Opening Science Gateways to Future Success: The Challenges of Gateway Sustainability

Nancy Wilkins-Diehr[1]
San Diego Supercomputer Center
University of California at San Diego
9500 Gilman Drive
La Jolla, CA 92093-0505 USA
+1-858-534-5118
wilkinsn@sdsc.edu

Katherine A. Lawrence
School of Information
University of Michigan
North Quad
105 South State Street
Ann Arbor, MI 48109-1285 USA
kathla@umich.edu

ABSTRACT
To make research and development investments where they will have the most impact, it is critical to understand why some science and engineering gateway or portal projects change the way that science is conducted at a fundamental level in a given community. This paper provides some initial reflections on a June 2010 focus group with the goals of uncovering some of these characteristics of success and generating practical insights that draw on the strength of multidisciplinary perspectives. We identify five key tensions that are challenges to gateway sustainability, and we offer policy recommendations that could benefit future gateway projects.

Categories and Subject Descriptors
H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Data sharing, Web-based services.* H.5.3 [**Information Interfaces and Presentation (e.g., HCI)**]: Group and Organization Interfaces.

General Terms
Management, Measurement, Performance, Design, Reliability, Experimentation, Human Factors, Standardization.

Keywords
Science and engineering gateways and portals, Web interfaces, multidisciplinary projects, sustainability, funding.

1. INTRODUCTION AND BACKGROUND
More and more research begins in the digital realm. Protein structures are created electronically, networked sensor data tracks the path of a hurricane, and computer simulations model the formation of the universe. This increasing amount of digital information and the ease of sharing this information and collaborating are changing the face of science and provide many opportunities to address the most challenging problems facing scientists and engineers today. The advanced cyberinfrastructure (CI) that supports research and discovery grows ever more powerful each day, yet these resources are not as accessible as desktop computing and the Internet. Frequently, use is mediated through science gateways or portals that facilitate the process of operating and understanding these powerful tools.

A science gateway as defined here is a community-developed set of tools, applications, and data that is integrated via a portal or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a targeted community. Access to advanced CI tools through portals or science gateways can significantly increase the productivity of researchers; however, these tools must have some longevity if they are to change the conduct of science in fundamental ways. Designing effective tools requires an investment of time, effort, and money, but most projects cannot be funded at high levels in an ongoing way due to constraints on traditional funding mechanisms. Consequently, understanding the types of science and engineering problems as well as the types of communities who can most benefit from applied, persistent CI will lead to informed investment decisions and will have a broad impact on many fields.

This paper presents some preliminary results from an applied research study of portals and science gateways. Our goal was to understand why some projects change the fundamental ways that a community conducts its research while other projects do not. Through a series of innovative, cross-disciplinary focus groups, we are identifying the social and technical enablers and barriers of successful projects. This paper draws some initial conclusions from the first focus group session, conducted in June, 2010.

2. METHODS
2.1 Participants
We spent several months analyzing national and international reports on cyberinfrastructure [1,3-28,30,32-38] in order to identify a small but representative group of projects to include in the first focus group session. Representation across National Science Foundation (NSF) directorates was a goal, but we also wished to include international projects and some projects in the humanities. Participants were identified from CI reports as well as through recommendations from NSF's CI Coordination

[1] The first author is the project's principal investigator, but both authors contributed equally to this work.

Committee (CICC) and scholars in the field.

This first focus group's goal was to consider the characteristics of successful and unsuccessful gateways. Fifteen participants represented a diverse assortment of science gateways, portals, and platforms. Attendees had been or were currently directly involved in a gateway or portal-like project for science. There was considerable variety in project characteristics including domains served, user interfaces, the use of computational/data resources, the size and expertise of the user base, and the age of the project. Represented projects included:

- The Adler Planetarium (www.adlerplanetarium.org)
- CIPRES (Cyberinfrastructure for Phylogenetic Research, www.phylo.org)
- Drupal (www.drupal.org)
- Earth System Grid (www.earthsystemgrid.org)
- FLOSSmole (Free, libre, and open-source software project, www.flossmole.org)
- GISolve (www.gisolve.org)
- GridChem (www.gridchem.org)
- iPlant (www.iplantcollaborative.org)
- Linked Environments for Atmospheric Discovery (www.leadproject.org)
- MyExperiment (www.myexperiment.org)
- nanoHUB (www.nanohub.org)
- National Science Digital Library (www.nsdl.org)
- National Snow and Ice Data Center (www.nsidc.org)
- TAPoR (Text Analysis Portal for Research, portal.tapor.ca)
- VORTEX WINDS (A Virtual Organization to Reduce the Toll of Extreme WINDS on Society, www.vortex-winds.org)

2.2 Focus Group Process

The focus group was scheduled for one day, from lunchtime on the first day to noon on the second day. The session began with a warm-up exercise through which participants identified characteristics of their gateway or portal on specific dimensions, such that they could see how they fit among their colleagues and get a sense of who else was present. Characteristics uncovered included size and sophistication of the user community, age of the project, and degree to which the gateway was compute and/or data intensive.

