Science gateways today and tomorrow:
positive perspectives of nearly 5,000 members of the research community

Katherine A. Lawrence
School of Information, University of Michigan, Ann Arbor, MI, USA

Michael Zentner
Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN, USA

Nancy Wilkins-Diehr
San Diego Supercomputer Center, University of California, San Diego, CA, USA

Julie A. Wernert
Research Technologies, Indiana University, Bloomington, IN, USA

Marlon Pierce
Research Technologies, Indiana University, Bloomington, IN, USA

Suresh Marru
Research Technologies, Indiana University, Bloomington, IN, USA

Scott Michael
Research Technologies, Indiana University, Bloomington, IN, USA

This is a pre-publication version of the following article which has been published in final form at wileyonlinelibrary.com:


1 Correspondence to: Katherine Lawrence, School of Information, University of Michigan, 4322 North Quad, 105 S. State St., Ann Arbor, MI 48109-1285, USA; kathla@umich.edu
Abstract

Science gateways are digital interfaces to advanced technologies that support science/engineering research/education. Frequently implemented as web and mobile applications, they provide access to community resources such as software, data, collaboration tools, instrumentation, and high-performance computing. We anticipate opportunities for growth within a fragmented community. Through a large-scale survey, we measured the extent and characteristics of the gateway community (reliance on gateways, nature of existing resources) to understand useful services and support for builders and users. We administered an online survey to nearly 29,000 principal investigators, senior administrators, and people with gateway affiliations. Nearly 5,000 respondents represented diverse expertise and geography. The majority of researchers/educators indicated that specialized online resources were important to their work. They choose technologies by asking colleagues and looking for documentation, demonstrated reliability, and technical support; adaptability via customizing or open-source standards was another priority. Research groups commonly provide their own resources, but public/academic institutions and commercial services also provide substantial offerings. Application creators and administrators welcome external services providing guidance such as technology selection, sustainability planning, evaluation, and specialized expertise (e.g., quality assurance, design). Technologies are diverse, so flexibility and ongoing community input are essential, as is offering specific, easy-to-access training, community support, and professional development.

Key words

Science/engineering gateways, portals, Web interfaces, software development, cyberinfrastructure, survey

Grant sponsor

This material is based upon work supported by the National Science Foundation under Grant Numbers 1216668, 1216723, 1216730, 1216733, 1216815, and 1216870.
1. INTRODUCTION

Over the last two years, science gateways have passed important milestones. Science gateways are defined broadly as digital interfaces to advanced technologies that support science and engineering research and education. These gateways, frequently implemented as web and mobile applications, provide access to community resources such as software, data, collaboration tools, instrumentation, and high-performance computing. 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 National Science Foundation (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 [2] have thousands of regular users. Also, significant effort has been placed into frameworks and infrastructure to support gateway development at scale, including HUBzero [3], iPlant [4], and Apache Airavata/SciGaP [5] in the US and WS-PGrade/SciBUS [6] 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. A small focus-group study previously identified a definite need and desire among the science-gateway building community to have greater access to the knowledge and experience gathered by fellow community members and the specialized services that support and interact with them [7, 8]. To investigate opportunities for growth—both of gateways and of a more integrated development community—and to measure, for the first time, the full extent and characteristics of this community, we have undertaken a very large-scale 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. Additionally, in order to establish the anticipated demand for support services for those who build gateways, we wanted to identify how much researchers and educators rely on gateways for their work. To do this, we engaged end users of gateways, builders of gateways and similar specialized resources, and administrative-level decision-makers who provision technology and human resources in support of research and education. This subset of stakeholders was chosen to represent potential “customers” of a center of gateway expertise. Within this sample, there were likely individuals with other stakeholder interests as funders or infrastructure providers, but these were incidental to the primary purposes of our study. To our knowledge, this is the largest such survey to explore the opportunities of and needs for science gateways today and tomorrow.

2. SURVEY DESIGN

This survey was conducted as part of an NSF-funded conceptualization grant to determine community requirements for a “Scientific Software Innovation Institute” to support science gateways. The grant funded a team from six US-based universities who participated in various capacities in the development, administration, and analysis of survey results. (For more information about our partnership, visit www.sciencegateways.org.) 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 that 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 application creator. Because a gateway can be a large effort with many stakeholder interests, it was important to get the perspective of each key stakeholder type.

2.1. Questionnaire

We asked survey participants about the importance of Web- or mobile-based interfaces in their fields and also about their participation in building such interfaces. At the start of the survey, we defined “science gateways” as we have at the beginning of this paper. After two “screening” questions to determine what role(s) (researcher and/or educator, administrator, application creator) was relevant for each respondent, we directed them to one or more of three sections. Questions posed to administrators and application creators about services they might use were all framed within the context of a “Science Gateways Institute.”

Administrators received a short selection of questions as to whether they expected that their institution would make use of planning services, building and maintenance consulting/contractor services, or expertise for accessing specialized resources.

