Opening Science Gateways to Future Success

A Research Project funded by the
National Science Foundation as
"Fundamental Cyberinfrastructure for
Productive Science and Engineering:
Identification of Barriers to and Enablers of
Successful Projects"

Grant Number OCI-0948476

Final Report

Nancy Wilkins-Diehr
University of California, San Diego
wilkinsn@sdsc.edu

Katherine A. Lawrence
University of Michigan, Ann Arbor
kathla@umich.edu

November 2012

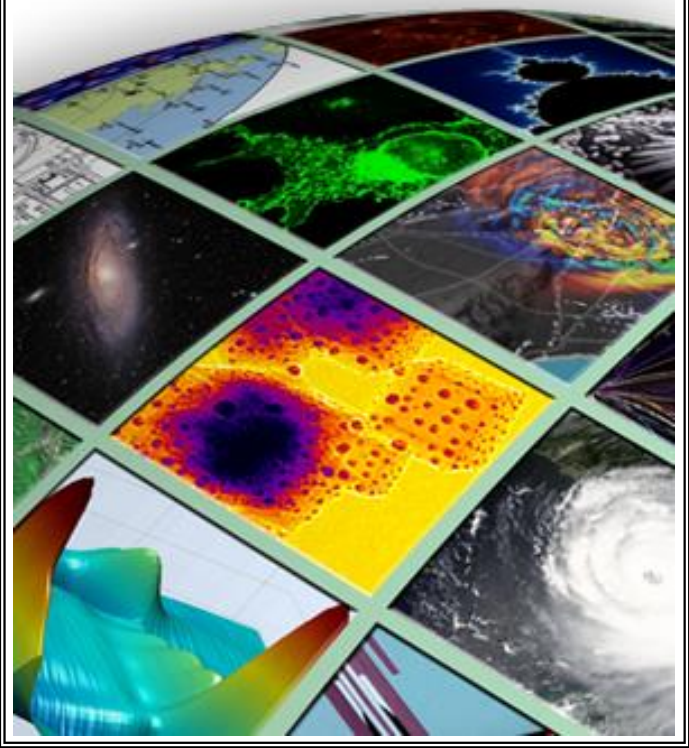

Table of Contents

Introduction

The existence of science and engineering gateways, often built with advanced cyberinfrastructure (CI) tools, can significantly increase the productivity of researchers facing the most difficult challenges. However, designing the most effective tools requires an investment of time, effort, and money. The potential impact of these gateways on science is unlimited, but the potential for transformative science relies on our ability to understand why some gateway projects change the way science is conducted at a fundamental level in a given community while others do not.

In this study, we looked for the characteristics of successful projects that would warrant sustained funding. Our goal is to help the National Science Foundation (NSF), which funded this study, make better informed decisions about how to fund and support gateways for the transformation of science and engineering. We also hope to highlight some of the challenges faced by the individuals whose hard work behind the scenes often goes unrecognized, and in doing so, offer some guidance and considerations for other stakeholders in gateway communities. This report should be a work in progress, and we welcome your questions and comments.

Methods

To achieve these goals, we conducted five full-day focus groups over two years. Each focus group invited approximately 15 participants. While we could have conducted individual interviews in a more cost-effective way, we believe conducting focus group studies in which diverse groups of people exchange ideas is more effective for eliciting novel observations and solutions to the well-established challenge of sustainability. Our activities were specifically designed to encourage the cross-fertilization of ideas and to move participants from concrete examples to generalizable principles.

In the first year, we largely studied projects that receive federal funding from the NSF. During the second year of our grant, we summarized and studied the results of our first two focus groups in June 2010, and we used what we learned to design and implement three more one-day focus groups in June 2011. Findings from that work led us to look beyond federally funded activities—both at technology projects and also at those who evaluate and fund those projects—for participants with knowledge and experience that would be relevant to NSF-funded science gateways.

The topics of the five focus groups were:

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.)

The activities and questions for each of the focus groups was customized to suit the expertise and interests of the participants. While traditional focus groups typically engage the participants in a one-to-many, facilitator-driven structure, these focus groups explored 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. Details of these activities are described in Appendix A.

Participants

Our five focus groups involved two main types of participants. Focus Groups One and Three included people who had been involved with a specific gateway or portal project. Focus Groups Two, Four, and Five included people with broader expertise and awareness about issues associated with the development, selection, and funding of gateways and portals. Focus Group Five specifically included people from federal agencies. A total of 66 participants attended our focus groups.

Selection of Focus Group One participants was based on national and international reports on cyberinfrastructure (see Appendix B for a complete list). We looked for projects mentioned in CI reports, those recommended by the NSF's CI Coordination Committee (CICC), and those recommended by scholars in the field. We then worked to include projects sponsored by all NSF directorates, but also international projects and some projects in the humanities. There was also considerable variety in project characteristics: user interfaces, the use of computational/data resources, user base, age of the project, etc. To identify participants for Focus Group Two, we used a snowball method of asking colleagues and experts for recommendations of both possible participants and people who could recommend others. For Focus Groups Three and Four, we drew on suggestions of specific people and projects as well as suggestions of broader areas of expertise (e.g., non-profit marketing, open-source projects) that were made by the first two focus groups. To identify participants for Focus Group Five, we asked our program officer for recommendations within and outside of NSF, and then we combed the websites of all other potentially relevant federal agencies (e.g., DOE, DOD, NOAA, EPA, etc.) to identify units (and people within those units) that were concerned with gateway-type technologies and/or research in science/engineering.