The first idea-generation exercise focused on the question "On your project, how did you handle [topic X] that contributed to the success of your gateway/portal?" The topics included:

- 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

Each of these topics was written at the top of a flip chart. Using the "wandering flip charts" technique, we asked participants to find a flip-chart topic that interested them, discuss with others at that location, and write their ideas, including the name of their project when relevant. After a half-hour of wandering among the flip charts, the participants were asked to revisit the charts and, on a second sheet, identify the social or technical enablers that made these ways of handling these topics possible. Finally, each participant was given colored sticky dots (6 each of two colors) and asked to vote for those enablers that they considered to be most essential (must have) versus optional (nice to have). At the end, as a group, we discussed the outcomes of the exercise.

The second main idea-generation exercise focused on lessons learned using the "World Café" method. People sat in small groups at tables with paper tablecloths for note taking and they were given the following assignment:

> 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?

After 25 minutes, all but one participant from each group 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. An example of one group's tablecloth of notes is shown in Figure 1.

Figure 1. Tablecloth image representing participants' thoughts on gateway development.

The third exercise was a brainstorming activity to identify the external forces that influence the sustainability of portals/gateways, plus the opportunities and challenges that these forces present. We offered a list of possible categories that could be considered:

- 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

This exercise was conducted as a large group and led to productive discussions. We concluded the focus group by soliciting input about next steps, including future invitees (individuals and groups), and further questions to consider. All participants filled out an anonymous evaluation as well.

2.3 Analysis

The content of the focus group discussions was recorded by notetakers and the written materials were archived with photographs. All the materials were reviewed and collected into a document that was subsequently read and coded for key themes by each of the authors of this paper. The authors then compared their analysis of the data and synthesized their findings into a single set of themes. The results of the "wandering flip charts" exercise is displayed in tables in the results section, and key themes are reviewed in the discussion section that follows.

3. RESULTS

The following eight tables summarize the *enablers* identified by the focus group participants, with the number of votes each received. They have been reordered to reflect the number of votes for each enabler as "essential" and then as "optional."

Table 1. Productivity

Enablers	**Essential**	**Optional**
Effective software engineering process	4	
Constant open discussion of tools new and old and willingness to experiment/review	3	3
Quality of service guarantee	3	2
Availability of excellent software engineering tools to research organizations at no cost		1
Strong researchers prototyping innovative tools		
A clientele generally amenable to tools in their research		
Show increased productivity/profit of sufficient benefit to justify paradigm shift in corporate project management		
Innovative, unique software: Rappture toolkit and Maxwell's Daemon		

Table 2. Contents

Enablers	**Essential**	**Optional**
DOE, NSF, NASA, NOAA, JISC, EPSRC, support for data management and activities	5	3
Metadata availability	4	
Usage data	2	2
Create a sense of the use and importance of contributing. Also, not static, contents as vitality	1	1
Educating providers on the value of their data to others and placing their data in accessible formats or systems		4
Willingness by science communities to help parse/format/export data in such a way as to enable use in existing framework		2
Digital libraries, projects, e-journals, individuals		1
An international mandate to provide community service		
Disasters: motivate communities to work together to stop senseless loss of life and property		
Analytical ultracentrifuge data [specific to particular gateway]		
Establishment of format, structure of metadata conventions		

Table 3. Target Audience

Enablers	**Essential**	**Optional**
Training Ph.D. students and post docs as liaisons to home communities	3	3
Community excitement	2	1
Learning from social science/research about the user community and how they behave	1	3
Long, deep engagement with climate community	1	2
Scholarly associations, journals, libraries	1	1

Great content, e.g. Prof. Supriyo Datta Theory of Molecular Conduction (>15,000 users)		
Seeding content		
Look for which audiences seem most needy, but also respond not just to our sense of need, but the user's sense of need		
Access to resources that are unique		
Existing organizations and communities		

Table 4. Technology

Enablers	**Essential**	**Optional**
Sufficient funds for full-time professional software developers	9	2
Agile development process	3	2
Interoperability with proprietary tools "standard" in field	1	2
Pushed ourselves to get capability into user's hands as quickly as we could	1	
Advances of Web technologies for highly usable gateway environments		
Willing to throw away code		
Willing to borrow/reuse open source code		
Planning for the future: what can you reuse, what will need to be custom, making it easier to extend		
Engagement of users: try to get out of Not Invented Here Syndrome		
Drupal		
Open source projects		
Refactor for efficiency		

