# Science Gateways: Sustainability via On-Campus Teams

Sandra Gesing
University of Notre Dame
Notre Dame, IN, USA
sandra.gesing@nd.edu

Katherine Lawrence
University of Michigan
Ann Arbor, MI, USA
kathla@umich.edu

Maytal Dahan
The University of Texas at Austin, TX, USA
maytal@tacc.utexas.edu

Marlon E. Pierce
Indiana University
Bloomington, IN, USA
marpierc@iu.edu

Nancy Wilkins-Diehr
University of California, San Diego
La Jolla, CA, USA
wilkinsn@sdsc.edu

Michael Zentner
Purdue University
West Lafayette, IN, USA
mzentner@purdue.edu

***Abstract*—The challenges for creators of specific science gateways are manifold, and the expertise needed for well-designed science gateways is very diverse. The sustainability of science gateways is crucial to serve communities effectively, efficiently and reliably. One measure to achieve greater sustainability of science gateways is establishing on-campus teams. Researchers are served more efficiently since the support by experienced developers reduces individual project investments, and a team can bring the diversity of required expertise for a well-designed science gateway. This paper goes into detail about the challenges and the benefits of on-campus groups and of sharing resources across a campus. We provide four successful cases, describe the services of the Science Gateways Community Institute (SGCI) to support the process in building such groups, and recommend strategies for using free campus resources.**

***Keywords—science gateways, science gateways community institute, sustainability, on-campus groups***

## I. Introduction

The importance of science gateways in research and education is evident in surveys (for example, [1] that was answered in 2014 by 4957 researchers across domains) and in mature, widely used frameworks such as HUBzero® [2] and Apache Airavata [3]. Also funding bodies have recognized their importance; science gateways are not only directly mentioned in solicitations by the National Science Foundation (NSF) [4], National Institutes of Health (NIH) [5] and Department of Energy (DOE) [6] but the NSF additionally funds the Science Gateways Community Institute (SGCI) [1,7]. SGCI supports researchers in enhancing their science via gateways connecting complex infrastructures as well as science gateway creators in efficiently creating and sustaining science gateways.

Software sustainability in general has gained more attention in the last 10 years in academia than ever before. This is evident in several trends such as the increasing number of software carpentry events, the workshop series Working Towards Sustainable Software for Science: Software and Experiences (WSSSPE) [8], solicitations of funding bodies providing funding for software sustainability, and funded institutes such as SGCI in the US and the Software Sustainability Institute (SSI) [9] in the UK. Researchers across a variety of domains recognize that their research heavily relies on scientific software or even produce scientific software themselves to achieve their research goals despite the fact that they may not be professionally trained in software engineering. In 2009 a survey [10] was published with answers from 2000 researchers.

- 91% answered that using scientific software is important for their own research
- 84% answered that developing scientific software is important for their own research
- 53.5% answered that they spend more time developing scientific software than they did 10 years ago
- 38% spend at least one fifth of their time developing software

Given these numbers, it is obvious that researchers can save time and effort by being supported via professional software developers and thus can more efficiently focus on their own research. We assume that we can approximately extend the survey numbers to science gateways as a specific subgroup of scientific software, and we additionally consider that not only professional software developers but experts in usability, security, quality assurance, project management, etc. are crucial for a well-designed science gateway [1]. Having on-campus groups with diversity of expertise and knowledge serving a variety of projects will ideally lead not only to achieving sustainability of science gateways but also to lowering the cost for single science gateways due to synergistic effects between projects and the sharing of resources across departments. This paper goes into details for four successful use cases, which are teams at four of SGCI's partner institutions. Subsequently, we will present challenges and strategies for establishing on-campus teams, sharing resources, and using free resources, and we conclude by giving an

NSF Award Id ACI-1547611

overview of SGCI's work to support the creation of on-campus teams.

## II. Related Work

Extreme Science and Engineering Discovery Environment (XSEDE) [11] has successfully established the model of campus champions to help researchers on their campus to gain access to XSEDE resources and receive support via such champions for effectively working with high-performance computing on a national scale. Additionally, XSEDE supports developers to connect their science gateways to XSEDE resources; in 2016, 77% of XSEDE users were accessing and using XSEDE resources via science gateways. Thus, in this case, the mixture of on-campus support and the support of science gateways has shown a great impact. The consortium Advanced CyberInfrastructure - Research and Education Facilitators (ACI-REF) [12] has a similar goal of establishing campus champions in a coordinated network. Its mission is to leverage existing resources and support their local campus researchers while also unifying member institutions under common objectives. Such facilitators are also supporting the development of science gateways and are on campus. While the facilitation already has shown impact in diverse projects, metrics and surveys are currently a work in progress.

A further group has been initiated and is led by Internet2 [13] to support the work of cyberpractitioners, including their career paths and incentives at institutions. This external advisory group suggests measures and best practices and aims at bringing developers to researchers to achieve an efficient and effective way to create cyberinfrastructure, which is a superordinate concept of science gateways. The group was started at the end of 2016, and best practices and outreach measures are under development. The initiation resulted out of the grant "EAGER: Fact-Gathering and Planning for a National-Scale Cyberpractitioner Program." [14]

The goal of WSSSPE working groups is sustainability of scientific software in general, considering among other issues the career paths of software developers at academic institutions and building a community around scientific software. While science gateway developers are a subgroup, incentives and career paths at academic institutions would be beneficial for on-campus teams for developing science gateways.

