal as well as students and members of the broader community.

Gateway developers and users — regardless of their domain area — have a lot in common but have had few venues for exchanging experiences. The newly expanded Gateways annual conference (formerly the Gateway Computing Environments workshop series) will be an opportunity for gateway creators and enthusiasts to learn, share,

Watch the recorded keynotes and panel from Gateways 2016!
Posted on our YouTube Channel.

http://sciencegateways.org/gateways2016/
Workforce Development: Keep the best and the brightest in the sciences

4 Focal Areas

Providing Financial Support
- Enabling students learning gateway skills
- Including internship experiences

Promoting Gateway-Related Career Paths
- Student-related conference programs

Campus opportunities
- Job boards

Establishing Center for Training and Education at ECSU
- Vigorous schedule of on-site and virtual training
- Development of training and course curricula about science gateways technologies

Integrating Gateways into Course Content
- Providing broader access to high-end resources

Partners

Google Summer of Code (GSoC)
SGCI Institute Areas

National Organization for the Professional Advancement of Black Chemists and Chemical Engineers (NOBCChE)

Molecular Science Software Institute

Association of Computer/ Information Sciences and Engineering Departments at Minority Institutions (ADMI)
Opportunities

• Get advice (sustainability, usability, cybersecurity, other) on your existing gateway
• Work with SGCI to build a gateway for you
• Find a working gateway or gateway development software in our catalog
• Learn how to set up a gateway group on your campus
• Keep up to date on gateway developments
  – webinar series, gateways in the news, google scholar feed, case studies, annual conference, blog posts
• Enroll a student in an internship program to learn gateway development
• Partner with SGCI on your own projects, technologies and events
Challenges for SGCI

• Many researchers/developers are not aware that they need/create “Science Gateways”
It’s a Fan!
It’s a Wall!
It’s a Spear!
It’s a Snake!
It’s a Tree!
It’s a Rope!
It’s a Science Gateway

It’s a Research Portal

It’s a Collaboratory

It’s a Virtual Lab

It’s e-Science eResearch

It’s a Cyberinfrastructure
Challenges for SGCI

• Many researchers/developers are not aware that they need/create “Science Gateways”

• Sustainability beyond funding
Sustainability

• Sustainability plans should include securing expertise, manpower, talent, money

• Problems:
  – Universities/institutes would like to dictate the prices
  – PIs have their “day jobs” - maybe they don’t want to be entrepreneurs and they don’t want to start businesses
  – Research groups don’t have the money to pay or are not willing to pay
  – Software should be open access

• Distinguish between software and services

• Software is free and open access => Communities all over the world can set up their own instances but they also have to maintain it themselves for changes/adaptions and support for user communities

• Services should be chargeable: adaptations, support/maintaining community, usage of resources => a marketplace for services

• Services cannot handle discontinuity, this is bit different for software which is static and inherently has a longer shelf life
Sustainability

• Start as early as possible at the community to ask for money (small amounts) and have price tags from the beginning on services - even without charging - to make the community aware of the cost or show cost of comparable services sold in industry
• Incubator programs
• Achieve policy changes at universities/institutes to support sustainability via demands from funding bodies
• Technology transfer groups supported by state/funding bodies’ individuals
• Using the ecosystem of companies writing STTR grants etc.
• Crowdfunding
• Career paths for developers
• on-campus developer groups
• Involving PhD students/postdocs
Challenges for SGCI

• Many researchers/developers are not aware that they need/create “Science Gateways”

• Sustainability beyond funding

• Attracting and selecting “right” customers
Customers of Training

What you brought to seminar and what it says about you:

- Stuff to take notes: First year. Foolishly thinks he'll ever need notes again.
- Reading material: Third year. Just here for show.
- Didn't bring anything: ABD/Postdoc. Has nothing better to do.
- Laptop: Young Assistant Professor. Working on three proposals at the same time.
- Playing with latest Gadget/Gizmo: Full Professor. Loves new toys.

Jorge Cham © 2008

WWW.PHDCOMICS.COM
Challenges for SGCI

- Many researchers/developers are not aware that they need/create “Science Gateways”
- Sustainability beyond funding
- Attracting and selecting “right” customers
- Attracting partners to contribute to repository
Partners

- Often already existing open-access repository
- A larger audience probably means more customers, which might be too much for a support team

=> Solutions tailored to each partner!
Next Event

HICSS-50 Minitrack on Science Gateways and Portals (January 2017)

The Hawaii International Conference on System Sciences (HICSS-50) will be January 4-7, 2017, at Hilton Waikoloa Village on Hawaii’s Big Island. We are organizing a Science Gateways and Portals minitrack at the conference. The conference provides a unique and highly interactive environment for researchers to exchange perspectives and ideas in various areas of information, computer, and system sciences.