Table 5. Community Engagement

Enablers	**Essential**	**Optional**
Very, very, easy to use; always, always, always up (99.7%)	8	2
Stable and full-time central organization. Needs time to grow and develop. Trust must be built. Identify locus of continuing responsibility	1	
Customizable upon user interest	1	
Charismatic leader who is well known and respected in the community		5
Online help/documentation, intuitively simple interface		3
Incorporation into legally-binding standards and continuing education - licensure in discipline		
Funding specifically for training and outreach		
Credit and attribution mechanisms		
Include tools/gateway as part of coursework/workshops		
Video tutorials		
Instructors using in graduate classes		
A well-connected and dedicated PI team		

Table 6. Culture

Enablers	**Essential**	**Optional**
Desire to share best practices/tools	1	4
Ever increasing data size for analysis and limitation of desktop analysis tools	1	1
Lessen the gateway development footprint		1
Data set is too large to work alone		1
Well-respected (scientifically) champions for cause/tool		
Try to infer learning and information from all activity.		

Audience-specific sharing expectations		
Societal climate research needs necessitate collaborative work		
Changes in private sector: integrated project delivery, Building Information Modeling (BIM), and joint venture requirements		
Provide data and metadata handling		
Desire for efficiency and reproduction of work		

Table 7. Partnerships

Enablers	**Essential**	**Optional**
Community buy-in/endorsement	3	2
Excited partners - providing services, using services	3	
Clear understanding of roles of partners/how partnership functions	2	1
Gateway advanced support by TeraGrid	1	3
Frequent all-hands meetings to determine direction		2
Endorsement/leadership of domain professional organizations and standardizing bodies		1
Core meetings (frequent), open meetings to all community (annual)		1
Clear quid pro quo - or the sense that one will emerge, plus endurance in time.		
Programs like TeraGrid science gateways, other collaborative work with EU counterparts		
Universities		

Table 8. Structure of Gateway-Building Organization

Enablers	**Essential**	**Optional**
100% professional, full-time software development team with an exceptional leader and an exceptionally demanding/knowledgeable customer	7	2
Professional programmers interacting with domain specialists	2	
Charismatic, involved, evangelical leaders, supporting public/scientists/evangelists	2	
"Company culture," user is (almost) always right	1	3
Community ownership of governing body, content and editorial boards (democracy or meritocracy that rewards active users)		3
Dedicated user support in house		2
Team subscribes to the growth and vision of the gateway and has understanding of long term goals		
Multidisciplinary team		
Universities		

4. DISCUSSION

There were several recurring themes or issues emphasized throughout the session: funding, project goals, tools, community engagement, and rewards and recognition. This was evident both in the table summaries above as well as in workshop discussions. Both are covered in this discussion. Many of these themes have implications with no straightforward resolution due to the trade-offs inherent in choosing one approach versus another or the challenges of achieving certain standards. We present each of these thematic tensions in the remainder of this section.

4.1 Funding: Development vs. Operations

Development timelines—and the funding associated with them—were a source of debate across the group. For example, some participants thought that five years was too short to develop infrastructures and that staff are "smelling the barn by year three," anticipating their exit and becoming less attentive to the thoroughness and quality of their work. Others thought it was imperative to show results much sooner and that this could be done more easily by employing professional software developers (Table 4, 8). Some observed that funding agencies may need to designate projects in particular ways to fund them for more than 5 years, for example the Major Resource Equipment and Facilities Construction (MREFC) or the Science and Technology Centers at the NSF.

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. Clearly, different projects have faced different development environments and have required different funding strategies.

Likewise, some participants thought costs were higher up front for development, while others were mindful of the constant need to adapt (e.g., changing data volumes, changing technologies, changing user goals) and therefore felt that the need for continued development funding as well as maintenance funds is essential. Connected to this issue is the option, as some gateways have done, to rebuild tools from the bottom up when they see the audience needs and the technology are changing. This, too, shifts how funding might be distributed.

Regardless, participants widely agreed that it is useful to divide the development timeline into periods with synchronized funding—for example, a requirements-analysis period, a development period, a testing period, an operations period, a reevaluation period, etc.—each with its own funding stream. It may be that this is too complex a funding structure to be implemented by agencies, but perhaps calling for descriptions of such periods with intermediate goals would be helpful in the evaluation of gateway-building proposals.

Many found that funding the initial stages of a project can be considerably easier than funding the more operational or maintenance phases. Participants lamented the challenge of getting funding for "unsexy" things—software development that is perceived as "easy," not cutting edge, or maintenance of existing useful tools. In order to fund such activities it will be important to capture the significant impact they have, for example statistics showing that large numbers of researchers rely on a gateway daily to conduct their work more effectively or large numbers of papers citing the use of the gateway. This requires a certain important focus by the gateway principal investigator (Table 8).