For a complete list of projects and organizational affiliations represented in the focus groups, see Appendix C.

Recommendations and Considerations

We describe the results of our analysis of the focus groups below, grouped as recommendations or considerations for various stakeholders in science gateways. These four main stakeholder groups include:

- Leadership and Management Teams
- Technology Developers
- Outreach Teams and Interested Community Members
- Funding Organizations

Our recommendations have been extracted from our presentations and articles. For additional detail and discussion, we recommend reading or viewing the following:

"Roadmaps, Not Blueprints: Paving the Way to Science Gateway Success," by K. A. Lawrence and N. Wilkins-Diehr. Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond (XSEDE 2012). Article No. 40, pp. 1-8; July 16-20, 2012. ACM, New York, NY. doi>10.1145/2335755.2335837. This article summarizes our complete study, synthesizing the input of all five focus groups. See the PDF on ACM's Digital Library.

"Opening Science Gateways to Future Success," video by K. A. Lawrence and N. Wilkins-Diehr. This 6 ½ minute video was created for our XSEDE 2012 paper presentation and is available on YouTube: http://www.youtube.com/watch?v=4ziEt0LRxEA&feature=youtu.be

"Opening Science Gateways to Future Success: The Challenges of Gateway Sustainability," by N. Wilkins-Diehr and K. A. Lawrence. Gateway Computing Environments Workshop (GCE), 2010, pp.1-10; November 14, 2010. IEEE Computer Society (Xplore Digital Library). doi: 10.1109/GCE.2010.5676121. This is a report on the first focus group conducted in June, 2010. See the PDF on IEEE.

"Making science gateways a success," by K. Lawrence and N. Wilkins-Diehr. Featured research report at *International Science Grid This Week,* January 26, 2011. This is a summary of our GCE Workshop paper. See the online article.

Appendix D lists some additional resources that we found to be useful and insightful. Our website (sciencegateways.org) also contains links to these articles and other content.

Overview

Gateways represent a delicate partnership between researchers in a science or engineering domain and computer scientists. The domain-specific researchers have a vision of how technology can advance their basic research challenges while the computer experts are interested in the fundamental challenges and opportunities presented by the gateway's goals. Typically, these groups of academic researchers come together to build something that serves

both their research objectives, though sometimes the construction of the gateway is handled by software engineers without a research agenda. If the partnership succeeds, the software can be adopted widely by a broader research community.

This partnership and context is important to remember as a background for the challenges faced by gateways. Academic funding and reward systems are not geared toward the production of stable, operational software. Nevertheless, gateways could only emerge from the vision of leading researchers who are steeped within an academic discipline. Consequently, gateways are led by people who often are unprepared for the business, technology development, community-building, and fundraising demands of these projects.

Although each gateway presents unique scientific and operational challenges, speak to enough of these veterans, and you will find certain common patterns that they share. All projects pass through specific stages of development: research, development, deployment, and operations (or maintenance, and some become institutionalized as infrastructure). First, gateways typically begin as a research project that tests viability or demonstrates proof-of-concept. Later, a gateway is developed into a stable product for deployment. At the operational phase, it must be maintained, requiring different skill sets and organizational structures and ideally a lower cost. Finally, as infrastructure, it is meant to have durability, yet most funding for science does not support that type of permanence, and leaders find themselves searching for scarce funds.[1] Our recommendations below are structured around these common issues and concerns.

Leadership and Management Teams

Design your governance to represent multiple strengths and perspectives.

Gateway governance fares best by fulfilling a set of specific roles, each satisfying certain needs:

- To begin, a project needs a forward-thinking **leader** (or leaders) able to generate significant community interest and plan for the life of the gateway. The ideal leader welcomes latecomers and offers ownership to anyone who wants to participate. Leaders should understand either the technology or the domain science (or both).

- **Project managers** need to oversee and manage a large project with the level of attention that a PI typically cannot. This person may be trained as a project manager or may have project-related expertise.

[1] More specifically, the research phase of a project might include literature reviews, requirements gathering, market and competitive analysis, and community engagement. The development phase includes prototypes, usability studies, and opportunities for user feedback. The deployment portion includes productization and marketing components. The operations or maintenance phase might describe planned cost reductions and transitions to non-research funding sources.

- A full-time **outreach "evangelist"** can build the community; this could even be a PhD student from the domain. A "user committee" can also help determine direction of the infrastructure and serve as proselytizers to the wider community.

- The **sponsoring organization** should have a clear identity and credibility, such that it knows the audience and its capabilities.

- Rather than organizing site visits, use project **advisory boards**—composed of representatives from all stakeholder groups (both current and future constituencies)—to provide useful guidance on matters of science as well as business. For some very large projects, the NSF has required a board of directors, along with an infrastructure to manage the work to be done.

Plan for change and turnover in the future.

Projects should include a strategy for the inevitable turnover that will and should occur over the life of a project. Over time, the project will need a different leadership, structure, management, and people as the project moves from a "startup" phase to a more mature and stable form. On your staff chart, indicate what would happen if key people left the project and what deliberate plans for succession you have considered.

Also, be willing to move to new venues when your current form is not effective. Is the gateway a means or an end goal? The fate of a project ultimately is to be either disused or subsumed, so identifying how it might be subsumed is critical. Acknowledge "exit plans" and look for post-NSF funding partnerships to develop before it is too late.

