

# Science Gateways and Data

## A Large-Scale Survey of Community Use and Needs



### Abstract

With the rise of science gateway use in recent years, we anticipate there are many opportunities for growth. We describe a large-scale survey that received nearly 5,000 responses. Our goal was to understand what type of support services might be provided to gateway communities. The importance of data services featured prominently among respondents working in research & education.



Read our full paper:  
<http://sciencegateways.org/surveyarticle/>

### Nearly 5,000 Responses Indicate Significant Interest

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. The total sample size was nearly 29,000, and the 4,957 participants in our online survey (by email invitation) represent a response rate of approximately 17%, exceeding our 10% target.

Respondents represent a broad range of disciplines (Figure 1). Respondents were primarily faculty and

research scientists, but also included members of higher education leadership, graduate students, and technology developers. Some 57% of respondents (n=2,819) report having participated in some capacity in the creation of desktop, mobile, or web applications.

The projects created by respondents employ many different types of people. The respondents who have participated in development projects have served in multiple roles, weighed heavily toward principal investigator (78%), with the next most common roles being domain-based experts (or content specialist, 21%) and advisory board or steering committee members (20%).

### Authors

**Katherine A. Lawrence**  
 University of Michigan School of Information

**Michael Zentner**  
 Purdue University

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

**Julie A. Wernert**  
 Research Technologies, Indiana University

**Marlon Pierce**  
 Research Technologies, Indiana University

**Suresh Marru**  
 Research Technologies, Indiana University

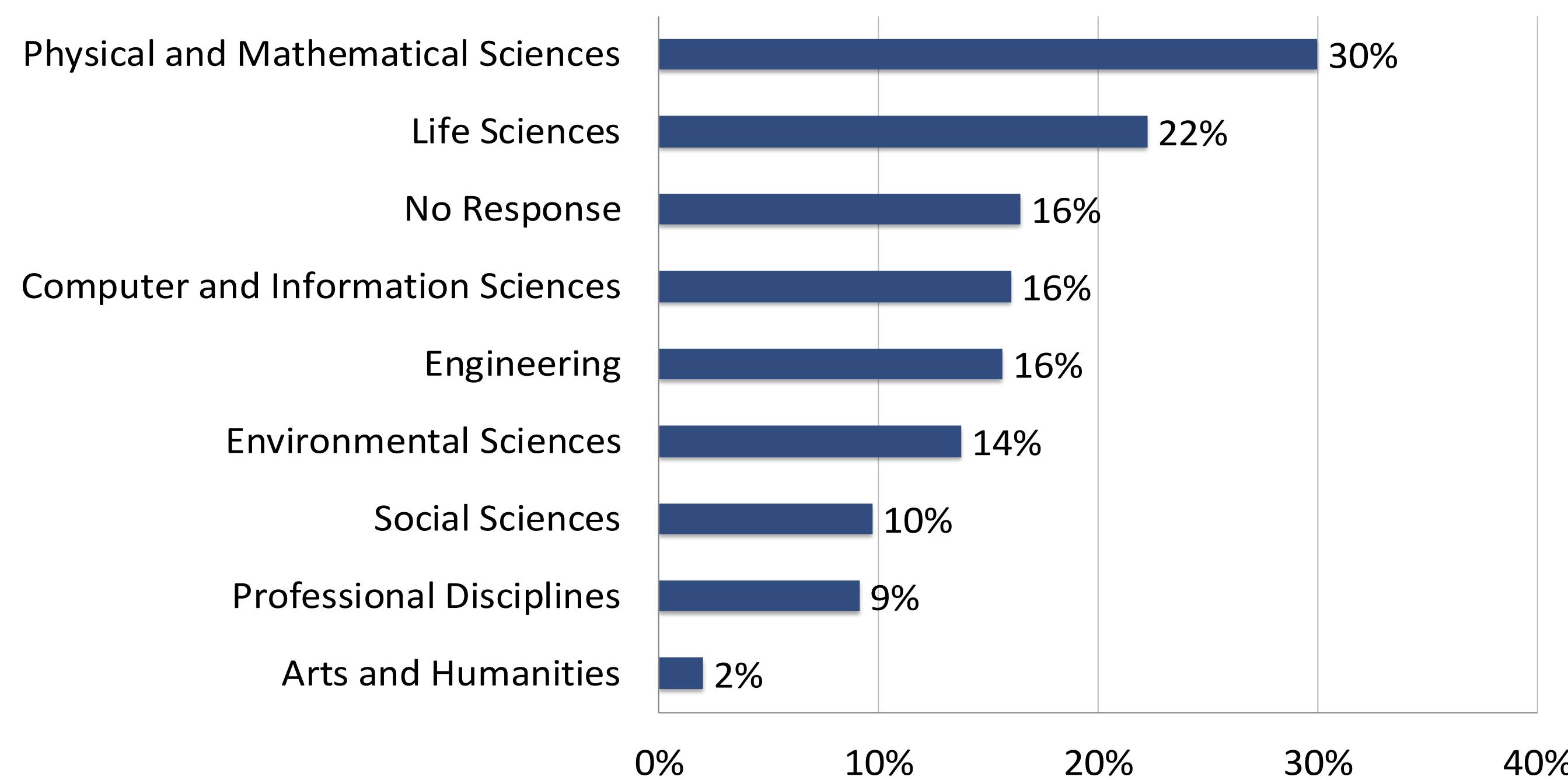
**Scott Michael**  
 Research Technologies, Indiana University