4.2 Project Goals: Research vs. Production

A second thematic tension was the contrast between research versus production. Most portals and gateways are funded as research projects, yet they require the construction and maintenance of infrastructure in order to support the research productivity of their target community. Both academic institutions and the NSF reward "new" research; young professors receive tenure through publications on new topics while NSF primarily funds basic research. As one participant noted, "Building a new road is charismatic; pothole fixing is not."

This tension between research and production manifests itself in several ways. Gateway development efforts generally involve cross-disciplinary teams. If both the developers and the end users are expecting to use the gateway to promote research interests there may be a conflict. Developers may address new research topics in the course of gateway design in order to further their academic goals. Resulting gateways may be more complex than necessary, less reliable, and may not meet the goals of the domain science community for whom they were designed. Focus group participants noted that sometimes simple tools are all that is needed to enable cutting edge science, but we "make the easy things hard." Participants felt it was necessary to focus on what works, not what is technologically interesting as shown by the prioritization given to quality of service in Table 1 and also to provide robust content, for example data collections that follow metadata standards developed by a community and professionally curated (Table 2).

Analogies to industry arose during the course of the discussions. It was noted that the research and development arm of a company is often separated from the production arm. The two have different goals and different metrics of success by which they are judged. It may benefit gateway programs to be split into components, each with its own success criteria. Gateways might have success criteria that emphasize both their utility to the community (number of users, up time, user satisfaction ratings) as well as their utility in enabling research (number of citations).

Participants also made the analogy between the NSF and a venture capitalist, who funds early concept work but with a roadmap to productization. Some participants mentioned the possibility of establishing connections to mission-oriented agencies such as NASA and NOAA as a means of productization, tightening inter-agency connections in areas where a gateway could have mission-specific uses once it reaches maturity.

4.3 Tools: Standardized vs. Open-Source vs. Custom

Participants expressed several different viewpoints related to the construction of gateways, specifically who should build them and how. Because most science and engineering gateways are housed in academic settings, students are more plentiful than professional software developers. Likewise, the academic settings (and their associated budget constraints) favor using inexpensive tools and techniques for building gateways. The question is whether such methods are ultimately more cost effective and whether they jeopardize the success of the projects.

For example, several participants cited the use of professional software developers (Table 4, 8), a well-developed software engineering process, and industry-grade software as being critical to their success (Table 1). An identified core development group and in-house user support were essential (Table 8), though sometimes that group may function within a larger consortium. Unfortunately, participants also stressed that good developers were hard to find.

As an alternative, some cited the use of open-source tools as a way to keep costs low (Table 4), however many said that relying on such tools can incur additional development costs. For example some tools—particularly those that are also created through research grants—-may change platforms unexpectedly or do not support backwards-compatibility between versions. Likewise, open-source products can be left hanging and may be unreliable as the original developers lose interest. Participants warned that there are no guarantees that open-source code will be robust and sustained and cited the inherent difficulty of knowing about relevant software for their applications. Several emphasized the need to interoperate with both industry and open-source tools that users already trust and use in their day-to-day work (Table 4). Examples mentioned included Excel, Matlab, and Moodle.

Finally, some felt that willingness to tear everything apart and start from scratch was sometimes necessary in gateway development, while others thought that having a plan from the outset that supported software re-use was important (Table 4). We suspect that some combination of all of these scenarios may be appropriate depending on the project. They key is to retain a close connection to the users and their goals as a way of motivating choices about tools and techniques (Table 3, 4).

4.4 Community Engagement: Delivering What the Users Want

Not one of the participants disputed the value and necessity of building gateways and portals that are focused on usability and reliability. As one participant said, the best way to engage a community is if the gateway is "Very, very, very easy; always, always, always up." The primary tension represented by this pervasive theme is the question of how feasible it is to build a site that fully meets the community's needs. Most gateway projects aspire to make their users say, "Gosh, I was relying on that gateway for my work. If it is gone, I'm unable to do my work." Participants observed that many projects get close to this ideal but don't quite reach it before their funding ends and a new project starts.

One way that projects work toward this ideal is by keeping the users' concerns addressed, studying usage patterns (Table 5, 6), and using technology to develop highly usable systems such as user-centric Web environments with start-to-finish capabilities. The most successful gateways address usability from the outset and have "charismatic, involved, evangelical leaders " (Table 8).

Gateway development begins with real user experiences but must be prepared to adapt. Gateways and tools must align with a user's usual workflow and not add extra steps. One participant noted that the existence of cyberinfrastructure will change communities just as the arrival of roads changed communities. Another stated "Deploying the tool fundamentally changed way that people could work, which necessitated restarting from the ground up for the next version; reconfigured the whole idea of user-centricity. They [the developers] *thought* they were user-centric to start, but the definition/reality changed over time."