Application creators were asked what types of resources for research and/or education they had helped to create. To identify these individuals, we asked them if they had participated in the creation of applications (e.g., software, websites, portals, science gateways, mobile interfaces, etc.) that support research and education in any of a variety of roles (e.g., principal investigator [PI], software developer, advisory board member, etc.). For such projects (defined as the resources these application creators had helped to create), we also asked what types of people worked on their projects, and what types they would have liked to add to their team. They were also asked if they could anticipate needing help in any of the planning, building/maintenance, or resource access areas about which we asked the administrators, and they were asked to prioritize the services/support types that were most critical. They were asked a series of questions about training and keeping up to date with Web-based technologies. A subset of developers with more technical expertise were screened and selected to answer some questions about their anticipated use of hardware and software resources such as APIs, Web-hosting and computing, and repositories as well as what technologies they have used to build their own Web-based resources.

We asked researchers/educators questions about what Web-based applications that provide access to specialized resources were important to their work, who provides or develops such resources, and whether they use mobile devices to access the resources. They were also asked how they learn about new technologies, discoveries, and innovations in their fields and what factors influence adoption of those technologies.

At the end, all participants were asked questions about a few topics. Those who are responsible for hiring or supervising people with expertise in technology development or maintenance were asked about the importance of several common concerns. All were asked what were the most important decision-making factors (besides cost) for adopting a new technology. Demographic questions included their primary area(s) of expertise, type of institution with which they were affiliated, and whether that institution served a significant number of minority students. All participants were also asked if they would like further information or contact from our project group, and we provided a section for open-ended comments.
2.2. Population sample

The survey sample was collected from three primary sources: National Science Foundation–funded principal investigators (PIs) (25,975 or 90% of sample), senior administrative members of EDUCAUSE and CASC (1,553 and 108 individuals respectively, for a total of 1,661 or 6% of sample),\(^2\) and individuals who have previously expressed interest in gateway initiatives (1,076 or 4% of sample). Because we wanted to gauge demand for gateways and gateway-development services, we did not screen survey recipients from NSF, EDUCAUSE, or CASC for prior interest in or use of gateways.

The NSF PIs were limited to those who had received funding from any directorate within the last 24 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. (Duplicates, foreign email addresses and anyone with an nsf.gov email address were removed.)

To reach administrators, we purchased from EDUCAUSE a filtered sample of their nearly 50,000 members. We targeted those identified in the role of CIO (Chief Information Officer), CTO (Chief Technology Officer), VP (Vice President or Vice Provost), and high-level director by filtering the list to include individual members (as opposed to organizational members and/or non-member affiliates) designated as their organization’s primary representative. The list was further filtered to include only representatives from US institutions. The roles represented included primarily CIOs, IT directors/assistant directors, and CTOs, but also included lesser numbers of members of general executive management (e.g., president, chancellor, provost, chief academic/business officer, vice president/chancellor student affairs/admissions, etc.), university chief librarians, academic deans and faculty members, campus functional leaders (e.g., registrar, bursar, general counsel, controller, etc.), and various IT managers. CASC has 79 member institutions, each with one or two representatives. We included all of these representatives in our sample.

The individuals with prior interest in gateways included participants from previous focus groups and workshops plus volunteers on our Website. This sample also included people who have attended or participated in conferences or email lists targeted at those who build or support science gateways, including the Gateway Computing Environments workshops, XSEDE Gateways, XSEDE Campus Champions, and HUBbub (the user conference for HUBzero).

A small number of recipients were from outside the US. The vast majority of survey recipients and respondents were based in the US due to the nature of NSF funding requirements. Our team would like to direct future efforts toward incorporating input from those funded by other federal agencies and from international communities.

We expected that some people would be sampled from multiple sources given that members of the academic-oriented population often wear many hats. After assembling this large sample, duplicates were identified and removed. There were 436 total duplicates (some records were repeated 3 times). In many cases, the duplicate was someone on a mailing list (e.g., one of our “prior interest” sources), so the record with less information (typically only an email address)

\(^2\) EDUCAUSE (www.educause.edu) and CASC (www.casc.org) are both US-based nonprofits that promote the use of information/computing technology in higher education and scientific research. Their membership includes high-level administrators concerned with the implementation and use of cyberinfrastructure at institutions for research and education, so these were our sources for a population of administrators.
was removed. If the duplicates were entirely identical, the person cleaning the records randomly selected which record remained.

The final total sample size was 28,712, and our 4,957 participants represent a response rate of approximately 17%. The response rate from NSF-funded PIs was 18% (4,533 of 25,975 sampled), from EDUCAUSE and CASC was 10% (163 of 1,661 sampled), and from our “prior interest” group was 24% (261 of 1,076 sampled). This exceeded our 10% target rate of response. As we discuss later in the paper (Section 3.2.1), 99% (3,952 of 4,004 question respondents) of the researcher/educators indicated that at least one type of gateway-type resource was important to their work, so nearly all respondents in this category are users of gateways and therefore possibly biased in favor of their use. While many people opt out of surveys because they are busy and not sufficiently motivated by the request, the high response rate of 17% is a promising indicator of interest. It is possible that our data may have been biased toward those who anticipate some indirect benefits to their work by responding to the survey, as evidenced in the higher response rate among those from the “prior interest” group and the lower response rate from the administrative sample, but this represents a small percentage of the overall responses.

