

# Science Gateways and the Importance of Sustainability

Nancy Wilkins-Diehr  
San Diego Supercomputer Center  
wilkinsn@sdsc.edu

Katherine Lawrence  
University of Michigan  
kathla@umich.edu

Linda Hayden  
Elizabeth City State University  
haydenl@mindspring.com

Marlon Pierce, Suresh Marru  
Indiana University  
{marpierc, smarru}@iu.edu

Michael McLennan, Michael  
Zentner  
Purdue University  
{mclenna, mzentner}@purdue.edu

Rion Dooley, Dan Stanzone  
Texas Advanced Computing Center  
{dooley, dan}@tacc.utexas.edu

The Web has a major impact on modern life. We do our banking, make travel arrangements, research health topics, go shopping and connect with friends and family via the Web.

This fundamental impact extends to the scientific realm as well. Modern science now depends on the Web. Truly impactful websites are created by scientists all the time. Some are developed to fulfill the needs of small research teams. Others are built to address the needs of a large community. Most are completely open and publicly accessible. Increasingly, they are accessible via mobile devices. We call these Web and mobile interfaces science gateways.

Formally, a science gateway is a community-developed set of tools, applications, and data collections that are integrated through a portal or a suite of applications [Wilkins-Diehr 2007]. Science gateways enable entire communities of users with a common scientific goal to use digital resources through a single interface, even when such resources are geographically distributed. The digital resources in this context could be anything from a highly tuned parallel application running on a supercomputer to a catalogued and cross-referenced data collection with built in analysis capabilities to a forum for sharing and rating educational course content and user-contributed analysis tools. Science gateways provide value-added interfaces to access these shared resources.

Science gateways can have varying goals and implementations. Some expose specific sets of community codes so that anonymous scientists can run them. Others may serve as a “metaportal,” a community portal that brings a broad range of new services and applications to the community. A common trait of all types is their interaction with

back-end digital services to provide value-added capabilities to the end user.

Gateways can be viewed as a specific type of software and therefore susceptible to all the software ecosystem challenges described in the WSSSPE call for participation. In fact, many times these challenges are even exacerbated for projects with a presence on the Web. It is often difficult to anticipate the growth user communities. Once there is that public exposure, projects that were planned to serve individual groups can “go viral” and become very valuable to a large community. This can come as a surprise to the developers. If a gateway doesn’t require a user to log in, it can be difficult to ascertain how many scientists are relying on it until it is decommissioned and the user community protests. Gateways can also have many moving parts with many dependencies for the infrastructure to which they provide access. Each of those parts can have its own sustainability challenges. Gateways are part of a nimble, dynamic, and ever-evolving ecosystem.

The TeraGrid Science Gateway program began in 2004 when TeraGrid directors observed that NSF supercomputers could have a much greater impact if they could be integrated into an increasing number of sophisticated, community-designed Web portals in addition to the historical access by individuals through the command-line. The gateway program envisioned researchers interacting with the same familiar Web interfaces, but now enjoying vastly increased analysis capabilities. Front-end gateway development in the TeraGrid program, however, was always initiated by and funded by research communities and not by the TeraGrid itself. The TeraGrid helped only with back-end integration of high-end resources. As a result, staff observed the dynamic nature and finite lifetime of gateway projects over the 7 years of the program. Often, popular gateways with sizeable user communities would fold

because the 3-year research effort that funded them had concluded.

These experiences led to a small EAGER study to understand the characteristics of successful gateways and the programs that fund them so that there could be better planning from the start. “Fundamental Cyberinfrastructure for Productive Science and Engineering: Identification of Barriers to and Enablers of Successful Projects” ran from 2009-2012. There were 66 participants in five full-day focus groups on the following topics:

1. Characteristics of successful gateways
2. Fields ready for transformation with appropriate gateways in place
3. Research initiatives that have been successful and sustainable in multiple fields and through multiple funding sources
4. External perspectives on the evaluation criteria and compelling features of potentially successful and sustainable technology projects, and expert opinions on the feasibility of new models for sustaining science and engineering portals and gateways
5. The viability of our preliminary findings and identification of additional factors and barriers that should be considered in the implementation of any recommendations emerging from this study (This group included representatives from NSF and other federal agencies.)