Retaining a focus on content is important as well, as some participants have seen that it can be easy to alienate researchers. Highly desirable content from respected researchers is important (Table 3). Content that is "very Web 2.0," such as user-contributed tagging, rating, and credit and attribution systems (Table 5) is also valued. Also key are agency support for data management, common metadata standards, and the willingness of a community to adhere to specific formats and share data (Table 2, 6). Along these lines, some projects hired gatekeepers to make sure that the metadata and data curation were properly structured and maintained. One participant noted that "gardens have to be tended every day. We need gardeners, and it is hard to fund this."

However, while technology can be hard to implement, some participants expressed their surprise at discovering how much *more* difficult it was to understand humans, particularly to engage them in using a gateway. Many portals must reconcile and welcome different types of users, such as casual users and contributors, who have different needs and goals. Some gateways may need to incentivize their contributors or the users who serve as early testers. To develop community interest in a gateway, participants described targeting students, connecting with scholarly associations (Table 7), and leveraging long, deep involvement with the community (Table 3). Other tactics included providing popular products from well-known scientists (Table 3) and partnering with leaders in industry, government, and education with credibility to speed technology adoption. Beyond these approaches, gateways hope that if tools are intuitive, an ecosystem will spring up around them and community support will develop naturally.

4.5 Rewards & Recognition: Traditional vs. New

A final tension is how to reconcile traditional, established forms of academic recognition with the need for new incentives to motivate gateway development. 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 evaluation.

Participants discussed how the publishing reward system might itself be changing because of the arrival of gateways. Gateways have accelerated the pace of progress, so old metrics of professional success and advancement such as journal publications and citations may be supplanted by new methods and products. Just as digital resources are forcing the publishing industry to adapt—for example, some books are outdated as soon as they are published—academic tenure and promotion may, too, have to adapt. One participant describes how contributions to a gateway such as highly rated curricula, tutorials, and applications or answers to other users' questions can demonstrate a certain value for those seeking tenure. Unfortunately, universities are slow-moving creatures, and revision of the tenure and promotion system does not appear to be changing anytime soon, but that does not rule out consideration of appropriate mechanisms for recognition of contributions through gateways.

So, what reward mechanisms are needed for longer term infrastructure development? First, we need to identify what metrics we should use to measure success. Metrics can also help funding agencies evaluate the sustainability of a gateway and whether it merits additional funding. Some examples of metrics suggested by participants include:

- Number of users
- Citations of the gateway as a tool used for research (much as software is cited)
- How data is being used (publications based on data accessed through the gateway in some way, either as a dataset or through computational resources provided by the gateway)
- Data integrity: Degree to which a respected curation process is used in making datasets available in the gateway
- How money is saved or efficiencies are gained (e.g., enhancing efficiency of individuals using the gateway)
- Availability of a gateway (uptime)
- Alignment with the user community to be served (e.g., through community surveys, or perhaps percentages of users funded by a given directorate or percentage that are members of a given professional society). It is important to measure both user satisfaction, but also the fraction of the user community that is served.
- Is it mission-driven? It needs to provide an obvious service. This could also be assessed through community surveys.

One participant noted that “ideas that are good will take off; we need to stop thinking about how research or scholarship should be.” In the same way, gateways are often good ideas that go beyond current notions of what research or scholarship *should* be and forge new directions and paradigms.

5. RECOMMENDATIONS

One goal of our study is to provide useful evaluation criteria to support longer-term funding decisions. An outcome of this focus group was that we identified other initiatives and opportunities through which NSF and other funding agencies could proactively support the sustainability of science gateways and portals. We list these in the following sections.

5.1 Funding: Research Enabling Research

5.1.1 Recognizing Gateways as Enablers of Research

NSF must consider the value of gateways as enablers of research, not just as research projects themselves. NSF’s strategic outcome goals of discovery, learning, research infrastructure and stewardship [29] indicate a recognition of the importance of infrastructure. We believe that the Foundation needs to focus on funding proven, successful gateways that enable new research in a sustained way.

5.1.2 MREFCs as a Model for Long-Term Funding

Participants recognized that for many NSF programs, funding beyond 5 years is not possible. Perhaps some components of long term funding programs such as Major Research Equipment and Facilities Construction (MREFC) could be used as a model. For example, an FY10 NSF budget description says:

> In order for a project to be considered for MREFC funding, NSF requires that it represent an exceptional opportunity that enables research and education. In addition, the project should be transformative in nature in that it should have the potential to shift the paradigm in scientific understanding and/or infrastructure technology... Projects under consideration for MREFC funding must undergo a multi-phase review and approval process.... As a general framework for priority setting, NSF assigns highest priority to ongoing projects, which are those that have received funding for implementation and where outyear funding for the full project has already been included in a Budget Request to Congress [31].

There are a number of components of this that may be tailored for a sustained gateway program: the potential for paradigm shifts, the multi-phase review and approval process, and the continuation of ongoing projects. Participants mentioned a multi-phase funding process that is used in NSF’s Directorate for Education and Human Resources Informal Science Education program [2] which includes pilot studies, followed by planning, implementation, and evaluation. One could see this cycle continuing. As one participant said, “‘build once, use forever’ is not realistic. Funding is needed to address evolving needs.”