Recruit a development team that understands both the technical and domain-related issues.

Successful gateway building requires expertise in a variety of areas, so look for team members or enlist consultants with knowledge of:

- web technologies
- database programming
- data management and curation
- grid computing
- security
- high-performance computing
- usability and design
- community outreach
- the specific domain and its community of researchers

Consider how you will pay for the project after the initial funding.

Diverse sources of funding and partnerships may be necessary to allow sustainability. You may want to involve multiple partners from the very beginning: multiple federal agencies, professional societies, non-profits or foundations, university-based technology transfer offices.

The caveat is that every organization has its own agenda, so you must specify how your goals align with each partner's mandates. Focus group participants offered specific advice for working with venture capital-type funders: make sure the deliverables, conditions, expectations, and milestones are clearly spelled out.

Projects might also look at their utility to wider communities who can foot the bill:

- User fees in the form of micropayments or institutional subscriptions.
- Ad placement by a relevant organization.
- Sponsorship by an association or university.
- Finding multiple institutions or archives that can host data in a distributed way. This will the financial and technical load for any one site and allow some degree of budgeting for the indefinite support of the content.

Some participants advised to "start as you mean to go on" and charge from the beginning if you plan to charge later. This may make it more difficult to grow an audience, but it will prevent community disenchantment when you no longer have base funding.

Measure success early and often.

Gateway work enables research over the longer-term, but it is a very different type of research product and does not fit the usual template for academic evaluation. Effective metrics can help funding agencies evaluate the impact and sustainability of your gateway and whether it merits additional funding. Measurement requires forethought and ongoing attention, may need to be built into the technology from the beginning, and should match the time scales associated with the project. It is better to pick two measures and follow them rigorously. Careful analysis can also help you set project direction as you observe how users engage with your gateway.

You may need to address "success" metrics from the perspective of the science or engineering domain that will use the gateway as well as from the perspective of the technology-focused team that is developing the gateway. These groups have different needs and potentially conflicting timelines. Consider involving social scientists to monitor a wide variety of metrics. Social scientists may be able to provide a perspective on use and usability that your own development team lacks. An external assessor who is familiar with the development of a project will understand the nuances of your progress better than someone reviewing periodic quantitative measures, so stay connected to your assessor to communicate more effectively about your progress in a timely way, particularly if your plans are going off track.

Here are some common quantitative metrics and the tradeoffs associated with some of them:

- Number of users: User registration can be a barrier to use, yet without registration, it is difficult to track the number of unique and returning users, so numbers of users may not be feasible or accurate. Moreover, volume is not necessarily a measure of importance but the fraction of the user community that is being served demonstrates impact more fairly.
- Alignment with the user community and user satisfaction: Called the "net-promoter score" in market research, this speaks volumes about your impact. Automating the use of the site supports responsiveness to changing user needs.

- Different types of contributions and their volume, downloads of content: Track how the data generated by the site is being used.
- Availability (uptime) of the gateway or hours of use of specific resources (e.g., equipment, computing resources, etc.)
- Citation rates for datasets or software
- Metadata quality or data integrity: This is the degree to which a respected curation process is used in making datasets available in the gateway. You can use metadata to make newer and archived data searchable and accessible for the long term; this allows you to demonstrate value of the site to funders for the long term.
- Money saved or efficiencies improved for individuals

Qualitative data is also important. Demonstrate your impact through stories and progress reports: Why is this gateway helping members of the community?

Technology Developers

Recognize the benefits and costs of hiring a team of professionals.

Professional software developers and a well-developed software engineering process have been critical to the success of some projects, but others still advocate for hiring graduate students. Professionals can be more efficient and may be more appropriate for overseeing software services, but even if professionals are preferred for a project, they may be difficult to find and their participation offers no secure career path when sustainable funding is uncertain. Regardless, as a project is weaned from initial NSF funding, leaders must consider how to transition to professional maintenance staff, in terms of finding, training, and funding them.

An additional consideration is whether your project will be “siloed” in one institution. Arranging partnerships with other institutions, despite the extra work it entails, may support knowledge transfer as the team composition changes. This may also provide access to broader types of expertise or sources of professional programmers.

Two other crucial professionals on the development team are usability experts and data curators. Usability professionals help the community speak for itself rather than depending on the imagination of computer scientists who are unfamiliar with the science or engineering domain. Social scientists (as an alternative or supplement to usability professionals) can help monitor an audience’s interests and needs and figure out how to keep them engaged.

Data curation is likewise dependent on experienced professionals. One common challenge for data integration is to enable users to recast the data at a later time. For example, Uniform Resource Identifiers allow objects hosted by gateways to be presented elsewhere. Where it ultimately “lives” does not matter, just that people can access it through the portal. Data professionals, working with skilled programmers, can identify practical ways of ensuring both the quality and sustainability of the gateway’s content.

Demonstrate your credibility through stability and clarity of purpose (but remember to match your end product to your goals).

Reliability, clarity of focus, and elegance of design convince new users to adopt a gateway as part of their research repertoire. Unfortunately, the software packages that underlie many gateways are themselves research projects, and these have been completely redesigned in the name of research, causing havoc for the gateways that rely on them. Using a stable underlying middleware, or one that remains backwards-compatible when altered, is key.

Nevertheless, recognize if your planned efforts are beyond what is necessary. If your gateway is still research or proof-of-concept, what you create may not need to be as refined and stable.