2.3. Implementation

The implementation process was reviewed by an Institutional Review Board to ensure that the survey was conducted ethically and that respondent data were protected. Responses were kept confidential. 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. The complete survey text, invitation email messages, and instructions are posted at http://hdl.handle.net/2027.42/110982.

2.4. Analysis

We received 4,957 responses, offering a diverse set of perspectives, disciplines, and work roles. Because many people selected more than one role, these respondents would have received multiple subsets of questions. Based on the initial two filtering questions, which determined work role (e.g., faculty, research scientist, application creator, administrator, and post-doctoral fellow) and participation in the creation of applications that support research and education, we directed the respondents to one or more of 3 sections: Administrator (n=447, 9% of total respondents), Application Creator (n=2,819, 57% of total respondents), or Researcher/Educator (n=4,538, 92% of total respondents). All respondents received a set of questions at the end, which included demographic-type questions.

Of the larger set of respondents, some went no further than the first page, which we eliminated as “implicit refusals.” Other respondents answered a portion of the questions and then “dropped off,” perhaps deciding that they had devoted enough time to the survey. Some respondents selectively answered questions throughout the survey, and the questions that they skipped were treated as “not answered.” In each analysis, we report the number of respondents who qualified for a given question as well as the number of respondents who actually answered the question, eliminating “dropped off” and “not answered” as appropriate.
Percentages and other statistics are reported as a percentage of the actual number of answers. In some questions asking respondents to rate a number of individual items (e.g., what specialized resources are important to them), some indicated their responses to a subset of the items. For these questions, the number of respondents reflects the total number of people who answered at least one item in that question, treating the items they skipped as an indication of disinterest. We believe that this allows us to compare items more fairly by using the same “n.” This has led us to update some specific figures and analyses from this data that was previously presented at a workshop.

3. RESULTS AND DISCUSSION

3.1. Respondent demographics

The high degree of community interest has been reflected by the number and variety of responses to our survey. We received 4,957 responses, 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 physical and mathematical sciences (30%, n=1,486) and life sciences (22%, n=1,102). Computer and information sciences (16%, n=794), engineering (16%, n=775), and environmental sciences (14%, n=682) were also well represented.

Because this study was funded primarily to identify the needs of NSF-funded researchers, our responses were largely from US institutions. Responses covered the US, as shown in Figure 2. Yellow dots indicate a single response, green indicates 2–10 responses, blue 10–50, and red more than 50 responses. Future work remains to clean up a small number of institutions that were not automatically located by Google and to combine institutions when slightly different names were used. This happened infrequently because most institutions were gathered from the NSF awards database where names are standardized. We feel this map is largely representative of survey responses.

Respondents were comprised primarily of faculty and research scientists, but also included members of higher-education leadership, graduate students, and technology developers. Respondents were asked to identify themselves by role. The selections offered were administrator, faculty, research scientist, technology developer, or post-doctoral fellow. Our 4,957 respondents selected 5,848 roles for mean of 1.2 roles per respondent. Respondents received survey questions tailored to their roles.

The largest category of respondents identified themselves as faculty members (n=4,093, 83% of 4,957), followed by research scientists (n=798, 16%), administrators (n=447, 9%) and technology developers (n=253, 5%). For those identifying as technology developers, a portion

---

3 Earlier analysis conducted for a workshop presentation mistakenly aggregated all responses within a larger disciplinary category. For example, an individual indicating mechanical and electrical engineering was counted twice rather than once in the category of engineering. Recalculation has reduced and slightly rearranged proportions but has not substantially altered the overall composition of the respondent pool.

4 The question option for self-identifying as an administrator parenthetically clarified “in the office of a CIO, CTO, IT director/service provider, or other high-level director.” While our sampling intended to reach administrators through EDUCAUSE and CASC, our respondents who self-identified as administrators came from all sources: 145 from EDUCAUSE and CASC (32% of the administrator group), 245 from NSF (55%) and 57 from the “prior interest” group (13%).
also identified themselves as research scientists (n=136, 54% of 253), faculty (n=116, 46%), or administrators (n=57, 23%). The smallest role groups represented in our sample were post-doctoral fellows (n=100, 2% of 4,957) and “other” (n=157, 3%). Despite a smaller number identifying as developers by role, some 57% of respondents (n=2,819 out of 4,957) across all roles report having participated in some capacity in the creation of desktop, mobile, or Web applications. An additional 9% (422 of 4,957) who have never participated in the creation of applications indicate that they hope to do so in the future.

We asked all survey participants to describe the type of institution with which they were affiliated (allowing them to check all that applied). Because the majority of our survey sample was composed of NSF-funded researchers, most respondents (3,829, or 77% of 4,957 total respondents) came from academic institutions. Three-hundred twenty-two respondents (6% of 4,957) came from non-profit organizations, 112 (2%) from corporate entities, and 105 (2%) from government organizations and contributed to diverse institutional viewpoints, albeit a small portion of our sample. About 10% (427 of 4508 respondents to the question) identified their academic institutions as Minority-Serving Institutions, while a nearly equal portion (373 respondents or 8%) did not know. Minority-Serving Institutions (MSIs) are designated as such or enroll more than half their students from minority groups that may include American Indian, Alaska Native, Black (not of Hispanic origin), or Hispanic.

