Who Cares About Science Gateways? A Large-Scale Survey of Community Use and Needs

Katherine A. Lawrence
School of Information
University of Michigan
kathla@umich.edu

Nancy Wilkins-Diehr
San Diego Supercomputer Ctr., Univ. of Ca., San Diego
wilkinsn@sdsc.edu

Julie A. Wernert
Research Technologies
Indiana University
jwernert@iu.edu

Marlon Pierce
Research Technologies
Indiana University
marpierc@iu.edu

Michael Zentner
Rosen Center for Advanced Computing, Purdue University
mzentner@purdue.edu

Suresh Marru
Research Technologies
Indiana University
smarru@iu.edu

ABSTRACT

With the rise of science gateway use in recent years, we anticipate there are additional opportunities for growth, but the field is currently fragmented. We describe our efforts to measure the extent and characteristics of the gateway community through a large-scale survey. Our goal was to understand what type of support services might be provided to the gateway community.

Categories and Subject Descriptors

H.3.4 [**Information Systems**]: Systems and Software – Distributed systems; J.0 [**Computer Applications**]: General; K.4.3 [**Computers and Society**]: Organizational Impacts – Computer supported collaborative work; K.6.3 [**Management of Computing and Information Systems**]: Software Management – Software development.

General Terms

Management, Design, Human Factors

Keywords

Science/engineering gateways/portals, Web interfaces, software development, high-performance computing, cyberinfrastructure

1. INTRODUCTION

Over the last two years, science gateways have passed important milestones. Gateways incorporate advanced resources—high-performance computers, data streams from sensors or instruments, curated data collections, and the software to capture, organize, analyze, and visualize this data—through web or mobile interfaces. Historically, most users accessed these resources by downloading and maintaining their own software or using complex programming languages through a command-line interface. Both the NSF funded XSEDE project (https://www.xsede.org/gateways-listing) and the DOE-funded National Energy Research Scientific Computing Center (NERSC; http://portal.nersc.gov/) now report that the number of users accessing their resources via science gateways surpasses the number of users accessing resources via the command line. Science gateways such as Galaxy [1] and NanoHUB [3] have thousands of regular users. Also, significant effort has been placed into frameworks and infrastructure to support gateway development at scale, including HUBzero [5], iPlant [6], and Apache Airavata/SciGaP [4] in the US and WS-PGrade/SciBUS [2] in the EU.

While these are important milestones that indicate the importance and health of science gateways in research and development, we believe there is much room for additional growth. To investigate this hypothesis and to measure, for the first time, the full extent and characteristics of the gateway community, we have undertaken a community survey. Our goal in conducting this survey was to understand what type of support services might be provided to the gateway community by a center of gateway expertise. To our knowledge, this is the largest such survey on this topic.

2. SURVEY DESIGN

In developing questions for the survey, we took an inductive approach, beginning with in-depth interviews of experts. These experts helped us identify participants and questions for a series of focus groups. Interactions at the focus groups refined the questions we wanted to pose to the very wide survey population. During a 7-month planning effort, we developed 36 questions that branched in different ways depending on whether a recipient was an administrator, researcher or faculty member, or technology developer. Because a gateway can be a large effort with many stake-holder interests, it was important to get the perspective of each key stakeholder type.

We asked survey participants about the importance of these inter-faces in their fields and also about their participation in building such interfaces: What roles did people play in the planning, build-ing and operation of web applications? What types of skills were needed in the projects?

2152-1093/14 $31.00 © 2014 IEEE
DOI 10.1109/GCE.2014.11

IEEE computer society

What types of skills were missing? What types of support services might benefit from a community offering? These are some of the questions investigated in our survey.

2.1 Population Sample

The survey sample was collected from three primary sources: National Science Foundation–funded principal investigators (PIs) (90% of sample), senior administrative members of EDUCAUSE and CASC (6%), and individuals who have previously expressed interest in gateway initiatives (4%). The NSF PIs were limited to those who had received funding within the last 18 months for at least $100,000. We focused primarily on NSF PIs because our conceptualization grant was to identify interest within the US-based NSF-funded community, and the timing and dollar criteria were selected to ensure that the PIs were active and that the grants were not small, workshop-type funding. Our team would like to direct future efforts toward incorporating input from those funded by other Federal agencies and from international communities. The individuals with prior interest in gateways include participants in focus groups and workshops plus volunteers on our website. The total sample size was nearly 29,000, and our more than 4,300 participants represent a response rate of approximately 15%. This exceeded our 10% target rate of response.