Leverage the work of others.

Industry-grade software may be optimal, but the budget constraints of academic settings favor using inexpensive tools and techniques (such as open-source software) for building gateways. Decide what is more cost effective for your project. Simple, commoditized tools have been used to enable cutting-edge science, so look at what tools and methods already exist or what is absolutely necessary. That said, reuse is sometimes not supported by funders, and you may be subject to software changes that you cannot control (see above). Remember also that tools that begin as open-source require ongoing maintenance and attention and must be funded with that in mind.

Plan for flexibility.

Agile development of a modularized product is likely the most flexible approach for managing a gateway project. The mindset of an agile development project is to “fail fast,” and iterate, learning from failures until achieving a functional product. In particular, technology can change in the time it takes to launch a project, so projects must be willing to anticipate and adapt to change. As described above, changing technology introduces stability issues, too. In particular, once you release your product to the community, be prepared for input and reactions from other stakeholders and adapt your plans accordingly.

Outreach Teams and Interested Community Members

Identify an existing community before you begin.

The community that will use your gateway needs to exist before the gateway is created. It also helps if the leader of your project is recognizable to and well-respected by that community so that he or she can help champion your gateway. The participants you engage at the start should be the same people you hope to engage for the long term.

Make it clear what your gateway is doing.

Be sure to communicate what your gateway does in a clear, crisp way. It is effective to have a simple “pitch” that makes it instantly clear what your gateway can do, yet be careful not to oversimplify and thereby degrade the nuanced value of what your gateway offers to the

research community. A science gateway should push a clear scientific challenge, enabled by technology but without technology as the focus. This clarity will appeal to community members, funders, and partners by establishing your credibility and your well-researched understanding of why you're doing the project, who it serves, and what gap it fills.

Know and show why your community would want to participate.

Do a thorough requirements analysis to define your community's specific and current goals, needs, requirements, and ways of working. Articulating your community's needs is not straightforward. As your gateway takes shape, the tools themselves can change the ways that researchers do their work, requiring your development team to revisit and redesign requirements.

Once you know what your community wants, make sure they'll want to use your gateway. First and foremost, people need to see that you have placed a priority on usability and reliability. Retaining a focus on content is important as well, as some gateways have found that it can be easy to alienate researchers. In addition, helpful users need incentives that make participation easier or even advantageous. Some gateways have successfully recognized and motivated participation with certificates, awards, or other visible validations of contributions or reputation. Make sure that latecomers to the community will not feel left behind.

Enlist your community to find solutions.

Finally, your community could identify relevant best practices beyond its own domain that could be applied to the features and functions of your gateway. For example, you might look at large commercial websites or established academic endeavors with significant longevity to identify good ideas to borrow and adapt. Remember that you must be prepared to adapt in response to technological, user, funding, or research changes and trends.

Funding Organizations

Support the lifecycle of technology projects.

Technology projects—and gateways in particular—follow a lifecycle that is different than a typical research project. As we described in the overview to this section, gateways move through several development stages. Just as venture capitalists (VCs) review projects at specific validation checkpoints, federal funders could provide a formal process for funding a gateway as it moves from research to development to deployment to operations (or maintenance). There are a few caveats when applying staged funding to the sciences. The markers for monetary success in the VC world may be easier to ascertain than those for scientific success. All stakeholders need a clear understanding of what is being rewarded as a project moves from one stage to the next. The better these expectations are defined up front, the better the gateway's chance of success.

A few other aspects of the technology development lifecycle suggest that gateways need to be supported in new ways. First, research and productization tend not to be compatible and

achievable in a short time frame. In particular, users become frustrated and are unwilling to use a gateway for their domain-science research if the gateway is in an experimental development stage at the same time. Second, that research-to-production transition takes time and involves different expectations. Producing something reliable and useful often takes longer than the common 5-year timeline for funding. Funders could consider “transition” grants that support the move to being a “product.” Finally, the changing technology behind gateways raises the cost of maintenance to be proportionately more than physical infrastructure (e.g., telescopes) might require. Some funders, such as UK-based JISC, award smaller, consistent amounts of money to keep a project going. While this model would not be practical for all technology projects, it could support highly experimental projects or projects that have a long start-up, and it could allow funders to have more projects in the pipeline. It could also support the more operational or maintenance phases of a project in a controlled manner.

Design solicitations to elicit—and reward—effective business plans.

How solicitations are crafted and reviewed influences science gateway design and success. Problems with solicitations have included insufficient attention to new, non-science components of the proposal; excess jargon and boilerplate language; and underfunding business planning milestones.

Taking non-science components seriously. Some programs have required impact statements, sustainability plans, and reuse descriptions, but these and other additions need careful consideration. In particular, the funding organization should align its evaluation system to appropriately recognize and reward clear efforts in non-science areas. For example, new requirements must be given due consideration; if software reuse is a requirement, then incentivizing innovation is counterproductive. If innovation is needed, require the proposal to include descriptions of other software products that have addressed similar topics, and why it cannot be reused.

Of course, new components of proposals require reviewers with the relevant expertise to evaluate proposals effectively at each stage. A program officer can set the stage by selecting diverse review panels and explaining expectations. Researchers, who are typical panel members, may not have the knowledge to evaluate a gateway that is destined to be a stable software product, so include a complement of reviewers with expertise in management, sustainability, and governance. Other federal agencies might also play a part in evaluating NSF projects, particularly to evaluate projects for subsequent funding in agencies that support operations.

