Cyberinfrastructure value: a survey on perceived importance and usage

Praneeth Chityala  
Pervasive Technology Institute, Indiana University  
Bloomington, Indiana, USA  
pkchitya@iu.edu

Julie A. Wernert  
Pervasive Technology Institute, Indiana University  
Bloomington, Indiana, USA  
jwernert@iu.edu

Claudia M. Costa  
Pervasive Technology Institute, Indiana University  
Bloomington, Indiana, USA  
clcosta@iu.edu

Craig A. Stewart  
University of Illinois at Urbana-Champaign  
Urbana, IL, USA  
casncsa@illinois.edu

ABSTRACT
The research landscape in science and engineering is heavily reliant on computation and data storage. The intensity of computation required for many research projects illustrates the importance of the availability of high performance computing (HPC) resources and services. This paper summarizes the results of a recent study among principal investigators that attempts to measure the impact of the cyberinfrastructure resources allocated by the XSEDE (eXtreme Science and Engineering Discovery Environment) project to various research activities across the United States. Critical findings from this paper include: a majority of respondents report that the XSEDE environment is important or very important in completing their funded work, and two-thirds of our study’s respondents developed products (e.g., datasets, websites, software, etc.) using XSEDE-allocated resources. About one-third of respondents cited the importance of XSEDE-allocated resources in securing research funding. Respondents of this survey have secured approximately $3.3B in research funding from various sources, as self-reported by respondents.

CCS CONCEPTS
- Computer systems organization → Architectures;  
- Human-centered computing → Collaborative and social computing; Collaborative and social computing;  
- Applied computing → Operations research; Education; Education;

KEYWORDS
XSEDE, CI – Cyberinfrastructure, NSF

1 INTRODUCTION
Across the world, increasingly tight government and university budgets prompt increasingly stringent requirements for scientific research funding. Similar concerns persist in private-sector research and development organizations, which demand early evidence of viability and profitability. This issue is particularly acute for higher education institutions in the United States due to ever-increasing competition for federal research funding, coupled with financial pressures related to projected decreases in university enrollments [4]. Helping to bridge this gap, XSEDE is an organization that facilitates, allocates, and supports access to NSF-funded and other cyberinfrastructure (CI), providing researchers the secure computational and data storage resources required for their research. Such resources and services (e.g., supercomputers, collections of data, and computational tools) aid innovation and discovery [5].

This paper uses survey methods to ascertain the value of XSEDE to those who use its services, focusing on the opinions of principal investigators (PIs) who have received a resource allocation from September 1, 2016, to August 31, 2021. The majority of these allocations are for the use of computing systems funded by the NSF through separate funding actions (i.e. independent of XSEDE) to Service Providers. Some PIs also received allocations of consulting and programming effort directly from XSEDE. Thus, the services investigated here are a mix of system, software, and application services. To better understand the relationship between success in competing for grant awards and plans for executing the activities described in the funded proposals, the study inquired about the use of advanced cyberinfrastructure resources allocated by, provided by, or accessed via the XSEDE CI environment.

2 METHODS
This study was reviewed by the Indiana University (IU) Institutional Review Board prior to its initiation and was determined to be "expedited," meaning there were no sensitive data collected in the study. All participants provided their informed consent prior to taking part in the study.
The study population was identified from the XSEDE allocation database and consisted of PIs who received resource allocations during XSEDE2 program years, beginning September 2016 through August 2021. Only PIs with projects classified as "research" or "XSEDE Rapid Response" (a special allocation category for COVID research) were included. The survey was conducted via email with personalized links. Population members received an initial letter of invitation on January 13, 2022, followed by four reminder messages, the last of which being on February 4, 2022. The questionnaire consisted of 7-20 questions, depending on answers to screening questions. After accounting for bounced or undeliverable emails, the effective survey population consisted of 1595 members. Of these, 700 agreed to participate and went on to complete or partially answer questions in the survey. The questionnaire consisted of 7-20 questions, depending on answers to screening questions. After accounting for bounced or undeliverable emails, the effective survey population consisted of 1595 members. Of these, 700 agreed to participate and went on to complete or partially answer questions in the survey. The effective response rate was 44% (700 out of 1595)[3]. Such a high response rate speaks to the importance PIs place on the value of XSEDE in their research activities.

Lastly, but not least, the data used for our analysis are either from our survey or the XSEDE database. Our dataset [1] consists of three files: survey responses (section 3.1 to 3.3), field of science (Table 1), and grant awards (Table 2).