2.2 Implementation

Participants were invited to participate by email. Initial invitations were sent in late May 2014. They could opt out by visiting the survey site and indicating that they did not wish to participate. Those who did not opt out received reminders until they participated or until the survey closed in July. The maximum number of contacts was four, including the initial invitation, two general reminders, and a final one that the survey was closing.

3. RESULTS AND DISCUSSION

The high degree of community interest has been reflected by the number and variety of responses to our survey. More than 4,300 people responded, representing a broad range of disciplines (Figure 1), including the humanities (e.g., visual and performing arts, history, and linguistics) and social and communications sciences (e.g., economics, geography, and anthropology). The majority of respondents represent fields in the life, physical and mathematical, and computer and information sciences. Engineering and environ-mental sciences were also well represented.

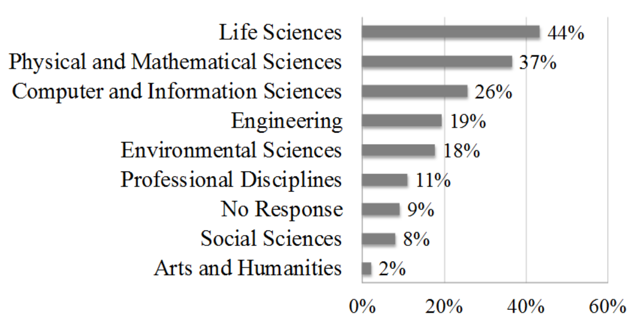

Figure 1: Primary areas of current domain expertise. Respondents could select all that apply; 91% of 4,382 responded, generating 7,376 total responses (mean=1.82 domains/respondent).

Respondents were comprised primarily of faculty and research scientists, but also included members of higher education leader-ship, graduate students, and technology developers. Some 57% of respondents report having participated in some capacity in the creation of desktop, mobile, or web applications. An additional 8% who have never participated in the creation of applications indicate that they hope to do so in the future.

The respondents who have participated in development projects have served in multiple roles, ranging from principal investigator to web designer to outreach and education specialist. The developer group (n=2471) selected 4881 roles, producing a mean of 2 roles per respondent. The aggregate responses indicate a breadth of experience, but skew heavily toward the perspective of principal investigator (79%), with the next most common roles being domain-based experts (or content specialist) and advisory board or steering committee members (22% each). The remaining roles (in order of frequency) were graduate-student or post-doctoral programmer (18%), website or user experience designer (15%), professional software developer or technology specialist (13%), outreach or community/educational engagement specialist (12%), and user support team member (10%).

That said, projects employ many different types of people. We provided a list of eight common types of staff members on software development projects and asked participants to indicate whether, on their projects they (a) had this type of staff, (b) wished they had this type of staff, or (c) did not need this type of staff (Figure 2). Student or post-doc programmers were by far the most prevalent (65%), followed by project managers (43%) and professional software developers (41%). Least available but most desired were quality assurance/testing experts and usability consultants. Graphic designers were prevalent but also needed. The presence or absence of these types of staff members may reflect the types of staff that are commonly funded at academic institutions. In addition to the eight types of staff members, we asked respondents to indicate any other roles. Some of the more common roles indicated across the diverse responses included content or domain experts; instructional designers; technical writers; librarians; computer scientists; software or system architects; and IT support.

Even more interesting are the wide variety of capabilities inside science gateways (Figure 3). Education tools (18%), computation-al tools (16%), data analysis tools (including those for visualization and data mining; 16%), and data collections (15%) are the most common. Future analysis of our data will look at possible correlations between capabilities created and specific domains.

Within our sample, we asked specific questions of the respondents who had participated in creating web- or mobile-based applications. We were interested in how they anticipated needing help with their development projects. Many indicated a high interest in help with many of the functions associated with build-ing a gateway (Table 1). The top two areas of interest include evaluation, impact analysis, and web analytics (53%) and planning how to adapt technologies (49%). Other services of interest include usability, web/visual/graphic design, choosing technologies, and visualization (each 48%), echoing some of the staff member roles that respondents wished they had. In most areas of potential support, at least 40% indicated that some help might be needed.