5.1.3 Partnerships and Novel Industry Connections

Participants also mentioned fees, such as micropayments, but it’s not clear that this would generate sufficient funding to sustain a gateway or that academic researchers using a gateway would have mechanisms to cover such costs. Nevertheless, the idea of partnerships and novel connections with industry were repeatedly raised as opportunities for funding gateways, particularly beyond their initial exploratory, development phase.

5.2 Project Goals: An Avenue for Entrepreneurship

5.2.1 Sustainability Through “Startup” Support

Gateways could follow a transition model based on how NASA and NOAA manage projects to make sure that projects persist. NSF needs to consider a more entrepreneurial mindset that provides more of a bridge between the initial idea and the outcome that needs to be sustained. Sustainability models need to be built in from the beginning, but researchers need logistical, “incubation” support for that as they are not trained in the business of “startups.” Typically, the visionary leader with the charisma to get a project started and engage a community is not the same person who knows how to make operations real and sustainable. NSF needs to encourage gateways to identify both those kinds of leaders along the way.

5.2.2 Connections to Potential Partners

Intra-agency (e.g., across directorates) and cross-agency support (e.g., with DOE, NASA, NOAA, Department of Commerce) may be needed, along with connections to industry and international partners, to encourage a sustainable pathway as well as productization.

5.2.3 Learning from Failures

As part of the “entrepreneurial support” structure for new gateways, NSF needs to anticipate a percentage of failures, accompanied by a proper process for “burying” those failures. Failures will be inevitable, but even so, we need to capture lessons from those and retire them in an appropriate way in order to better inform subsequent gateway projects.

5.3 Tools: Supporting the Builders

5.3.1 Training Future Professionals

As we noted, good software developers are hard to find, but gateways might provide an opportunity for students in the educational pipeline to get experience with this type of computational science. This would be a supplement (or alternative) to including dedicated professionals in the software construction, maintenance of metadata and data curation, and ongoing support and maintenance.

5.3.2 A Gateway Consortium

Participants were also enthusiastic about the idea of forming a gateway consortium (or a social network, association, or federation) that would provide a software repository with tagging and reviewing capabilities. Additionally, such a consortium would be a place where lessons learned are captured and keys to sustainability aren’t lost or reinvented over and over.

5.3.3 Stable Software Foundations

Many gateway projects rely on underlying software layers that are themselves research projects. Unfortunately, these projects are often released and then redesigned in the name of research. Gateway developers are eager for foundational software and platforms that either do not change often or that have compatible interfaces across versions.

5.4 Community Engagement: A Multidisciplinary Affair

5.4.1 Multiple Avenues for Creating Usability

Any new project leadership team needs to recognize that the developers may not be the best equipped to understand the users nor to design with usability in mind. Projects need to engage the community in multiple ways to get them to try the site and come back, and social scientists can help with that. For example, Unidata engages "user committees" that meet regularly and help determine the direction of infrastructure development.

5.4.2 Gathering Metrics

Another obvious opportunity is to instrument a gateway so that usage can be analyzed. For example, metrics gathered through instrumentation could answer "What are people doing when they use a tool? What is slowing people down right now?"

5.5 Rewards and Recognition: Rethinking Incentives

Although NSF cannot single-handedly push reform of tenure and promotion, they can consider how to introduce standards that could be used as ways of reward infrastructure development. Likewise, they could offer ways of documenting meaningful participation in a gateway so that researchers can list their work in a way that is analogous to publication.

6. CONCLUSION AND FUTURE WORK

This study offers a small first step toward identifying the pathways to the sustainability of science and engineering gateways. We found five key areas of concern: funding, project goals, tools, community engagement, and rewards and recognition. Participants offered many generative ideas that could help alleviate some of the tensions associated with these issues.

Immediately after the focus group described here, we conducted a second focus group with the goal of identifying disciplinary areas that are ripe for the development of a science gateway. With that data in hand, plus the results of the focus group described here, we intend to conduct two additional focus groups to elaborate further on our understanding of these issues and to provide additional resources and guidance to NSF and, we hope, the science and engineering gateways that they support.

7. ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under Grant No. OCI-0948476. We wish to thank our NSF Program Officers Susan Winter and Jacqueline Meszaros and the other awardees on this "Catalyzing Virtual Organizations" EAGER grant who provided substantive guidance as we developed this project. The early seeds of this project were developed through a white paper that Nancy co-authored with Kelvin Droegemeier, Jay Alameda, and Dennis Gannon. Cory Knobel offered many useful suggestions on the identification of participants and design of the focus group process. Other helpful recommendations of participants were provided by Mark Parsons, Annamaria Carusi, Kevin Schawinski, Phillip Westmoreland, Ruth Pordes, and Peter McCartney. Melissa Chalmers provided outstanding note-taking support during the focus group. Of course, most of all, these insights would not be possible without the thoughtful and enthusiastic participation of our focus group participants, who attended as representatives of their projects.