Words, words, words. Solicitations need more clear and simple language. A solicitation filled with jargon and boilerplate often results in proposals filled with the same. Need we say more?

Provide funds for crucial milestones. Another way to support effective business plans is to provide planning-process funding or bonuses for specific targets: paying user experience specialists to assess the user community first, conduct a competitive analysis, and propose a sustainability plan.

Recognize the benefits and limitations of both technology innovation and reuse.

Funders must find a cost-effective balance between a need for standards and a need for competition that introduces alternate methods. Consider how funding could enable technological flexibility in research without the rigidity of road-type infrastructure.

Expect adjustments during the production process.

A complication for gateways is that technology and goals can change so quickly that something envisioned at the beginning of a multi-year project may need significant changes midway through. As we suggested above, synchronizing funding to the stages of the development process would allow for evaluation and adjustments at intervals. Milestones that indicate progression from one stage to another might include market analysis, technology deliverables, blueprints, community deliverables, formation of advisory groups, business plans, and sustainability analyses. In our focus groups, participants liked the idea of "roadmaps, not blueprints," allowing a project to pursue a set of clear goals without the rigidity of fixed specifications. Contracts could describe an interrelated timeline, milestones, risk analysis, and budget.

Copy effective models from other industries and sectors.

Many excellent models for supporting technology development exist both within and outside the National Science Foundation. For example, the NSF's informal education programs fund a pilot study followed by planning, implementation, and evaluation phases. This is a functional structure for turning projects into sustainable, operational products. Several other models were proposed and well received by focus group participants: incubators, consortia, and app stores.

Incubators. High-tech and Internet-based businesses have been well supported by the incubator model. Typically provided by venture capitalists, an incubator offers pooled expertise or services to a set of startup projects that could not otherwise afford or access such resources. Gateways would likely benefit from a pool of project management or business development expertise including financial experts, strategists, legal professionals, and sustainability specialists. An incubator for gateways might also provide a pool of developers with specific expertise to support gateway projects. While this would require overhead on the part of the NSF, the improved planning by gateways could ultimately allow more efficient spending.

Consortia. Just as professional associations, non-profit organizations, and users of common resources (e.g., IEEE, Apache, OGCE, XSEDE) represent communities with common needs and objectives, gateway developers are part of a community that is dispersed due to diverse disciplines and funding sources. A gateway consortium could provide a venue to share challenges and capture lessons learned (from failures and successes); resources for connecting newcomers with more experienced developers; and a software repository to share reusable components. Such a consortium should have criteria for participation in the community but be open enough to include self-identified participants.

"App" store. Just as Apple products and Android phones have marketplaces for customized applications, a similar system for gateways would allow developers to find and reuse useful components. The system would also provide feedback to developers and a means for the better components to rise to the top.

Encourage partnerships that support gateway sustainability.

NSF's mission limits it to primarily funding start-up endeavors while other federal agencies fund operational projects. For example, NASA and NOAA have operational roles and curate data for the research community. Cross-agency partnerships, industrial collaborations, university involvement, and partners willing to fund valuable gateway services are all ways to extend funding and expand audiences. Industrial collaborations should require careful consideration, because proprietary content may restrict the openness of the academic process.

That said, it is not clear that gateways enabling basic research can always find a home elsewhere. It may be in NSF's best interest to continue to fund tools that fundamentally enable research. One way might be to treat gateways as facilities such as MREFC (Major Research Equipment and Facilities Construction) or MRI (Major Research Instrumentation) projects, not research projects.

References

Brown, J. and Isaacs, D. 2005. *The World Café: Shaping Our Futures Through Conversations That Matter.* Berrett-Koehler Publishers, Inc., San Francisco, CA.

Heath, C. and Heath, D. 2010. *Switch: How to Change Things When Change Is Hard.* New York, Crown Publishing Group.

Lawrence, K. A. and Wilkins-Diehr, N. 2012. Roadmaps, Not Blueprints: Paving the Way to Science Gateway Success. *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond* (XSEDE 2012). Article No. 40, pp. 1-8; July 16-20, 2012. ACM, New York, NY.

Wilkins-Diehr, N. and Lawrence, K. A. 2010. Opening Science Gateways to Future Success: The Challenges of Gateway Sustainability. *Gateway Computing Environments Workshop (GCE), 2010,* pp.1-10; November 14, 2010. IEEE Computer Society (Xplore Digital Library).

Appendix A: Format of Focus Groups

We describe each of the focus group activities in general terms, with examples of the specific questions or tasks we used to direct each unique focus group. (Expanded from Lawrence and Wilkins-Diehr, 2012.)

Spatial Warm-ups

Spatial warm-up exercises are an icebreaker that also helps start the group conversation by introducing participant-specific content relevant to the day's topics. Because the participants are standing and moving around, it also gives them "physical memory" of the information they learn about each other. For example, the group gathers in an open floor area, and we identify sides of the room as north, south, east, and west of a map of the United States and beyond. We ask them to silently move to where they currently live, and then each person identifies themselves, their affiliation, and location. We follow this with other specific dimensions (plotted along an X and Y axis or along a continuum) so that they can see how they fit among their colleagues. Dimensions for the first and third focus groups included age of their portal or gateway, number of users vs. sophistication of those users, and the amount of data vs. computation capacity their portal/gateway provides. Other exercises (with different focus groups) helped people identify their sources of funding, expertise, types of projects, and disciplines by moving into clusters.

