Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 14:06:29
REFERENCE NO: 249

Author Names & Affiliations

- Nancy Wilkins-Diehr - San Diego Supercomputer Center
- Michael Zentner - Purdue University
- Maron Pierce - Indiana University
- Maytal Dahan - Texas Advanced Computing Center
- Katherine Lawrence - University of Michigan School of Information
- Linda Hayden - Elizabeth City State University

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Multiple: computational science, cyberinfrastructure, organizational behavior

Title of Submission

The Importance of Science Gateways in the Research Landscape

Abstract (maximum ~200 words).

Science gateways, also known as virtual research environments, virtual laboratories, or web portals have enhanced the scientific landscape since the origin of the Web. They connect disparate components of research cyberinfrastructure — data collections, instruments, computational power, and often most importantly, people. They provide avenues through which many can access top-tier resources and contribute to scientific discovery regardless of geographic location or institutional status. A 2014 survey found that more than two-thirds of respondents depend on gateways to do their academic work; more than half also engaged in the development of gateways. This points to a need for well-designed science gateway capabilities and for more broadly funded activities for their creation, potentially benefitting diverse fields of research, such as astronomy, genomics, and ecology. Investment in gateways leverages other investments by deploying shared resources, eliminating duplicate efforts, and helping researchers use their time more efficiently and effectively. Recent investments in gateways, such as the Science Gateways Community Institute, which serves as a hub of expertise, technology, and community, are encouraging movements toward recognizing the value of gateways. Additionally, gateways in the classroom can serve as an educational resource and a motivator for students to pursue careers in gateway development.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

Several studies have highlighted the importance of gateways to the research community. A 5000-respondent survey in 2014 [1] of NSF principal investigators, as well as academic CIOs and CTOs, captured feedback from diverse domain and geographic areas. This represented a 17% response rate for an untargeted survey, indicating wide interest. Respondents were comprised primarily of faculty and research scientists, but also included members of higher-education leadership, graduate students, and technology developers.
Eighty-eight percent of respondents indicated that they depend on gateways to conduct their work. Gateways are in fact widely used by the community to access many types of specialized resources such as data collections (75%); data analysis, visualization and mining (72%); computational tools (72%); and tools for rapid publishing or finding articles of interest (69%). Ten different categories indicated interest from 39% or more of respondents.
Further, 57% of respondents participated in the development of gateways and 8% more intended to do so in the future. Respondents served in multiple roles in these projects: principal investigators, web developers, outreach specialists, subject matter experts, advisory committee members, and more. Again, these gateways were built for a variety of purposes. Leading categories included interfaces to educational tools (18%), computational tools (16%), data analysis tools (16%) and data collections (15%).
The research challenge here is that there is a widespread need for well-designed science gateway capabilities and a need for more broadly funded activities for their creation. In in-depth focus-group discussions conducted prior to the survey, participants described how gateways can address several specific grand challenges. Example fields include astronomy, genomics, and ecological systems.
Astronomers are attracted by the ease and efficiency of gateways. Analysis of “a billion-item sky survey” would be more efficient, less error-prone, and more reproducible, enabling astronomers to spend less time developing infrastructure and more time asking questions and getting answers.
Genomics work consists of time-consuming data mechanics. The data are noisy, and the tools are difficult to calibrate. So data must be curated and integrated, but processing non-numerical data or providing domain-specific annotations is expensive. A gateway that could link disparate data sets would add tremendous value. In particular, such advances would enable scientists to ask questions across data sets or within very large data sets. A gateway would also allow researchers to connect with others who have rare samples or identify potential collaborators working with similar types of data.
To help scientists understand ecological processes and systems at scale, this field needs to bring together experimentalists, observationalists, and modelers to enable model validation with experimental data. Newer projects like NEON and the Ocean Observatories Initiative provide high-resolution spatial and temporal sensor data that are ideal for this work. Researchers need to validate instrument measurements, too. The big challenges are social (large-scale communication) and technological (organization and format of data). Gateways could address both.
There is no shortage of research questions which could be advanced by improved interfaces to data and analysis capabilities.

[1] Lawrence, K. A., Zentner, M., Wilkins-Diehr, N., Wernert, J. A., Pierce, M., Marru, S., and Michael, S. (2015) Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community. Concurrency Computat.: Pract. Exper., 27:4252-4268. http://dx.doi.org/10.1002/cpe.3526.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

Expertly designed science gateways can amplify investments in other research infrastructure, providing highly usable interfaces to the diverse, interconnected components of expensive cyberinfrastructure — scientific instruments, streaming sensor data, supercomputers and more. Gateways can also create cost savings for the thousands of researchers working in independent laboratories by deploying shared capabilities.
In one example, many independent researchers may be funded to conduct molecular dynamic simulations to investigate cutting-edge questions. All of these researchers need access to software, and all need access to computing capabilities. They can each invest in developing their own codes and deploying and maintaining their own computing clusters, or they can put more of their funding into the human labor to conduct the inquiries, using optimized software and resources provided through a science gateway. This latter solution is a more efficient use of resources and applies researchers’ time toward the activities that create the most added value. However, it can’t happen if the gateways themselves are not funded.
Science gateways are often seen as infrastructure and not pure research. They can experience significant ongoing funding challenges, yet they increase the utility of other research investments, sometimes very significantly. Lack of funding for this important piece of the puzzle can sabotage much more substantial investments in research and infrastructure, pushing the research community to proceed with their

work in less cost-effective ways, by duplicating efforts in laboratories throughout the country.
There have been significant recent investments in this regard that we believe will have a continued place in 2030 and beyond. The Science Gateways Community Institute provides centralized expertise and a means for gateway developers to learn about technologies and other gateways as well as to share ideas with other developers by participating in conferences and publications . We encourage NSF to support even more far-reaching ideas going forward; for example, broaden the burden of supporting gateways through a call to create university gateway centers of excellence. With the right NSF incentives, more universities will begin to think of gateways as core infrastructure (that contribute to university branding and visibility) and make long-term commitments to their success.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

Attracting and retaining students from diverse communities for the important work of science gateway development is important in keeping the US at the forefront of science and engineering. Targeted outreach, training programs, internships and real career paths on campuses can all contribute to this goal. In addition, gateways themselves are an important piece of the educational process and are used in classrooms from high schools to graduate level courses. This can be widely expanded by encouraging the development of instructional materials highlighting gateways and curricula used in education in a public gateway catalog. In this way educators, regardless of their location and affiliation, can access top-quality resources such as supercomputers and instruments. The use of gateways also broadens the reach of cutting-edge research into the classroom.

Consent Statement

- “I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode).”