Attendees came from leading organizations worldwide, such as digital humanities projects, astronomy gateways, citizen science projects, online journals, and private foundations that evaluate technology projects (<http://sciencegateways.org/projects/opening-science-gateways-to-future-success/participants/>). These focus groups employed a many-to-many, participative exchange of ideas and expertise among the participants in order to generate practical insights that drew on the strength of multidisciplinary perspectives.

We observed that millions of dollars are spent on gateways, but developers face several challenges:

- They often work in isolation even though development can be quite similar across domain areas.
- They bridge cyberinfrastructure — locally, campus-wide, nationally, and sometimes internationally.
- They need foundational building blocks so they can focus on higher-level, grand-challenge functionality.
- They struggle to secure sustainable funding because gateways span the worlds of research and infrastructure.

The study outlined tensions in the academic environment in which many science gateways are developed. Gateways and perhaps other software development efforts represent a partnership between researchers in a science or engineering domain and computer scientists. The domain researchers have a vision of how technology can advance their basic research challenges while the computer scientists can be motivated by cutting-edge technology changes. Sometimes these goals can be at odds [Lawrence, 2006]. Often there is little academic or financial reward for maintaining a robust, reliable gateway even if it enables thousands to be productive. This is changing, but slowly. Academic leaders can also be unprepared for the demands of production operation and long-term planning.

The study concluded that gateways can significantly increase research productivity, but designing the most effective tools requires time and money, so we must invest wisely. The impact of gateways can be increased significantly if several key stakeholder groups understand what makes the most successful gateways successful. Recommendations are summarized here, but are available in full in the report [Wilkins-Diehr, Lawrence, 2012].

Recommendations for **leadership and management teams**: design your governance to represent multiple strengths and perspectives, plan for change and turnover in the future, recruit a development team that understands both the technical and domain-related issues, consider how you will pay for the project after the initial funding and measure success early and often.

Recommendations for **technology developers**: recognize the benefits and costs of hiring a team of

professionals, demonstrate your credibility through stability and clarity of purpose (but remember to match your end product to your goals), leverage the work of others and plan for flexibility.

Recommendations related to **outreach teams and interested community members**: identify an existing community before you begin, make it clear what your gateway is doing, know and show why your community would want to participate, and enlist your community to find solutions.

Recommendations to **funding organizations**: support the lifecycle of technology projects, design solicitations to elicit—and reward—effective business plans, recognize the benefits and limitations of both technology innovation and reuse, expect adjustments during the production process, copy effective models from other industries and sectors, and encourage partnerships that support gateway sustainability

Successful gateways demonstrate value to large numbers of constituents and keep operational costs low. Because gateways can require a diverse set of expertise to remain viable in the long term, providing a pool of expertise that many can share can be a way to reduce costs and reduce reinvention.

The Science Gateway Institute has proposed just such a pool through a conceptualization award in NSF's Scientific Software Innovation Institutes (S2I2) program. The goal of the institute is to not only serve the National Science Foundation community, but serve as a focal point for gateway development nationally and internationally. In one example of international cooperation, the Institute and the International Workshop on Science Gateways will co-edit a special journal issue featuring submissions from workshops held by both groups.

The institute plans to offer several services and resources to support the gateway development community:

- An incubator service offering consultation and resources on topics such as business plan development, software engineering practices, software licensing options, usability, security and project management as well as a software repository and hosting service.

- A team of gateway developers to help research groups build their own gateways.
- A forum to connect members of the development community.
- A modular, layered framework that supports community contributions and allows developers to choose components.
- Workforce development to help train the next generation for careers in this cross-disciplinary area and build pools of institutional expertise that many projects can leverage

Of course the institute itself needs to plan for sustainability. How will other projects pay for services? When does it make sense for NSF to fund centralized services to make other projects cheaper to launch? How does one measure success? How does one design an organization that can evolve for the long term? These questions and more will need to be addressed in the strategic plan for the institute.