8. REFERENCES

[1] 2004. An Operations Cyberinfrastructure: Using Cyberinfrastructure and Operations Research to Improve Productivity in American Enterprises. Retrieved on 09/13/10 from http://www.optimization-online.org/OCI/OCI.pdf.

[2] 2008. NSF's Informal Science Education Program Solicitation. Retrieved on 9/21/10 from http://www.nsf.gov/pubs/2008/nsf08547/nsf08547.htm.

[3] Ad Hoc Committee for Cyberinfrastructure Research, Development and Education in the Atmospheric Sciences. 2004. Cyberinfrastructure for the Atmospheric Sciences in the 21st Century. Retrieved on 09/12/10 from http://www.cisl.ucar.edu/cyrdas/cyrdas_report_final.pdf.

[4] Ainsworth, S., Honey, M., et al. 2005. Cyberlearning Workshop Series. Retrieved on 09/13/10 from http://archive.cra.org/Activities/workshops/cyberlearning.

[5] Atkins, D. 2005. Cyberinfrastructure and the Next Wave of Collaboration. Retrieved on 09/13/10 from http://deatkins.blogs.com/dan_atkins_weblog/cyberinfrastructure_initiatives/Atkins%20%20Educause%20Final.pdf.

[6] Atkins, D., et al. 2003. Revolutionizing Science and Engineering Through Cyberinfrastructure: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure retrieved on 9/18/10 from http://www.nsf.gov/od/oci/reports/atkins.pdf.

[7] Atkins, D., Butler, R., Feldman, S., Foster, I., Kesselman, C., Livny, M., and Stevens, R. 2005. Planning for Cyberinfrastructure Software. Retrieved on 09/13/10 from http://www.nsf.gov/od/oci/ci_workshop/index.jsp.

[8] Awards made from NSF's Engineering Virtual Organizations (EVO) program. Retrieved on 9/18/10 from http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501057.

[9] Backlund, P., et al. 2002. Cyberinfrastructure for Environmental Research and Education. Retrieved on 09/13/10 from http://www.ncar.ucar.edu/cyber/cyberreport.pdf.

[10] Berman, F. and Brady, H. 2005. Workshop on Cyberinfrastructure and the Social and Behavioral Sciences. Retrieved on 09/13/10 from http://www.sdsc.edu/about/director/pubs/SBE/reports/SBE-CISE-FINAL.pdf.

[11] Carusi, A., and Reimer, T. 2010. Virtual Research Environment Collaborative Landscape Study. Retrieved on 9/18/10 from http://www.jisc.ac.uk/media/documents/publications/vrelandscapereport.pdf.

[12] Crawford, D., et al. 2007. NSF's Cyberinfrastructure Vision for 21st Century Discovery. Retrieved on 9/18/10 from http://www.nsf.gov/pubs/2007/nsf0728/index.jsp.

[13] Cummings, Jonathon, et al. 2008. Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations. Report from the

Building Effective Virtual Organizations workshops. National Science Foundation. Retrieved on 09/11/10 at http://www.ci.uchicago.edu/events/VirtOrg2008/VO_report.pdf

[14] Davis, J., Ahalt, S., Doyle, F., Evans, L., Grossman, I., and Kim, S. 2006. Workshop on Cyberinfrastructure in Chemical and Biological Systems. Retrieved on 9/18/10 from http://www.oit.ucla.edu/nsfci/default.htm.

[15] Deelman, E., and Gil, Y. 2006. Workshop on the Challenges of Scientific Workflows. Retrieved on 9/18/10 from http://vtcpc.isi.edu/wiki/index.php/Main_Page.

[16] Dolbow, J., Khaleel, M., and Mitchell, J. 2004. Multiscale Mathematics Initiative: A Roadmap. Retrieved on 09/13/10 from http://www.sc.doe.gov/ascr/mics/amr/Multiscale%20Math%20Workshop%203%20-%20Report%20latest%20edition.pdf.

[17] Edwards, P., Jackson, S., Bowker, G., and Knobel, C. 2007. History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures. Retrieved on 9/18/10 from http://www.si.umich.edu/pne/PDF/ui.pdf.

[18] Frischer, B., Unsworth, J., Dwyer, A., Jones, A., Lancaster, L., Rockwell, G., and Rosenzweig, R. 2005. Summit on Digital Tools for the Humanities. Retrieved on 9/18/10 from http://www.iath.virginia.edu/dtsummit/SummitText.pdf.