Minitrack Details for ST: Science Gateways and Portals

January 5, 2016 from 4:00-5:30pm in Kohala 3

- “Science Gateways: The Long Road to the Birth of an Institute” by Sandra Gesing, Nancy Wilkins-Diehr, Maytal Dahan, Katherine Lawrence, Michael Zentner, Marion Pierce, Linda Hayden, Suresh Marru
- “Maintaining a Science Gateway – Lessons Learned from MoSGrid” by Lukas Zimmermann, Richard Grunzke, Jens Krüger
- Panel on Science Gateways, including Maytal Dahan (SGCI), Jens Krüger (MoSGrid), Kelly Gaither (Texas Advanced Computing Center), and others

About the “Science Gateways and Portals” Minitrack

This Minitrack is part of the “Software Technology” track.

Science gateways are a community-specific set of tools, applications, and data collections that are integrated together via a web portal or a desktop application, providing access to resources and services for distributed data management and distributed computing infrastructures (DCIs). They offer easy and intuitive access to computing infrastructures and instruments irrespective of their location. The challenges in the area of gateways are manifold: from intuitive user interfaces and security features through efficient data and workflow management to parallelization of applications employing parallel and distributed architectures. In the last decade, quite a few
Team

Maytal Dahan
Maytal joined the Texas Advanced Computing Center (TACC) at The University of Texas at Austin in August 2002 as a Software Developer. She is involved in several software projects that develop middleware and web portal applications. Maytal is the project lead for the XSEDE User portal and managed the User Information and Interfaces in XSEDE. Maytal was Co-PI [...] 

Sandra Gesing
Sandra Gesing is a research assistant professor and computational scientist at the University of Notre Dame. Previously, she worked as a research associate at the University of Edinburgh, UK, and at the University of Tübingen, Germany, where she also received her PhD. Additionally, she has extensive experience working as a project manager and developer in industry. Her [...] 

Linda Hayden
Dr. Hayden holds a PhD in Mathematics and Education. She is presently the Associate Dean of the School of Mathematics, Science and Technology at Elizabeth City State University, and the Director of the Center of Excellence in Remote Sensing Education and Research (CERSER). The CERSER program works in partnership with federal agencies, other universities and [...] 

Katherine Lawrence
Katherine Lawrence is an independent scholar affiliated with the University of Michigan School of Information. She collaborates with individuals and groups as a researcher, writer, educator, and facilitator. Her most recent research project was “Opening Science Gateways to Future Success,” funded by the National Science Foundation’s Office of Cyberinfrastructure. The goal of this project was [...] 

Suresh Marru
Muru leads the science gateway program in the XSEDE project. He is a Co-Principal Investigator on the NSF funded Open Gateway Computing Environments and leads the service oriented architectures and scientific workflow efforts. Muru is deeply involved in various science Gateway efforts including GridChem, UtterScan, BioVLAB and OILAS projects. Earlier he was instrumental in success of the [...] 

Stephen Mock
Stephen Mock is director of Advanced Computing Interfaces (ACI) at the Texas Advanced Computing Center (TACC). Steve has been involved in developing web-based science gateways, portals, middleware, and web service APIs since 2000 for use in the high performance computing arena. He has worked on the iPlant Project, Agave, the XSEDE User Portal, the TACC [...] 

Marlon Pierce
I'm the assistant director at the Community Grids Lab at Indiana University. I lead projects in Grid Web Portal and Web Service applications development to support Geographical Information Systems, earthquake modeling applications, and chemical informatics applications. And some other stuff. You may have noticed I got a promotion—maybe the blog helped. My full and short [...] 

Nancy Wilkins-Diehr
Nancy Wilkins-Diehr is Associate Director at SOSIC and co-director of XSEDE’s Extended Collaborative Support program. She has been involved in science gateways and their interfaces to high-performance computing since 2005. Nancy received her Bachelor’s degree from Boston College in Mathematics and Philosophy and her Master’s degree in Aerospace Engineering from San Diego State University. She has [...] 

Michael Zentner
Michael Zentner is a Senior Research Scientist at the Rosen Center for Advanced Computing (RCAC) at Purdue University working on visualization and data analysis. Prior to joining Purdue, he was the founder and senior team member of several information technology companies, where he created innovative solutions for extracting patterns from data, collaboration, and constrained optimization. Dr. Zentner has [...] 

and more!
Acknowledgements

• The whole team of SGCI
• NSF award id 1547611
Thanks for your attention!

http://sciencegateways.org/
7 December 2016