3.2. Using Gateways

3.2.1 How do scientists use gateways?

Science gateways and other applications for research and education offer a wide variety of capabilities (Figure 3). Those respondents who have participated in the creation of some type of application that supports research and education reported creating a wide variety of tools and services. The most common were education tools (18%, n=1,396 of 7,805 identified application types), computational tools (16%, n=1,272), data analysis tools (including those for visualization and data mining; 16%, n=1,239), and data collections (15%, n=1,215).

The survey focused not just on those involved in developing Web applications, but also attempted to assess the importance of applications providing access to specialized resources. We asked those who identified as researchers and/or educators, “From your perspective as a researcher and/or educator, how important to your work are Web-based applications that provide access to the following specialized resources?” (Table 1). For accessing most types, at least 60% of respondents indicated that Web-based applications were “somewhat” or “very” important (4,004, or 88% of 4,538 researcher/educator respondents answered this question). As many as 75% (n=2,995 of 4,004) indicated that data collections were important to their research/education work, and data analysis tools (n=2,886) and computational tools (n=2,874) were additional top categories of important resources (72% each). Tools for rapidly publishing and/or finding domain-specific articles and data (n=2,773, 69%) and educational tools (n=2,663, 67%) were also important to many. Only 52 respondents (1% of 4,004 respondents) answered that every resource in the list was “very unimportant” or “somewhat unimportant” to their work.

The respondents were also asked to indicate their primary areas of expertise from among 59 fields organized in eight higher-level groups. These fields and groups were distilled by looking at the fields distinguished by NSF directorates, the “Fields of Study” list historically used in surveys of academic communities, and other departmental classifications. Types of resource categories were analyzed with respect to each of these expertise areas. Within each of the eight
high-level expertise groups, the proportion of respondents who indicated the resource category was “somewhat” or “very” important was calculated for each contained expertise area. The average and standard deviation of these proportions were calculated within each high-level expertise group (see Figure 4). Across disciplines, the resource category of Data Collections uniformly ranks high. The discipline category of Arts & Humanities had the highest mean level of importance in 5 out of 10 of the resource categories (Citizen Science, Collaboration Platforms, Data Collections, Rapid Publishing, and Workflows). However, for these resource categories, variability across contained expertise areas was generally higher for Arts & Humanities than for other high-level expertise groups, perhaps indicating less homogeneity of the disciplines contained within that group than those contained in other groups. Similarly, the expertise group of Professional Disciplines also exhibited a high degree of variability, possibly for similar reasons.

To determine what resource categories were in greater demand within specific disciplines, we calculated the percentage of respondents in each expertise area who considered a given resource category to be either somewhat or very important. We then calculated the overall average of these percentages for each resource category, as well as the standard deviation of the percentages. Very few resource/expertise combinations fit a conventional “outlier” definition of three standard deviations from the mean. However, at two standard deviations, we see some combinations that stand out (Table 1).

For those researchers and educators who use the resources described in this survey, 47% either use mobile apps to access those resources, or would like to use them (n=1,809 of 3,852). There are no specific expertise areas that said they “do not use mobile apps but would like to” that stand out above others.

3.2.2. How do people learn about and choose gateways?

Because building community is critical to a science gateway’s success, we wanted to learn how domain researchers/educators learn about new technologies. Respondents could make multiple selections. Of the 3,913 (out of 4,538) researcher/educator respondents to this question, 78% (n=3,042 of 3,913) indicated that they learned about technologies from colleagues, while 61% (n=2,378) learned at conferences and meetings, and 51% (n=1,993) learned through research publications. Fewer respondents learned from Web searches/specialty sites (38%, n=1,480) or from students (33%, n=1,273). Fewer than 10% learned from mailing lists (9%, n=371), advertisements/trade magazines (4%, n=160), or other methods (2%, n=63). It is interesting to note that the top three means of learning about new technologies surpassed Web searches, perhaps indicating that Google is not sufficient for locating specialized scientific interfaces that researchers trust.

After potential “customers” learn about science gateways or other new technologies, there are additional factors that determine whether they will adopt them for their work. Respondents were asked to select the three most important factors when considering the adoption of a new technology. The 4,180 respondents (out of 4,957 total respondents) selected 12,283 factors, producing a mean of 2.5 factors selected per respondent. The most important factor, identified by 49% (n=2,053 of 4,180) of respondents, was the availability of documentation (Figure 5). After that, the next several factors received a similar proportion of the votes. The ability to adapt or customize a tool (35%, n=1,461), its demonstrated reliability (31%, n=1,278), the availability of technical support (30%, n=1,242), and whether or not a package was open source (27%, n=1,146) were all ranked similarly. Some of this reflects the modern approach to working
independently: “Give me good documentation, let me make my own code modifications, and I’ll be happy.” Interestingly, the reputation of the technology creators (11%, n=453) received the lowest ranking, indicating that if a tool works and is well-supported, then who has created it is less of a factor in its success.