[19] Hart, D. Cyberinfrastructure - A Special Report. Retrieved on 9/18/10 from http://www.nsf.gov/news/special_reports/cyber/index.jsp.

[20] Head-Gordon, T., and Gordon, M., et al. 2005. CHE Cyber Chemistry Workshop. Workshop Report. National Science Foundation. Retrieved on 09/11/10 from http://bioeng.berkeley.edu/faculty/cyber_workshop.

[21] Kansas Geological Survey Report, 2004, Geoinformatics: Building Cyberinfrastructure for the Earth Sciences (2004). Retrieved on 09/13/10 from http://www.geoinformatics.info/.

[22] Keyes, D., et al. 2003. A Science-Based Case for Large-Scale Simulation. Retrieved on 9/18/10 from http://www.pnl.gov/scales/docs/volume1_72dpi.pdf.

[23] Loewenstein, R., Markowsky, G., et al. 2003. A Workshop Sponsored by the National Science Foundation OPP/CISE. Polar Science and Advanced Networking: workshop. Retrieved on 9/15/10 from http://www.polar.umcs.maine.edu/.

[24] Marlino M., Sumner T., and Wright M. 2004. Geoscience Education and Cyberinfrastructure, Digital Library for Earth System Education. Retrieved on 09/13/10 from http://www.dlese.org/documents/reports/GeoEd-CI.pdf.

[25] Maron, N., Smith, K., and Loy, M. 2009. Sustaining Digital Resources: An On-the-Ground View of Projects Today. Retrieved on 9/18/10 from http://www.ithaka.org/ithaka-s-r/research/ithaka-case-studies-in-sustainability/report/SCA_Ithaka_SustainingDigitalResources_Report.pdf.

[26] Mathematical and Physical Sciences Directorate of the National Science Foundation. 2004. Identifying Major Scientific Challenges in the Mathematical and Physical Sciences and their CyberInfrastructure Needs. Retrieved on 09/13/10 from http://www.nsf.gov/attachments/100811/public/CyberscienceFinal4.pdf.

[27] McCurdy, B., et al. 2001. Computation as a Tool for Discovery in Physics. Report to the National Science Foundation by the Steering Committee on Computational Physics. Retrieved on 09/12/10 from http://www.nsf.gov/pubs/2002/nsf02176/start.htm.

[28] Mish, K., et al. 2003. Cyberinfrastructure for Engineering Research and Education.. Retrieved on 09/13/10 from http://www.nsf.gov/eng/general/Workshop/cyberinfrastructure/index.jsp.

[29] National Science Foundation Investing in America's Future Strategic Plan FY 2006-2011. Retrieved on 09/13/10 from http://www.nsf.gov/pubs/2006/nsf0648/nsf0648.jsp.

[30] Novotny, M., Ceperly, D., Jayanthi, C., and Martin, R. 2004. Materials Research Cyberscience enabled by Cyberinfrastructure. Retrieved on 09/13/10 from http://www.nsf.gov/mps/dmr/csci.pdf.

[31] NSF FY10 MREfC budget. Retrieved on 9/21/10 from http://www.nsf.gov/about/budget/fy2010/pdf/30_fy2010.pdf.

[32] Powell, T., et al. 2004. Trends in IT Infrastructure in the Ocean Sciences. Retrieved on 09/18/10 from http://www.geo-prose.com/oceans_iti_trends/oceans_iti_trends_rpt.pdf.

[33] Simpson, T., Lewis, K., Chen, W., et al. 2006. Exploring Cyberinfrastructure Themes in Engineering Design. Retrieved on 09/12/10 from http://www.mne.psu.edu/simpson/NSF/EXCITED.

[34] Trouvé, A., Haworth, D., Miller, J., Su, L., and Violi, A. 2006. Workshop on Cyber-Based Combustion Science. Retrieved on 9/18/10 from http://www.nsf-combustion.umd.edu/.

[35] Walker, J., Carlson, R., et al. 2003. CyberInfrastructure (CI) for the Integrated Solid Earth Sciences (ISES). Retrieved on 09/18/10 from http://www.paleostrat.org/Documents/ises-ci%202003.pdf.

[36] Welshons, M., editor, et al. 2006. Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. Retrieved on 09/11/10 from http://www.acls.org/cyberinfrastructure/cyber.htm.

[37] Wilkins-Diehr, N., Alameda, J., Droegemeier, K., and Gannon, D. 2008. The Importance of Long-Term Science and Engineering Infrastructure for Digital Discovery. Retrieved on 9/18/10 from http://www.sciencegateways.org/sgwwhitepaper.

[38] Wooley, J. et al. 2003. Building a Cyberinfrastructure for the Biological Sciences. Workshop Report. NSF Directorate for Biological Sciences Advisory Committee (BIOAC). Retrieved on 09/11/10 from http://research.calit2.net/cibio/archived/CIBIO_FINAL.pdf