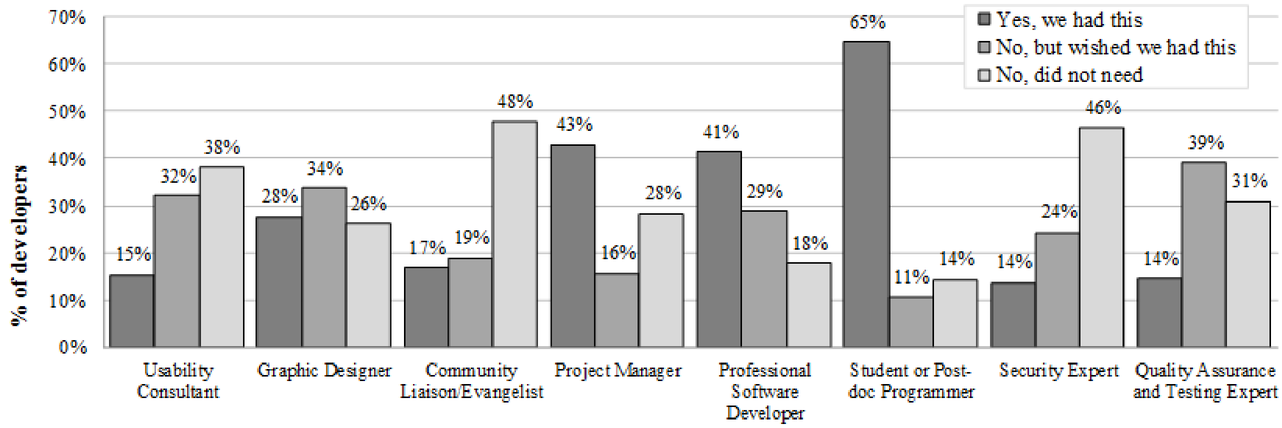

Figure 2: Types and desirability of staff members who have worked on projects (n=2471, 84-90% response rate across types)

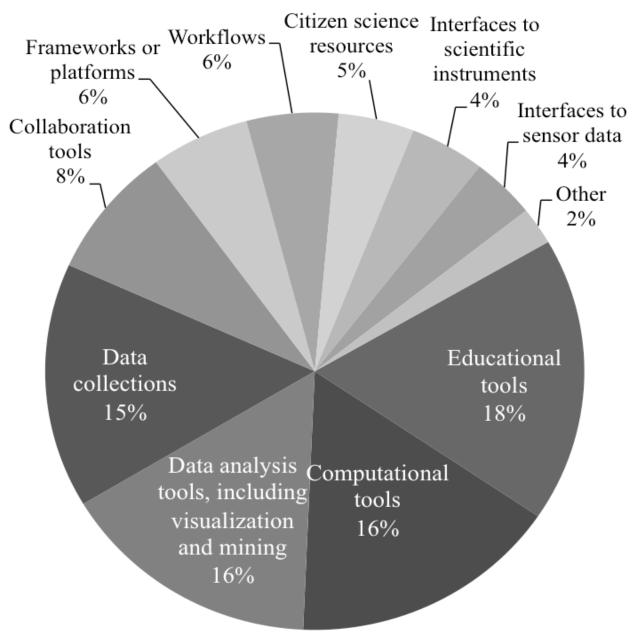

Figure 3: Types of applications created by respondents (n of application types=6581, by 2299 developers; mean=2.9 application types/developer)

When technology developers were asked what they use to build their web- or mobile-based applications, Drupal, Ruby on Rails, and WordPress were most commonly cited, but nearly 200 other development platforms, frameworks, and applications were cited at least once, including DreamWeaver, Java/ JavaScript, Python, and php/MySQL, as well as "home-grown" codes. This suggests that providing a one-size-fits-all, or even a one-size-fits-most, solution is not feasible; instead, a technical community forum should be fostered to share and extend these solutions in a collaborative way.

Finally, we asked web- and mobile-application developers what mechanisms they prefer for training their staff. They were allowed to indicate up to three preferences. Self-paced, online learning was by far the most popular (38%), followed by workshops or short courses (31%). Webinars (25%) and on-site custom training (22%) are also reasonably popular options. It is possible that less popular options such as MOOCs (Massive Open Online Courses; 12%) may become more popular for online learning as this teaching method grows more common, particularly if it could deliver the convenience of self-paced online learning with the specificity and instructor presence of a workshop.

The survey focused not just on those involved in developing web applications, but also attempted to assess the importance of applications for providing access to specialized resources. We asked those who identified as researchers and/or educators how important to their work were the Web-based applications providing access to specialized resources (Table 2). For accessing most types, at least 50% indicated that web-based applications were "somewhat" or "very" important.