3.2.3. Where are gateways found?

Participants who had indicated resource categories were somewhat or very important to their work were asked how those Web resources they use were created or accessed (e.g., commercial service, their own research group, a public or academic organization, or they do not know). They were allowed to choose more than one option, given that multiple parties may participate in creating the resources. Figure 6 shows that for every category, “my own research group” dominates except for scientific instruments. Scientific instruments are dominated by commercial services, which might suggest that instrument vendors play a key role. When looking only at externally provided resources (i.e., not created by “my own research group”), commercial services are more popular for some and public/academic institutions are more popular for others. Collaboration, rapid publishing, and scientific instruments were more noticeably supplied by commercial services whereas computational tools and data collections are dominated by public/academic institutions as the external resource. Although commercial services are used significantly, they are most often used in conjunction with other categories.

Citizen science and collaboration platforms offer surprising results. The large number of “do not know” responses for citizen science may indicate that large groups of users of citizen science sites may be unaware of how they were created. Citizen science had the lowest degree of commercial involvement and was strongly dominated by the researchers’ own research groups. It is also somewhat surprising that the researchers’ own research groups strongly dominate collaboration platforms when so many publicly and commercially available collaboration infrastructures already exist. This may be due to the interpretation of exactly what constitutes a collaboration platform (e.g., a shared network drive, Dropbox, Google Drive, HUBzero, etc.).

3.3 Building Gateways

3.3.1. Who builds gateways?

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 application creator group selected 5,366 roles (across 2,819 self-identified application creators), producing a mean of 1.9 roles per respondent (Figure 7). The aggregate responses indicate a breadth of experience, but skew heavily toward the perspective of principal investigator (78%, n=2,199 of 2,819), with the next most common roles being domain-based experts (or content specialist, 21%, n=603) and advisory board or steering committee members (20%, n=577).

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 8). Student or post-doc programmers were by far the most prevalent (70%, n=1,932 of the 2,756 respondents [which was 98% of 2,819 application creators]), followed by project managers (45%, n=1,250) and professional software
developers (44%, n=1209). Least available but most desired were quality assurance/testing experts (42%, n=1,156) and graphic designers (36%, n=979). Usability consultants were also desired (34%, n=932), but many also indicated that they were not needed (41%, n=1,134). 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.

Within our sample, the respondents who had participated in creating Web- or mobile-based applications were asked specific questions. 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 building a gateway. (Figure 9 shows their responses in blue.) The top area of interest is evaluation, impact analysis, and Web analytics (72%, n=1,541 of 2,153). Other important interest areas include planning how to adapt technologies (67% or 1,439 of 2,153), Web/visual/graphic design (67%, n=1,434), choosing technologies (66%, n=1,420), and usability services (66%, n=1,419), echoing some of the staff member roles that respondents wished they had. In all but one of the more than 20 areas of potential support, at least 40% indicated that some help might be needed.

We had 447 respondents who indicated that they had administrator roles, and we used their responses to determine which types of services were important to decision makers. Like the application creators, they were specifically asked if their institution would make use of a variety of planning services, building and consulting services, and services providing expertise in accessing resources.

We considered projected demand for a service as the number of respondents indicating one or more of “Would Definitely Use,” “Would Use at Minimal Cost,” and “Would Use at No Cost,” divided by the total number of responses. In decreasing order of demand, these services are shown in green in Figure 9. Overall, planning services populate the top in service types, with sustainability guidance and choosing or adapting to technologies topping the list. While there is non-trivial demand for all proposed services, there is clearly a ranking defined by the administrators more toward the strategic planning activities (indicated with lighter shades of color) and less on the execution (medium shades) or the resources (darkest shades).

3.3.2. How do people build gateways?

Those involved in creating applications (as defined in Section 2.1) were asked about their participation in developing Web applications vs. mobile apps. More than 78% (n=2,175 of 2,760 respondents) participated in the creation of Web applications. In contrast, only 17% (n=419 of 2,538) of them participated in creating mobile apps. Of those who helped create mobile applications, 93% (n=388 of 419) also helped create Web applications. At present, mobile application creation is limited, but 38% (n=952 of 2,538) of respondents are considering them for the future.

To understand the common technologies that are used or needed for building Web- or mobile-based applications, we filtered participants with this question: “In your projects, do you make decisions about and/or implement the Web-based technologies you use?” The 1,436 application creators who choose or implement their technologies most commonly cited WordPress (36%, n=522) and Drupal (29%, n=423), followed by Ruby on Rails (14%, n=198), Django (13%, n=191), and Joomla (10%, n=148). However, 36% (518 of 1,436 respondents)
selected “Other,” reflecting nearly 200 other development platforms, frameworks, and applications that were cited at least once (by the 364 who entered further details about “other” technologies), including DreamWeaver, Java/JavaScript, Python, and php/MySQL, as well as “home-grown” codes. A graphical representation of the proportions of these technologies in use is shown in Figure 10. Given the wide range of available technologies, it is not surprising that “Other” was selected by a third of the respondents, and no single technology dominated. We do note that PHP underlies all the listed choices except Ruby on Rails, Django, and Liferay; 81% (n=1,170) of the respondents are using some type of PHP-based technology, although this could have been biased by our choice of technologies in the question.