3 RESULTS
3.1 Importance of XSEDE in securing funds and completing funded work
The importance of XSEDE is tracked using two questions: What is the level of importance of XSEDE in 1) securing research funding and 2) completing funded work? The majority of respondents reported that XSEDE is important or very important in securing funds and completing funded work. Details are summarized in Fig. 1.

In addition, XSEDE’s Extended Collaborative Support Services (ECSS), often described as expert consulting services and a key service offering, was cited by 61% (106 of 175) of respondents who had used the service as either important or very important in completing their funded work.

3.2 Usage analysis of XSEDE-allocated resources
XSEDE provides cyberinfrastructure resources to support research activities in the two key areas: computational capabilities and data storage. Our survey captured two important metrics for the projects using XSEDE-allocated resources: 1) the percentage of computation performed on computational resources, and 2) the percentage of data stored on data storage resources. These questions are aimed at understanding the role of XSEDE-allocated resources and the individual’s research needs. Details are summarized in Fig. 2.

Further analysis indicated that 38% (265 of 700) of respondents used XSEDE-allocated resources for more than 75% of their project’s computational needs, and 19% (133 of 700) used XSEDE-allocated resources for more than 75% of their project’s data storage requirements. No respondents reported having not used XSEDE-allocated resources for their project’s computational needs; nine percent (62 of 700) of respondents indicated they had not used any XSEDE-allocated resources for their project’s data storage requirements. It was also reported that more than 89% (625 of 700) of respondents used their XSEDE allocation before it expired.

3.3 Research products developed using XSEDE-allocated resources
Our study also collected other important non-financial metrics, including the kinds of products developed using XSEDE-allocated resources. More than 66% (463 of 700) of responding PIs reported having developed products (in addition to publications) using XSEDE-allocated resources. Figure 3 illustrates the research products developed using XSEDE-allocated resources.

Further analysis showed that 61% (424 of 700) of respondents in completing funded work, and for 31% (220 of 700) for securing funds. One percent (6 of 700) of respondents reported that XSEDE resources are not at all important in completing funded work, and seven percent (51 of 700) reported XSEDE resources are not at all important for securing funds.

Another key finding is that 63% (444 of 700) of respondents stated that the funded work could not have been completed without the availability of XSEDE-allocated resources. Only seven percent (49 of 700) of respondents said they could have completed the funded work without XSEDE-allocated resources.
Table 1: Distribution of CI usage and product development across different fields of science

| Field of Science                        | Respondents | Products | Computation | Storage |
|----------------------------------------|-------------|----------|-------------|---------|
| Engineering and Technology             | 25%         | 27%      | 26%         | 27%     |
| Biological Sciences                    | 23%         | 21%      | 22%         | 21%     |
| Physical Sciences                      | 20%         | 9%       | 12%         | 21%     |
| Chemical Sciences                      | 12%         |          |             |         |
| Earth and Environmental Sciences       | 10%         |          |             |         |
| Computer and Information Sciences      | 5%          |          |             |         |
| Mathematics                            | 2%          |          |             |         |
| Social Sciences                        | 2%          |          |             |         |
| Other Natural Sciences                 | 1%          |          |             |         |
| Medical and Health Sciences            | 1%          |          |             |         |
| Other                                  | 1%          |          |             |         |

Figure 3: Percentage of respondents reporting that XSEDE was used to develop indicated products

3.4 Analysis across fields of science

XSEDE allocation data provided individual research activities and domains for 700 respondents representing 737 projects (multiple PIs had allocations on more than one project). This allocation data also contained grant values as self-reported by the researchers, but without the corresponding funding source. Corresponding grant values are available for only 495 of 700 responding PIs; grant values for the remaining 205 respondents are unknown and are considered as zero dollars in further analysis. Based on that data, we looked at the distribution of resources and products developed per domain area. The engineering and technology, biological sciences, and physical sciences domains command the greatest percentages of XSEDE resources in terms of computation and data storage and, in turn, report the greatest number of non-publication products produced. Further investigation shows that the percentage of products developed and resources used across various domains are similar in proportion to the percentage of respondents across various domains. Table 1 illustrates the CI usage and products developed over various domains (fields of science).

The total funding received by survey respondents for research activities supported by XSEDE was in excess of $3.3B as self-reported by the PIs. Approximately $2.3B of this ascribed to the 63% of PIs who said they would not have been able to complete their funded work without XSEDE-allocated resources. After linear extrapolation to the total population of 1595 XSEDE-allocated PIs, the total funding stands in excess of $7.5B. This speaks to the impact - financial, scientific, and otherwise - that the XSEDE project has had on the US scientific community. These grants are further distributed across fields of science as shown in Table 2.