Wandering Flip Charts with Voting

In this method, we pose a question with a set of topics. Each of these topics is written on a flip chart. We ask participants to find a flip-chart topic that interests them, discuss with others at that location, and write their ideas, including the name of their project or organization when relevant. After a set time of wandering among the flip charts, the participants are asked to revisit the charts and, on a second sheet, consider a second question that elaborates on the first (e.g., enablers of items identified in the first round). Finally, they are given colored dots and asked to vote for those enablers that they consider to be most essential (must have) versus optional (nice to have). The questions/topics used with the focus groups were:

Focus Group 1: *Question:* "On your project, how did you handle [Topic X] that contributed to the success of your gateway/portal?" then "What initial building blocks for your organization/project enabled these factors for success?" *Topics:* Productivity enhancement (tools), Contents (materials), Target audiences (who the project was for), Technology (behind the scenes or up front), Community engagement (how you got the audience involved), Traits or culture of the discipline, Partnerships, Structure of the organization building the gateway.

Focus Group 3: *Question:* "In what ways has [Topic X] contributed to your success?" *Topics: Audience*/Participation, Partnerships/Sponsorship, Organizational Design, Technology Selection, Shaping External Factors

Focus Group 4: *Question:* "When working with technology projects, what does your organization do to [Topic X]?" *Topics:* Set up expectations through solicitation guidelines for the content of their proposal?; What do you do to help get projects started and transition to new phases?; Require organizational structures that they set up? Governance? Advisory Panels?;

Solicit status reports at certain points in time?; Create support systems for connecting with and learning from peers?

Focus Group 5: *Question:* Please react to these success factors: How would you recognize that these factors exist? How have you seen them manifested? How have you measured them? How might they present a problem? *Topics:* Thirty-six key project success factors that had been distilled from the previous four focus groups, grouped into the categories of user community, business models, product design, staffing and software.

World Café

The "World Café" method is designed to promote small-group discussion and cross-fertilization of ideas (Brown & Isaacs, 2005). People sat in small groups at tables with paper tablecloths for notetaking and they discussed a topic. After 25 minutes, everyone but one participant from each group (serving as the table's host) rotated to different tables to discuss and cross-fertilize, returning to their original groups after 20 minutes. The groups then prepared to report back to the larger group, after which we discussed key themes across the groups. The topics for each group were set up as follows:

Focus Group 1: Take a walk down memory lane: What if you were telling someone starting up a portal project about what was easy and what was hard on your project. Tell each other stories about an aspect or an incident where you thought, "This is hard." How did you work through it? Did that work? Knowing what you know now, what would you have done differently?

Focus Group 2: Identify disciplinary areas that are ripe for a gateway. In the first round, you will describe your ideas to each other, then identify generalizable characteristics (e.g., data heavy, computationally focused, community oriented, etc.). In the second round, identify what is in the way of these gateway ideas being realized or why the gateway has not been built yet.

Focus Group 3: Imagine you have been brought in as consultants to a new gateway or digital resource. Reflect on your own organization or project, and consider what was difficult or did not go so smoothly. What would you do differently, and how would you advise a newcomer to avoid the same hard lessons?

Focus Group 5: Imagine you have been brought in as sustainability advisers to a new gateway or digital resource. Drawing on what you know as a federal employee and how government works, how would you advise a them to navigate from the development phase (funded by NSF) to an operational phase (funded by another agency and/or partner organization)? Consider such questions as: Who needs to be involved? How can your agency support them? What are the conditions that need to be in place for a new agency to take over? Are there compliance or standards issues that come into play?

Structured "Brainstorming"

Structured brainstorming is an idea generation activity done in pairs, groups, or as a large group, with the goal of addressing a specific question or series of questions or tasks. The questions/tasks for each group were as follows:

Focus Group 1: Identify external forces, opportunities, challenges. Possible categories to consider: Funding sources (including external evaluation, broader impact); Publication venues (journals, conferences); Evolution of scholarship (grand challenge questions, collaboration, disciplinary trends, societal engagement); Demand (patterns, preferences, demographic shifts); Technology (infrastructure, innovations, standards, R&D); Partnerships and federations (with industry, between universities, between agencies); Education trends; Other major forces.

Focus Group 3: In a perfect world, what would you like for your project/organization to make it even better? Identify concrete things (beyond money) that would make a difference then propose ways of making it happen using concepts from *Switch* by Chip Heath and Dan Heath (2010).

Focus Group 4: In pairs, identify features, characteristics, and criteria of exemplary technology projects that would transfer to the evaluation of other projects. How might these criteria might be measured? [A large group discussion followed reports from each pair.] Consider these questions: "How do these align with the criteria you use to evaluate projects in your organization?" "Do these criteria apply to different phases of a project?" and "How hard is it to measure these?"

Focus Group 5: Each pair of participants will consider two ideas that have been identified in the previous focus groups. You will receive two sheets of paper, each with one possible means of support for gateway projects. Consider the strengths and weaknesses of these ideas, what would be necessary to implement the ideas, and examples of how they might be implemented. Then hand your sheets to two other pairs and continue reflecting on the proposed ideas that you have received from others.