For these application creators who indicated that they make decisions about and/or implement the Web-based technologies they use (1,554 or 55% of the 2,819 application creators), we asked several more technically oriented questions. First, we asked how likely would they be to use various API resources relevant to building science gateways. APIs can be used to access remote services provided by other groups. Responders were asked their likelihood of using six types of API service categories. Overall, we found data services to be the most likely to be used, with 46% of all respondents (n=706 of 1,530) selecting “somewhat likely” or “very likely” to use. Security and information services were also more likely to be used; Table 2 shows the results for all items.

We noted these responses were weighted by the judgments of those identifying themselves as faculty and administrators. When considering respondents in roles of research scientist, postdoc, and technology developer, the percentages indicating they were somewhat or very likely to use services uniformly increased by 3%–6%, with the exception of data services. The order of services did not change for this smaller group. In particular, security services increased to 47% (n=174 of 372 respondents in these roles), nearly as high as data services. In additional analysis, we also noted a statistically significant spread in the likelihood of different fields of expertise to want different APIs. This needs further and more directed investigation for gateways that want to target specific areas, such as the Arts and Humanities.

Clearly there is more need to explore the technologies in use by gateway creators, given the large number of respondents choosing “Other.” While our questions mostly focused on front-end technologies, we recognize the importance of middleware and other back-end technologies. Investigating that plethora of technology options was beyond the scope of this survey, so we directed our questions toward the front end to better understand how back-end technologies might have to integrate. For gateways wishing to use APIs and associated services, it is necessary to provide language- and framework-independent means of accessing these services. The value in providing “turnkey” solutions based on existing frameworks needs closer examination. The importance of PHP as a programming language for the Web is also clear, and developers of gateway tools should take this into account. 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.

3.3.3. What are workforce development concerns?

Survey respondents whose work involves hiring or supervising people with expertise in technology development (n=1,538 or 31% of all 4,957 respondents) were asked about their employment concerns. Figure 11 shows that by a strong margin, the availability of people with cross-disciplinary abilities was the most important. Conversely, outsourcing opportunities ranked the least important.
When examined by role, administrators unilaterally responded with the highest degree of concern across items when all roles were compared proportionally. The greatest misalignments between administrators and other stakeholders are in two specific concerns. The administrators (and those who identified their work role as technology developers) were substantially more concerned about competing with industry for top performers than were faculty, research scientists, or post-docs. This may be because the pool of workers in academic departments is more specialized than those who might work in IT departments or developing technologies for multiple constituents. Identifying a pipeline of future staff showed a similar pattern of concern, likely for similar reasons.

We asked Web- and mobile-application creators what mechanisms they prefer for training their Web-based technology development staff (n=2,076 or 74% of 2,819). They were allowed to indicate up to three preferences. Self-paced, online learning was by far the most popular (50%, n=1,048 of 2,076), followed by workshops or short courses (41%, n=856). Webinars (33%, n=688) and on-site custom training (29%, n=600) are also reasonably popular options. It is possible that less popular options such as MOOCs (Massive Open Online Courses; 16%, n=334) 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.

Respondents were asked to select their top two choices from a list of information sources available or potentially available for keeping up-to-date on relevant Web-based technologies (2,077 or 74% of 2,819 application creators answered). We were not surprised to learn that online communities via email or Web-based forum (47%, n=966 of 2,077) or one’s own development teams (43%, n=891) or both (19%, n=392) were by far the most common choices. This pattern held across roles and areas of expertise. The third most popular response was a focused workshop or track as part of an existing conference (18%, n=371). Newsletters and independent, niche conferences were not as popular. These responses reinforce the need for gateway support efforts to reach out through existing or convenient venues. As follow up work, we hope to examine the specific types of online communities that are available, their popularity, and their perceived effectiveness.

4. CONCLUSION & FUTURE WORK

This extensive survey indicates that gateways are an active part of the science and engineering research and education landscape. Many scientists and educators depend on Web-based applications to access specialized resources, particularly for computational tools and data access and analysis. Educational tools and connecting with or collaborating with others in one’s own domain are also important functions. These resources are provided in many different ways, but people most often turn to their own research group to develop Web-based applications for accessing them. Logically, some resources, such as computational tools and data collections, are provided by public or academic institutions, whereas others, such as collaboration tools, scientific instruments, and rapid publishing mechanisms, are supplied commercially. The source of citizen science services are variable and not entirely transparent. Consumers of specialized resources are careful about how they adopt new technologies. Most learn about new technologies from colleagues as well as professional meetings and publications. While there is no dominant characteristic prompting them to adopt, high priorities are documentation, adaptability, reliability, and technical support.
The creators 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, possibly because they are not common to academic institutions. Enabling these creators to learn from experts and access specialists is perceived as a valued and needed service. Most significantly, application creators and administrators want guidance in selecting or adapting technologies and help with evaluation such as impact analysis or Web analytics. Some would also engage contractors to fill the project-team roles that are scarce at their institution.

Currently, to build and deliver their applications, developers are using the Web more than mobile apps, but more than one-third anticipate mobile apps in their future work. Technologies are quite varied, but they are dominated by technologies that are PHP-based. Data services (as APIs) are uniformly desirable by application creators across fields of expertise and roles. When excluding faculty and administrators from the pool of responses, the value of security services was also notable.