The engineering and technology category accounts for a major share of the grants secured, representing 33% of the total, whereas grants from biological sciences, physical sciences, chemical sciences, and earth and environmental sciences individually represent from 10% to 20% of grants dollars secured.

4 DISCUSSION

The high response rate of 44% suggests that PIs recognize the importance of the XSEDE resources they have used. Such a rate of response for a survey that did not provide incentives for completion is considered above average[2].

Survey outcomes suggest that XSEDE is providing a great deal of value to the work of US researchers. Over 80% of respondents said that XSEDE was important or very important in completing funded research. This underscores the value XSEDE provides in supporting research that is regarded as “worthy” in peer-reviewed funding decisions made by federal agencies in particular.

Survey results are in some ways consistent with the general history of NSF-provided national cyberinfrastructure resources, and in other ways quite different. Consistent with the origins of the NSF Supercomputing Centers program of the 1980s, survey results suggest that computation is still a more sought after service than storage. Conversely, while physics and chemistry are anecdotally viewed as disciplines that use a great deal of NSF-provided resources, the highest number of responses came from those in engineering and technology and biological sciences, with these two disciplines accounting for almost half of our respondents.

The most frequent types of products reported were datasets and software. Interestingly, almost a third of the respondents indicated that they had developed some sort of training resource based on or about the use of XSEDE. Inventions were reported as one of the
outcomes for three percent of the respondents. Patents represent a particularly valuable and potentially marketable product; one percent of respondents indicated that a patent had resulted from their use of XSEDE.

The value of funded research supported by XSEDE is especially notable in light of the project’s overall budget. XSEDE’s total 11-year budget is less than $300M. The total amount of grant funding that survey respondents reported having been supported by XSEDE was more than $3B.

5 CONCLUSION

XSEDE plays a crucial role in allocating required cyberinfrastructure resources for many research activities across the United States. This paper has discussed the survey on usage of XSEDE-allocated cyberinfrastructure resources over the 2016-2021 period, along with datasets that provide the self-reported grant funding secured by individual research activities and the domain of research as per NSF. The high response rate of 44% speaks to the value PIs derived from their use of XSEDE allocated resources and services. A large majority of respondents said XSEDE-allocated resources were important or very important in completing their funded work. About 63% reported that they would not be able to complete their funded research without XSEDE-allocated resources which accounts to about $2.3B. Nearly 66% reported that they had developed products (e.g., websites, software, datasets, etc.) using these resources, testifying to the tangible outcomes of these resources. This implies the pivotal role of XSEDE-allocated CI resources in their research activities.

ACKNOWLEDGMENTS

This work was supported in part by NSF award 1548562 (John Towns, PI) and by the IU Pervasive Technology Institute. Thanks to Ken Hackworth and Dave Hart for providing XSEDE allocations data. Thanks to Winona Snapp-Childs and Marlon Pierce for providing valuable comments. And thanks to Kristol Hancock and Tonya Miles for expert editing services. Any opinions presented here are those of the authors and may not reflect those of supporting organizations.

REFERENCES

[1] Praneeth Chityala, Claudia Costa, Julie Wernet, and Craig Stewart. 2022. Datasets: Cyberinfrastructure Value: Perceived Importance and Usage. Indiana University. https://hdl.handle.net/2022/27600
[2] Elisabeth Deutskens, Ko De Ruyter, Martin Wetzels, and Paul Oosterveld. 2004. Response rate and response quality of internet-based surveys: an experimental study. Marketing letters 15, 1 (2004), 21–36.
[3] Susan Morton, Dinusha Bandara, Elizabeth Robinson, and Polly Atatoa-Carr. 2012. In the 21st Century, what is an acceptable response rate? Australian and New Zealand journal of public health 36 (04 2012), 106–8. https://doi.org/10.1111/j.1753-6405.2012.00854.x
[4] Craig A Stewart, David Y Hancock, Julie Wernert, Thomas Furlani, David Lifka, Alan Sill, Nicholas Berente, Donald F McMullen, Thomas Cheatham, Amy Apon, et al. 2019. Assessment of financial returns on investments in cyberinfrastructure facilities: A survey of current methods. In Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning). ACM, Chicago, IL, USA, 1–8. https://doi.org/10.1145/3352186.3352228
[5] XSEDE. 2022. About XSEDE. XSEDE. Retrieved April 1, 2022 from https://www.xsede.org/about/what-we-do