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 Stanzone offered their insight and ideas to the survey design. Additional help was provided by Maytal Dahan, Stephen Mock and Yan Zhou.

*This project was funded by the National Science Foundation under grant numbers 1216668, 1216723, 1216730, 1216733, 1216815, and 1216870.*

### Gateways Are Growing

Currently 40% of XSEDE users access resources through science gateways. XSEDE's ECSS program helps with the programming needed to make supercomputer and data resources available through gateways, but there is much room for additional growth for science gateways that do not strictly make use of computing resources. Through an NSF conceptualization-phase software institute grant, we undertook a large-scale survey to understand both how gateways are used across domains and what kind of services they might need for more efficient development. We developed the survey inductively, through expert interviews, focus groups, and pilot testing. The final 36 questions branched differently depending on whether a recipient was an administrator, researcher or faculty member, or technology developer.



**Figure 1. Primary areas of current domain expertise. Respondents could select all that apply; 84% (4,141 of 4,957) responded, generating 6,689 total responses (mean=1.6 domains per respondent).**

### Almost Two-thirds Consider Data Accessed Through Gateways To Be Important

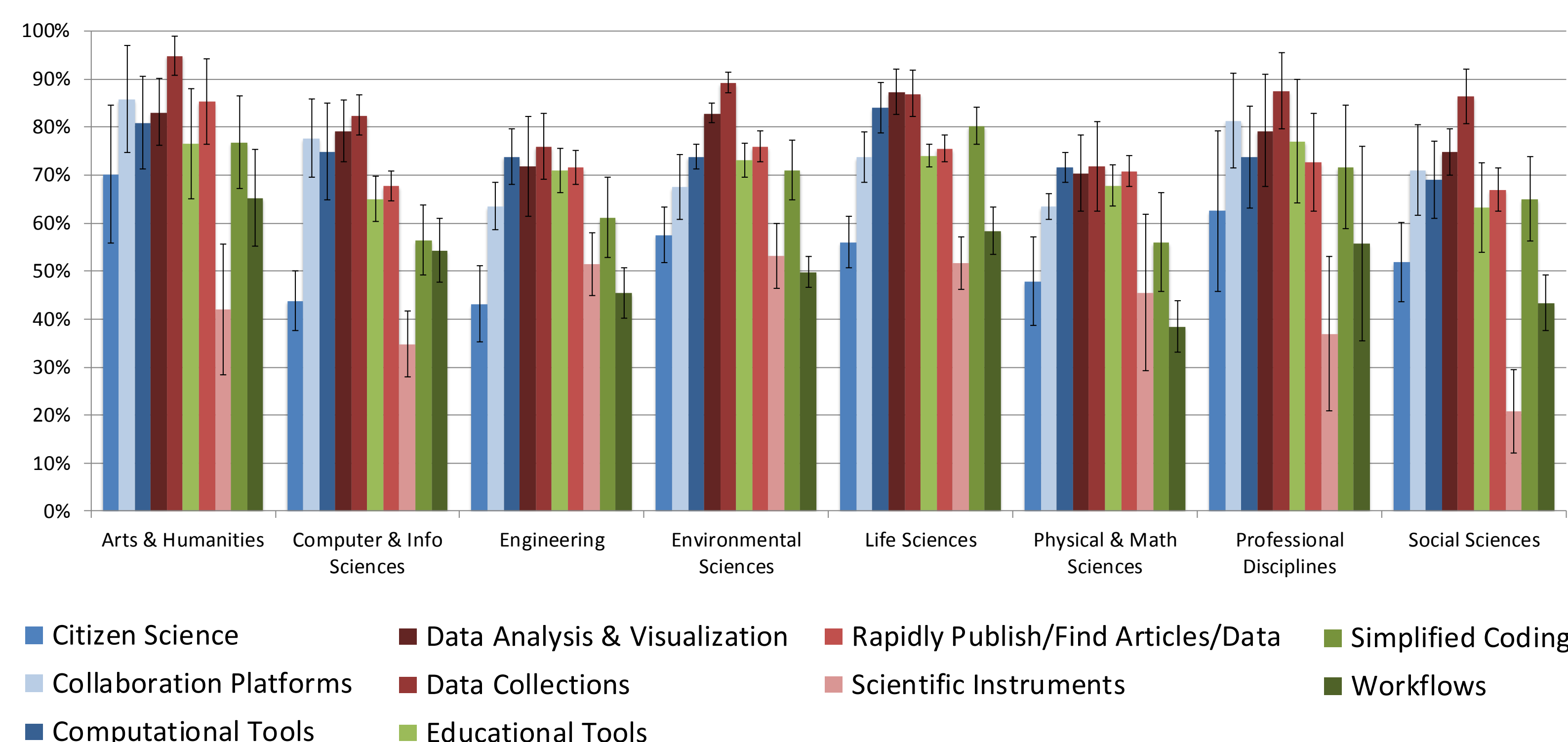
Gateways offer a wide variety of capabilities, as indicated by our application-creator respondents (Figure 2), and data services represent a third of applications created by respondents. We asked respondents who identified as researchers and/or educators how important to their work were the web-based applications providing access to specialized resources (Table 1). For accessing most types, at least 60% indicated that web-based applications were "somewhat" or "very" important, and data-related services were among the top choices.