In order to provide the best support and supply the necessary training for a community of gateway developers, we wanted to know workforce development concerns, preferred training mechanisms, and how people keep up-to-date with relevant technologies. The greatest concerns are developing staff with cross-disciplinary abilities and giving technology staff a viable career path. Self-paced, online learning (including Webinars) as well as face-to-face short workshop opportunities are the most preferred methods for receiving training, while online communities, development-team colleagues, and existing conferences are the best ways to provide updates about Web-based technologies.

From this survey and analysis, what can we conclude about the science gateways of tomorrow? Detailed requirements analysis, technical considerations, software and service architectures, and human-computer interaction research are usually used to answer these questions, but exploring at this level of detail was not the purpose of our survey. Furthermore, disruptive technologies (such as the swift adoption of mobile and cloud platforms by academic researchers in recent years) make forecasts of technology trends difficult.

Nonetheless, we believe that this survey has been able to identify important trends that are independent of technical considerations. The size and breadth of the response indicates that gateways are important to a considerable number of researchers. We believe the importance of gateways to research and education will continue to grow. It is thus important to bring gateway developers and users into a coherent community to overcome problems of sustainability, resource limitations, isolation, and scaling that are currently unsolved. From the interest indicated by respondents, we see the need for an open technical forum for gateway tool providers and developers to interact, exchange information, and share experiences.

Gateways must scale to all sizes of communities, which has implications for technology, staffing, and development methods. As an example, the number of students served by an online, course-integrated gateway to support data analysis and computational experiments in a single professor’s class may range from fewer than ten students (for highly specialized subjects) to thousands of students in MOOCs. Gateways in these instances must be easy to create and maintain for the length of a course (several months to a year). At the other end of the spectrum, discipline-specific gateways may have a goal from the beginning to serve entire science communities. These gateways need reliable infrastructure and also a workforce pipeline of developers, operators, and support personnel for the long term. The general popularity of data-centric APIs and services among survey respondents also indicates the need for an important shift in thinking for many gateways, which have focused on computational tools.
We conclude that gateways of the future will benefit significantly by participating in a more coherent community. Gateway providers need a reliable source for finding and contributing to high-quality science gateway software that fits their needs. They also need support for making their gateways successful. This includes technical support and also entrepreneurial support; gateways seeking to serve large user bases may need to think of themselves as startup companies and learn how to interact with their “customers” while keeping in mind how they will finance their work over time. Gateway providers also will need to foster a pool of talented developers who see satisfactory career paths working on science gateway technologies and supporting scientific research. The challenge that we see in the future is designing and implementing governance for a science gateway community that will be clearly led but open and flexible.

As we reviewed our results, we noticed some additional opportunities for future investigation. For example, our sample was largely PIs funded by the National Science Foundation, so of course our responses were biased toward the science and engineering communities in the US. Clearly there is a diverse and interested community in the arts and humanities as well as in the life and medical sciences, and we hope to identify how we might support gateway development in those communities and reach out to our international colleagues. We also have the opportunity to reach out to a more targeted audience of technology specialists to further investigate the requirements for complex back-end systems. Additionally, our survey attracted a large number of thoughtful responses to our open-ended comment area at the end of our questionnaire, and we need to follow up with these ideas and identify additional questions or topics to pursue. Finally, we would like to open this dataset to a broader, interested audience who may find additional insights and opportunities for building and sustaining the diverse and growing science gateway community.

ACKNOWLEDGMENTS

The entire team of the Science Gateways 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. Additional help was provided by Maytal Dahan, Stephen Mock, and Yan Zhou.
REFERENCES


# TABLES

Table 1. 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 (for first two columns, n=4,004, or 88% of 4,538). The third and fourth column is derived from the average and standard deviation of percentages for each expertise area by resource category (n=3,793). The areas that are notable at 2 standard deviations are listed below.

<table>
<thead>
<tr>
<th>Specialized Resources</th>
<th>Percent</th>
<th>Expertise Areas Disproportionately Interested</th>
<th>Expertise Areas Disproportionately Disinterested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collections</td>
<td>75%</td>
<td></td>
<td>Math &amp; Statistical Sciences, Physics</td>
</tr>
<tr>
<td>Data analysis tools, including visualization and mining</td>
<td>72%</td>
<td>Law &amp; Legal</td>
<td>Manufacturing, Math &amp; Statistical Science, Public Administration</td>
</tr>
<tr>
<td>Computational tools</td>
<td>72%</td>
<td></td>
<td>Comp &amp; Info Sci (Other), Public Administration, Anthropology</td>
</tr>
<tr>
<td>Tools for rapidly publishing and/or finding articles and data specific to my domain</td>
<td>69%</td>
<td>Languages &amp; Literature, Arts &amp; Humanities (Other)</td>
<td>Law &amp; Legal</td>
</tr>
<tr>
<td>Educational tools</td>
<td>67%</td>
<td>Languages &amp; Literature, Social Service</td>
<td>Economic or Political Science</td>
</tr>
<tr>
<td>Platforms for fostering group or community collaboration</td>
<td>63%</td>
<td>Languages &amp; Literature, Arts &amp; Humanities (Other), Social Service</td>
<td></td>
</tr>
<tr>
<td>Simplified interfaces that eliminate the need to learn coding</td>
<td>62%</td>
<td>Social Service</td>
<td></td>
</tr>
<tr>
<td>Citizen science and other public engagement resources</td>
<td>47%</td>
<td>Languages &amp; Literature, Architecture, Natural Resources, Social Service</td>
<td></td>
</tr>
<tr>
<td>Workflows that automate or capture tasks or processes</td>
<td>42%</td>
<td>Visual/Performing Arts, Social Service</td>
<td>Public Administration</td>
</tr>
<tr>
<td>Scientific instruments, such as telescopes, microscopes, or sensors</td>
<td>39%</td>
<td></td>
<td>Public Administration, Economic or Political Science</td>
</tr>
</tbody>
</table>
Table 2. Percentage of respondents who indicated that they were somewhat likely or very likely to use a given service (n=1,530).