4. CONCLUSION & FUTURE WORK

This extensive survey indicates that gateways are an active part of the science and engineering research and education landscape. Scientists and educators depend on web-based applications to access resources. The developers of these applications require an extensive array of skills and expertise to deliver their products. People who can provide these skills and expertise are not always available. Enabling these projects to learn from experts and access specialists is perceived as a valued and needed service. Because we have only just collected and cleaned our data, we plan to do additional analysis, looking at more complex associations. For example, we may investigate:

- What types of gateway resources are most popular with specific fields of study? This could include resources created or used by survey respondents, as they were asked as two separate questions.
- What fields are good candidates for large-scale gateways? What fields are currently not in need of support?
- Who builds the resources used in particular domains (e.g., commercial suppliers, public or academic institutions, or by the researchers themselves)? Do certain resources tend to be provided by certain types of sources?

Table 1: Percentage of mobile- or web-based application developers who would seek at least some help from a service provider (n=2471)

Proposed Service	Interest
Evaluation, impact analysis, website analytics	53%
Adapting technologies	49%
Usability services	48%
Web/visual/graphic design	48%
Choosing technologies	48%
Visualization	48%
Developing open-source software	47%
Support for education	47%
Community engagement mechanisms	46%
Keeping your project running	45%
Legal perspectives	45%
Managing data	44%
Mobile technology development	43%
Database structure, optimization, and query expertise	43%
Computational resources	43%
Data mining and analysis	42%
Cybersecurity consultation	42%
Website construction	41%
Software engineering process consultation	39%
Source code review and/or audit	38%
High-bandwidth networks	33%
Scientific instruments or data streams	32%
Management aspects of a project	28%

- What "generic" technologies are most needed by gateway developers?
- What are the prevalence and relevant capabilities of mobile devices for accessing high-end gateway-type resources? Can mobile devices better serve certain disciplines or resources than others?
- What development roles are most commonly staffed together and which have been needed but not employed?
- Where do people learn about new technologies and how do they decide to adopt them?
- What are the biggest challenges to hiring and maintaining gateway development staff?

5. ACKNOWLEDGMENTS

The entire team of the Science Gateway Institute contributed to the development of the survey described in this paper. In addition to the authors, Rion Dooley, Linda Hayden, Michael McLennan, and Dan Stanzione offered their insight and ideas to the survey design. This material is based upon work supported by the National Science Foundation under Grant Numbers 1216668, 1216723, 1216730, 1216733, 1216815, and 1216870.

6. REFERENCES

[1] Jeremy Goecks, Anton Nekrutenko, James Taylor, et al. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*, 11(8):R86, 2010.

[2] Peter Kacsuk, Zoltan Farkas, Miklos Kozlovszky, Gabor Hermann, Akos Balasko, Krisztian Karoczkai, and Istvan Marton. Ws-pgrade/guse generic dci gateway framework for a large variety of user communities. *Journal of Grid Computing*, 10(4):601–630, 2012.

[3] Gerhard Klimeck, Michael McLennan, Sean P Brophy, George B Adams III, and Mark S Lundstrom. nanohub. org: Advancing education and research in nanotechnology. *Computing in Science & Engineering*, 10(5):17–23, 2008.

[4] Suresh Marru, Lahiru Gunathilake, Chathura Herath, Patanachai Tangchaisin, Marlon Pierce, Chris Mattmann, Raminder Singh, Thilina Gunarathne, Eran Chinthaka, Ross Gardler, et al. Apache airavata: a framework for distributed applications and computational workflows. In *Proceedings of the 2011 ACM workshop on Gateway computing environments*, pages 21–28. ACM, 2011.

[5] Michael McLennan and Rick Kennell. Hubzero: a platform for dissemination and collaboration in computational science and engineering. *Computing in Science & Engineering*, 12(2):48–53, 2010.

[6] Dan Stanzione. The iplant collaborative: Cyberinfrastructure to feed the world. *Computer*, 44(11):44–52, 2011.

Table 2: Percentage of all researchers or educators who indicate that web-based applications providing access to specialized resources are either "somewhat" or "very" important to their work (n = 3958)

Specialized Resources	Importance
Data collections	63%
Data analysis tools, including visualization and mining	60%
Computational tools	60%
Tools for rapidly publishing and/or finding articles and data specific to my domain	58%
Educational tools	55%
Platforms for fostering group or community collaboration	53%
Simplified interfaces that eliminate the need to learn coding	51%
Citizen science and other public engagement resources	38%
Workflows that automate or capture tasks or processes	34%
Scientific instruments, such as telescopes, microscopes, or sensors	32%