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., their own research group, a commercial service, 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. For every category, "my own research group" dominated except for scientific instruments. Of particular interest, 58% of data collections and 50% of data analysis/visualization/mining tools were provided by their own research group. When looking only at externally provided resources (i.e., not created by "my own research group"), data collections are dominated by public/academic institutions as the external resource (39%) vs. commercial services (17%). Within each of the 8 high-level expertise groups, the proportion of respondents who indicated a resource

**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 (4,004, or 88% of 4,538)**

Specialized Resources	Percent
Data collections	75%
Data analysis tools, including visualization and mining	72%
Computational tools	72%
Tools for rapidly publishing and/or finding articles and data specific to my domain	69%
Educational tools	67%
Platforms for fostering group or community collaboration	63%
Simplified interfaces that eliminate the need to learn coding	62%
Citizen science and other public engagement resources	47%
Workflows that automate or capture tasks or processes	42%
Scientific instruments, such as telescopes, microscopes, or sensors	39%

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 (Figure 3).

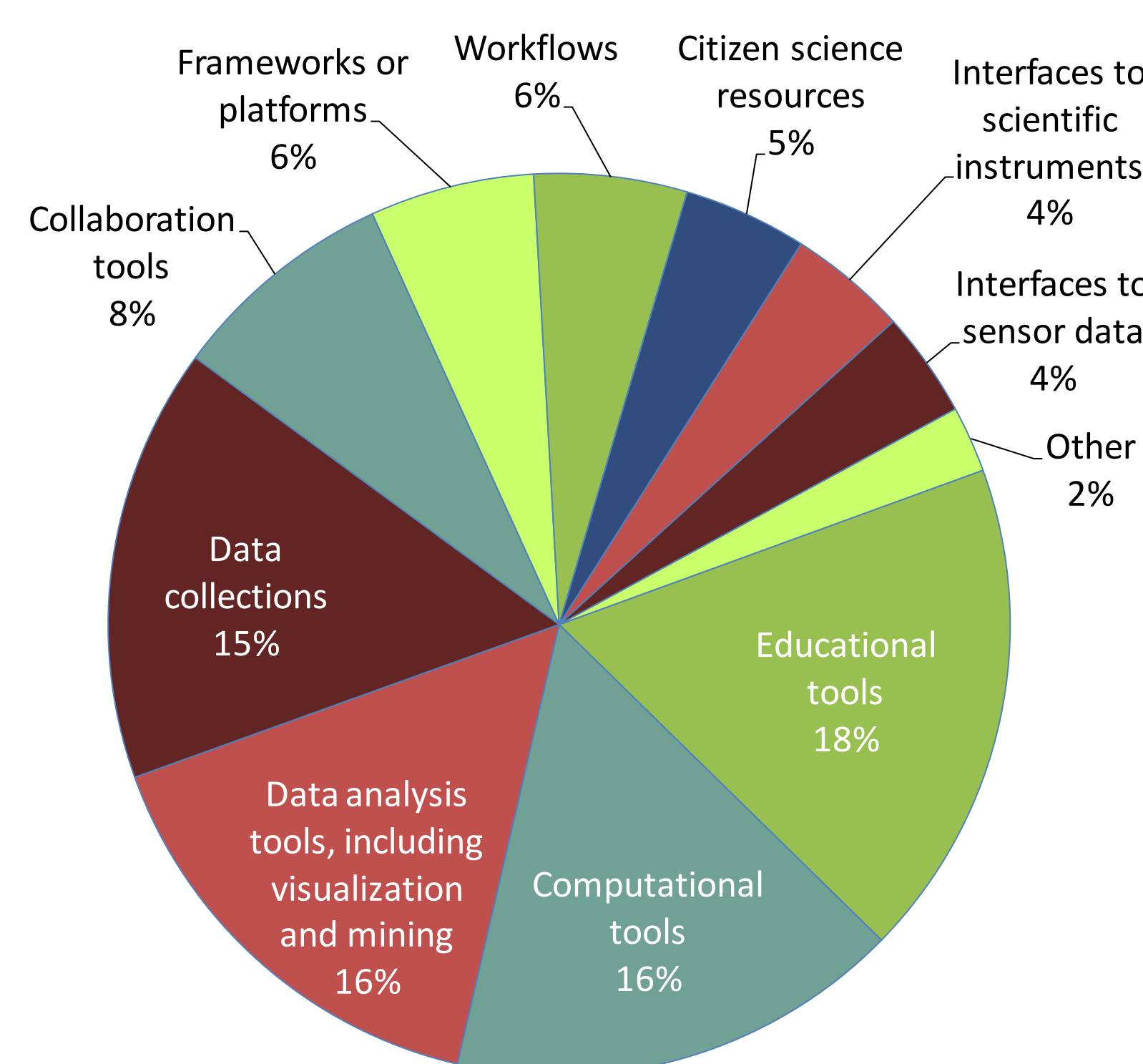


**Figure 3. Percent of respondents, within 59 expertise fields organized in 8 higher-level groups, who indicated a resource category was "somewhat" or "very" important. The average and standard deviation of their discipline-based percentages were calculated within each high-level expertise group (n=3793, which includes the overlapping set of those who answered both questions. Because respondents had to answer both questions to be included, the percentage of expressed interest is higher than other analyses which had more non-respondents. Note also that Arts & Humanities is minimally represented among NSF PIs).