<table>
<thead>
<tr>
<th>Services</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data services (e.g., metadata and collection management, access, movement, analytics)</td>
<td>46%</td>
</tr>
<tr>
<td>Security services (e.g., identity, authorization, profile)</td>
<td>39%</td>
</tr>
<tr>
<td>Information services for remote resources (e.g., monitoring, cataloging)</td>
<td>35%</td>
</tr>
<tr>
<td>Execution services (e.g., compute job, workflow)</td>
<td>29%</td>
</tr>
<tr>
<td>Event services (e.g., messaging, events)</td>
<td>28%</td>
</tr>
<tr>
<td>Accounting services (e.g., CPU and/or software usage, allocation)</td>
<td>24%</td>
</tr>
</tbody>
</table>
Figure 1. Primary areas of current domain expertise. Respondents could select all that apply; 84% (4,141 of 4,957 participants) responded, generating 6,689 total responses (mean=1.6 domains per respondent). “Professional Disciplines” included Architecture; Education; Information; Law; Legal Professions and Studies; Library Science; Management or Business; Natural Resources, Forestry, and Conservation; Public Administration; Social Service Professions; and "other."
Figure 2. Geographic location of the institutions of survey respondents (n=4,957). Yellow dots indicate a single response, green 2 to 10 responses, blue 10 to 50 responses, and red more than 50 responses. (Locations are determined by the institutions identified in the survey invitation list and filtered to include only respondents.)
Figure 3. Types of applications created by respondents (n of application types=7,805, by 2,756 creators (out of 2,819); mean=2.8 application types per application creator).
Figure 4. Percent of respondents, within 59 expertise fields organized in 8 higher-level groups, who indicated a resource category was “somewhat” or “very” important. The average and standard deviation of their discipline-based percentages were calculated within each high-level expertise group (n=3793, which includes the overlapping set of those who answered both questions. Because respondents had to answer both questions to be included in this analysis, the percentage of expressed interest is higher than other analyses which had more non-respondents).
Figure 5. Most important factors when adopting a new technology. Respondents could select up to 3 choices; 4,180 (84% of 4,957 participants) responded, generating 12,283 factors (mean=2.9 factors per respondent).
Figure 6. How Web-based applications for accessing specialized resources (identified as important to research/education work) are provided or developed. Respondents could select all that apply, and choices depended on what resources they had prioritized in a previous questions; between 32% and 62% (1,459 to 2,841 of 4,538 participants) responded.
Figure 7. Roles on application development projects. Respondents could select all that apply; 2,819 respondents in the application creator group selected 5,366 roles (mean=1.9 roles/respondent).
Figure 8. Types and desirability of staff members who have worked on projects (n=2,756 respondents or 98% of application creators).
Figure 9. Types of services relevant to administrators and application creators. Administrators (n=447, in green) were asked what types of planning services, building and maintenance consulting/contractor services, and expertise for resource access their institution would use. Application creators (n=2,819, in blue) were asked if they would seek help for a nearly identical array of services. The figure shows the percentage of administrators who would definitely use, use at minimal cost, and/or use at no cost such services, juxtaposed with the percentage of mobile- or web-based application creators who would seek at least some help from a service provider (n=2,153, or 76% of 2,819 creators). The shade of color indicates the phase or type of service as indicated in the key (e.g., the lightest colors are for planning services).

* For application creators, “Guidance with project sustainability” was labeled “Keeping your project running,” and “Choosing and Adapting to Technologies” were 2 separate items (rated 66% and 67% respectively).

** Administrators did not receive “Source code review” as an item.
Figure 10. This word cloud indicates the diverse assortment of technologies used by application creators to build web- or mobile-based applications. The most commonly mentioned technology by the 1,529 respondents was WordPress (34%), while an additional 34% indicated “Other.” All technologies mentioned at least 3 times in the “Other” category are included in this figure.
Figure 11. Top concerns for hiring and maintaining a highly qualified technology or science gateway development staff. Figure illustrates the percentage of “very” or “somewhat” important concerns indicated by all survey respondents whose work involves hiring or supervising people with expertise in technology development (n=1538).