Sales Pitch

For Focus Group 2, the second idea-generation exercise that built on the first day's discussions was the task of creating a sales pitch for a portal project. Each small group was asked to develop a one-minute "elevator pitch" for their hypothetical funders as well as convey the content that would be on the home page of their portal that would attract their desired audience. From these presentations, we discussed what elements and features appeared to be the most compelling or salient, which then extended to a broader conversation about what conditions are necessary for portals to be built and to be sustained.

Create a Solicitation

Participants gathered in groups to create or identify the key points of an NSF solicitation (or request/call for proposals or grant program that represents an ideal of what would be in a solicitation) to set up a portal/gateway for success.

Focus Group 4: If anything were feasible, both in terms of what the funder could provide and what the project should accomplish, what would be the key ingredients? Consider the following elements: Length and timeline of the grant; entry points for involvement by other parties; oversight or support provided by the funding agency; milestones and evaluation criteria; governance. [After group discussion of all the elements identified, participants were asked what changes were needed to make such a solicitation successful in a constrained environment by

applying the concepts from *Switch* (Heath & Heath, 2010).] Consider the questions: “What stood out? What else would you add? Where would we need to take first steps? What have you learned? Do you think you’ll try any of these ideas?”

Focus Group 5: [After the Wandering Flip Charts exercise] How could these identified success factors be incorporated into a request for proposals (RFPs)? Consider the following: What language would convey the importance of this and how it might be met? How prescriptive can a solicitation be? What would the funding agency have to do to monitor or support this requirement? How do you keep these from becoming overwhelming and make sure the applicants are not just paying lip service?

Appendix B: Reports Used as Background Research

The following reports were used to identify potential subjects and participants for our focus groups.

- Building a Cyberinfrastructure for the Biological Sciences (http://research.calit2.net/cibio/archived/CIBIO_FINAL.pdf)
- CHE Cyber Chemistry Workshop (http://bioeng.berkeley.edu/faculty/cyber_workshop)
- Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences (http://www.acls.org/cyberinfrastructure/cyber.htm)
- Computation as a Tool for Discovery in Physics (http://www.nsf.gov/pubs/2002/nsf02176/start.htm)
- Cyberinfrastructure for the Atmospheric Sciences in the 21st Century (http://www.cisl.ucar.edu/cyrdas/cyrdas_report_final.pdf)
- Cyberinfrastructure for Engineering Design (http://www.mne.psu.edu/simpson/NSF/EXCITED/)
- CyberInfrastructure and the Next Wave of Collaboration (http://www.educause2005.auckland.ac.nz/interactive/presentations/Atkins.pdf)
- Cyberinfrastructure for Engineering Research and Education (http://www.nsf.gov/eng/general/Workshop/cyberinfrastructure/index.jsp)
- Cyberinfrastructure for Environmental Research and Education (http://www.ncar.ucar.edu/cyber/cyberreport.pdf)
- CyberInfrastructure (CI) for the Integrated Solid Earth Sciences (ISES) (http://tectonics.geo.ku.edu/ises-ci/reports/ISES-CI_backup.pdf)
- Cyberinfrastructure and the Social Sciences (http://www.sdsc.edu/sbe/)
- Cyberlearning Workshop Series (http://www.cra.org/Activities/workshops/cyberlearning)
- Geoinformatics: Building Cyberinfrastructure for the Earth Sciences (http://www.geoinformatics.info/)
- Geoscience Education and Cyberinfrastructure, Digital Library for Earth System Education (http://www.dlese.org/documents/reports/GeoEd-CI.pdf)
- Identifying Major Scientific Challenges in the Mathematical and Physical Sciences and their CyberInfrastructure Needs (http://www.nsf.gov/attachments/100811/public/CyberscienceFinal4.pdf)
- Materials Research Cyberscience enabled by Cyberinfrastructure (http://www.nsf.gov/mps/dmr/csci.pdf)
- Multiscale Mathematics Initiative: A Roadmap (http://www.sc.doe.gov/ascr/mics/amr/Multiscale%20Math%20Workshop%203%20-%20Report%20latest%20edition.pdf)
- An Operations Cyberinfrastructure: Using Cyberinfrastructure and Operations Research to Improve Productivity in American Enterprises (http://www.optimization-online.org/OCI/OCI.pdf)
- Planning for Cyberinfrastructure Software (http://www.nsf.gov/od/oci/ci_workshop/index.jsp)

- Preparing for the Revolution: Information Technology and the Future of the Research University (http://www.nap.edu/catalog/10545.html)
- Polar Science and Advanced Networking: workshop (http://www.polar.umcs.maine.edu/)
- Research Opportunities in Cyberengineering/Cyberinfrastructure (http://129.25.60.81/%7Eworkshop/)
- Revolutionizing Science and Engineering Through Cyberinfrastructure: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure (http://www.nsf.gov/od/oci/reports/atkins.pdf)
- A Science-Based Case for Large-Scale Simulation (http://www.pnl.gov/scales/docs/volume1_72dpi.pdf)
- Summit on Digital Tools for the Humanities (http://www.iath.virginia.edu/dtsummit/SummitText.pdf)
- Trends in IT Infrastructure in the Ocean Sciences (http://www.geo-prose.com/oceans_iti_trends/oceans_iti_trends_rpt.pdf)
- Workshop on the Challenges of Scientific Workflows (http://vtcpc.isi.edu/wiki/index.php/Main_Page)
- Workshop on Cyber-Based Combustion Science (http://www.nsf-combustion.umd.edu/)
- Workshop on Cyberinfrastructure in Chemical and Biological Systems (http://www.oit.ucla.edu/nsfci/default.htm)
- NSF's Cyberinfrastructure Vision for 21st Century Discovery (http://www.nsf.gov/pubs/2007/nsf0728/index.jsp)
- Virtual Research Environment Collaborative Landscape Study (funded by the UK's Joint Information Systems Committee, report not yet released)
- NSF's Engineering Virtual Organizations (EVO) program (http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501057)
- History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures (http://www.si.umich.edu/pne/PDF/ui.pdf)
- Cyberinfrastructure - A Special Report (http://www.nsf.gov/news/special_reports/cyber/index.jsp)
- Science of Collaboratories Study (http://scienceofcollaboratories.org/)
- Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations (http://www.ci.uchicago.edu/events/VirtOrg2008/VO_report.pdf)
- Sustaining Digital Resources: An On-the-Ground View of Projects Today, Ithaka Case Studies in Sustainability (http://www.ithaka.org/ithaka-s-r/strategy/ithaka-case-studies-in-sustainability)
- The Importance of Long-Term Science and Engineering Infrastructure for Digital Discovery (http://www.sciencegateways.org/sgwwhitepaper)