**

### > 50% Want Support

We asked the respondents who had participated in creating web- or mobile-based applications how they would anticipate needing help with their development projects. Many indicated a high interest in help with many of the functions associated with building a gateway (Table 2). In all but one of the more than 20 areas of potential support, at least 40% indicated that at least 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 (note: respondents sometimes selected more than one option). 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.



**Figure 2. Types of applications created by respondents (n of application types=7,805, by 2,756 creators (out of 2,819; mean=2.8 types/developer).**

**Table 2. Types of services relevant to administrators and application creators. Administrators (n=447) were asked what types of services and expertise their institution would use. Application creators were asked if they would seek help for a nearly identical array of services (n=2,153, or 76% of 2,819 creators).**

Proposed Service	% Interest or Demand	
	Creators	Administrators
Guidance with project sustainability*	62%	79%
Choosing and adapting to technologies*	66%	78%
Data mining/analysis resources	58%	73%
Evaluation, impact analysis, website analytics	72%	72%
Visualization resources	65%	72%
Legal perspectives	61%	70%
Developing open-source software	64%	68%
Community engagement mechanisms	62%	68%
Scientific instruments or data streams	44%	68%
Mobile technology development	59%	67%
Data management resources	60%	66%
Usability services	66%	65%
Education support resources (e.g., K-12, online)	64%	64%
Computational resources (local, remote, cloud)	59%	64%
Database structure, optimization, and query expertise	59%	63%
Cybersecurity consultation	57%	63%
Web/visual/graphic design	67%	59%
High-bandwidth networks	45%	58%
Website construction	57%	52%
Software engineering process consultation	53%	52%
Source code review and/or audit**	51%	
Management aspects of a project	38%	52%

\* 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.