Many on the Science Gateway Institute team have had long-term involvement in gateway projects and so have their own experiences with sustainability. The Center for Remote Sensing of Ice Sheets (CReSIS) was established in 2005 to improve understanding of polar ice sheet changes through improved measurement and analysis. The last Intergovernmental Panel on Climate Change (IPCC) Assessment was unable to place an upper limit on sea-level rise estimates, as a result of incomplete understanding of ice sheets, so the need for a center clearly remains. CReSIS plans to address sustainability by strengthening partner relationships, increasing core support from institutions involved, and developing collaborative proposals (to NSF, NRL, NASA) by identifying areas of important future work where the center can contribute.

The Science Gateway Group at the Pervasive Technology Institute at Indiana University has worked with a great many gateways over the years. The group observes that successful gateways have a lot to teach other groups about sustainability. A well-established characteristic of any successful gateway is

that it has leadership willing to serve a community of scientists over pursuing personal research agendas. Many gateways also provide reproducibility and transparency by tracking the provenance of a user's online experiments in their data management systems. This is a core capability of CIPRES, UltraScan, GridChem, and QuakeSim gateways (to name just a few).

Arguably more scientific application communities should consider building gateways for these reasons: they can provide a comprehensive “Software as a Service” environment and relieve the burden on users for installing and maintaining sometimes complicated applications. Going beyond this, gateway environments are excellent ways to measure impact of software: the gateway can track who is using the software, what (to some extent) they are doing with it, and similar metrics. When gateways combine these metrics with community building, they have the potential to provide a stronger bond between developers and users than other software delivery approaches.

The HUBzero team observes that one approach to making science gateways sustainable is to attract and pool funding from multiple sources. HUBzero offers a service model for science gateway support and manages services through a recharge center where many funded projects can leverage the expertise of a common set of resources. Projects choose from a well-defined menu of support services posted on the hubzero.org web site, including hub operation (hosting), web design, and consulting. Each service has a fixed price established by Purdue University on a cost-recovery basis. This recharge center supports more than 27 projects from many different funding agencies, including the US National Science Foundation (NSF), the National Institute of Health (NIH), the Department of Energy (DoE), the Environmental Protection Agency (EPA), and some private foundations. All together, this funding supports a team of 25 staff working full time on the HUBzero science gateway cyberinfrastructure project. Pooling funding in this manner allows leveraging

efforts across multiple science gateway efforts. New successful features created for one hub project are integrated into the core software and thereby migrate to all others, and to the HUBzero open source release. For example, the collaborative “project” functionality was originally developed for the Purdue University Research Repository ([purr.purdue.edu](http://purr.purdue.edu)), and the “collections” capability for finding and posting interesting content was developed for STEMEdHub.org. Wherever possible the HUBzero team seeks to make such advances generic to drive their adoption across the diverse set of gateways based on HUBzero.

We hope these general and specific observations and contribute to the discussion of the important topic of software, and gateway, sustainability. Strides forward in this area will benefit the research community in many ways.

## REFERENCES

[Lawrence, K. A. 2006] Walking the Tightrope: The Balancing Acts of a Large e-Research Project. *Computer Supported Cooperative Work (CSCW): The Journal of Collaborative Computing*, 15(4): 385–411.

[Wilkins-Diehr 2007] Wilkins-Diehr, N. 2007. Special issue: Science Gateways-- Common Community Interfaces to Grid Resources: Editorials, *Journal of Concurrency in Computation: Practice and Experience*. Volume 19, Issue 6 (April 2007), pages 743-749.

[Wilkins-Diehr, Lawrence, 2012] Wilkins-Diehr, N. and Lawrence, K. A. 2012. Opening Science Gateways to Future Success. Final report for the National Science Foundation Grant Number OCI-0948476, November 2012. Available for download at [http://sciencegateways.org/wp-content/uploads/2012/06/Final\\_Report\\_OCI-0948476.pdf](http://sciencegateways.org/wp-content/uploads/2012/06/Final_Report_OCI-0948476.pdf)].