Appendix C: Organizational Affiliations of Focus Group Participants

The **gateways, portals, and technologies** represented by the participants in the first and third focus group included the following:

- National Science Digital Library
- iPlant
- Earth System Grid
- TAPoR (Text Analysis Portal for Research)
- GridChem
- GISolve
- Linked Environments for Atmospheric Discovery (LEAD)
- National Snow and Ice Data Center
- nanoHUB
- VORTEX WINDS (A Virtual Organization to Reduce the Toll of Extreme WINDS on Society)
- CIPRES (Cyberinfrastructure for Phylogenetic Research)
- MyExperiment
- FLOSSmole (Free, libre, and open source software project)
- Galaxy Zoo
- Drupal, a gateway-building technology
- ScienceForCitizens.net
- Protein Data Bank (PDB)
- Long Term Ecological Research Network Office (LTER) and University of Virginia
- Folding@Home
- Sage Bionetworks
- eBird, at the Cornell University Lab of Ornithology
- National Digital Information Infrastructure and Preservation Program (NDIIPP) in the Office of Strategic Initiatives at the Library of Congress
- Dark Energy Survey
- Library of Congress's World Digital Library
- Sakai
- 18thConnect
- Computational and Information Systems Laboratory (CISL) Research Data Archive, at the National Center for Atmospheric Research
- Networked Environment for Music Analysis (NEMA)

Participants in the second and fourth focus groups were selected for their broad expertise and awareness about issues associated with the development, selection, and funding of gateways and portals. They came from **organizations** including:

- Virtual Knowledge Studio for the Humanities and Social Sciences, Netherlands
- Centre for e-Science, University of Lancaster, UK
- Electronic Visualization Laboratory (EVL), University of Illinois at Chicago

- Illinois Center for Computing in Humanities, Arts, and Social Science (I-CHASS), National Center for Supercomputing Applications (NCSA)
- Cyberenvironments and Technologies group, NCSA
- Center for Public Policy, University of Houston
- Committee on the Conceptual and Historical Studies of Science, and Fellow at the Computation Institute, University of Chicago
- University of Michigan, School of Information and Provost's Office
- Mass General Institute for Neurodegenerative Disease
- CyberGIS Software Integration for Sustained Geospatial Innovation
- Common Solutions Group
- Hawkshurst Group
- Oxford e-Research Centre, UK
- Elsevier Labs
- Science House
- Open Health Tools
- Microsoft Research Connections
- Pew Internet & American Life Project
- Citizen Science Central, hosted by the Cornell Lab of Ornithology, Department of Program Development and Evaluation
- Andrew W. Mellon Foundation's Department of Scholarly Communications and Information Technology
- Ithaka
- New York Times

Federal agencies represented in our fifth focus group included one or more people from the following agencies:

- Department of Defense (multiple units)
- Department of Energy (multiple units)
- Institute of Museum and Library Services, Office of Library Services
- Library of Congress
- National Endowment for the Humanities, Office of Digital Humanities
- National Institute of Food and Agriculture (NIFA), US Department of Agriculture (multiple units)
- National Oceanic and Atmospheric Administration (NOAA), US Department of Commerce
- National Science Foundation (multiple directorates)
- United States Geological Survey, Core Science Systems

Appendix D: Additional Resources and Reports

Citizen Science Central Toolkit, an online, adaptive, and evolving resource for starting a citizen science website
http://www.birds.cornell.edu/citscitoolkit/

Ithaka Case Studies in Sustainability series (2008-2011)
http://www.sr.ithaka.org/research-publications/case-studies-sustainability-series
including

- Sustaining Digital Projects: An On-the-Ground View of Projects Today
- Revenue, Recession, Reliance: Revisiting the SCA/Ithaka S+R Case Studies in Sustainability
- Sustainability and Revenue Models for Online Academic Resources

Virtual research environment collaborative landscape study, a JISC-funded project (January 2010)
http://www.jisc.ac.uk/publications/reports/2010/vrelandscapestudy.aspx

"Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations," a report from the "Building Effective Virtual Organizations" Workshop (May 2008)
http://www.ci.uchicago.edu/events/VirtOrg2008/VO_report.pdf

Science of Collaboratories Study (2001-2004)
http://soc.ics.uci.edu/