The UK SSI was funded in 2010 and serves the UK's research software community as well as partners on international level. Its goal is to cultivate better, more sustainable, research software to enable world-class research. The institute has received funding to employ research software engineers (RSEs) and coined the term and role of RSEs in academia. In their current state-of-the-nation report [15] they emphasize the benefit of centralization of teams of RSEs at institutions: "High quality staff can be attracted by providing a shared and stable home for RSEs with many interesting projects available. By establishing a service function, the creation of high-quality, impactful code can be the sole goal of the RSEs, who are not distracted by the competing career demands experienced by researchers. By aggregating demand at a wider level than the individual research group, RSE resources can be made available to projects that do not have sufficient need to hire a permanent RSE of their own." This is directly transferable to campus-based, science-gateway creator teams.

SGCI collaborates or is directly involved in the above-mentioned institutes or projects and one of its goals is to show the benefits of on-campus teams for researchers, for institutions, and for the sustainability of science gateways, as well as to influence the academic landscape in this direction.

## III. Benefits and Successful Use cases

We analyzed four successful use cases of institutions with on-campus teams: the Center for Research Computing (CRC) at the University of Notre Dame [16], the HUBzero Team at Purdue University [17], the Science Gateways Research Center at Indiana University [18], and the Advanced Computing Interfaces Department at TACC at the University of Texas, Austin [19]. Each of the use cases shows different strategies, priorities, and metrics for success and different lessons learned. All four teams are well established and provide their diverse expertise to their institutions in addition to various external communities and institutions. The benefits of on-campus teams for science gateway creation can be summarized in the following major topics:

- Great visibility for the institution's research activities
- Synergy effects between projects
- Shared resources, costs, and expertise across departments
- Lower learning curves, thus faster spin-up of projects
- Expertise that is otherwise difficult for individual projects to obtain
- Ability to retain top-quality research computing support by providing interesting projects

The following subsections go into detail for each of the teams.

### *A. The Center for Research Computing at the University of Notre Dame*

The CRC has the mission to engage in computational science, foster multidisciplinary research, and provide advanced computational tools and services. It works to facilitate discoveries across science, engineering, the arts and humanities, social sciences, business and other disciplines. While it started in 2006-2008 as center with services for high-throughput and high-performance computing and basic user support, it was extended with services for cyberinfrastructure development to fulfill its mission via diverse expertise including web development, grant support and visualization. Table I shows its growth in ten years of operation in regard to number of clients, cyberinfrastructure projects, team members as well as funding distribution. The goal was achieved by the director of the CRC via a roadmap with four steps (the last one is a recurring step):

- Take the risk and hire people first
- Train people
- Generate / bring projects
- Assign people to projects and focus on getting more projects and more people

Table I demonstrates how this example roadmap worked out well for growing the team at the CRC. There was a centralized team from the beginning, which was extended by an evangelist who felt strongly about extending the services in the direction of cyberinfrastructure development. About 20 team members work on science gateways in diverse roles: from

web development to backend development, visualization and consultancy services. In this team, 70% are on soft money brought in via contributing to grants or via working on re-charge for projects to implement science gateways for them.

TABLE I. THE GROWTH OF THE CRC DURING TEN YEARS BY ADDING CYBERINFRASTRUCTURE DEVELOPMENT

| | **2006-2008** | **2017** |
|---|---|---|
| Number of clients | ~300 active users | ~1800 active users |
| CI Projects | - | 15-20 per year with ~35 faculty from various departments including social sciences and humanities |
| Team size | 7 FTEs | 45 FTEs |
| Funding sources of FTEs | Centrally funded | 70% by grants and re-charge |

### *B. The HUBzero Team at Purdue University*

The HUBzero Team achieved growth and success with a different strategy than the CRC. It started in 1996 with the software PUNCH [20] to support the nanotechnology community with a vision by an evangelist. It was extended in 2002 to nanoHUB [21] with a diversity of services and as science gateway. The lessons learned from this process for a specific community were spun out to the science gateway framework HUBzero, and since 2015 there has been a broad diversification and self-funded sustainability. HUBzero incorporates a foundation, a development group at Purdue, and a software framework with hubs hosted by Purdue for a variety of institutions, with foundation members running their own hubs and open-source hubs. Each of the different stages and strategies was led by an evangelist. The team has grown exceptionally to 25 full-time software professionals (+5 vacancies) with a diversity of expertise:

- Cybersecurity
- Web programming
- User experience design
- Scientific application development
- Analytics
- Middleware
- High performance computing
- System administration
- Customer service

The team services over 2 million total visitors annually and is entirely self-funded. This successful use case was driven by one use case that has been expanded in different directions: vertically on services for one community, horizontally to serve various communities. Similar to the CRC, this example shows that an evangelist was crucial for a campus-based group to be successful.

### *C. The Science Gateways Research Center at Indiana University*

The Science Gateways Research Center was founded in 2012 and started its work based on lessons learned from former XSEDE science gateway projects and the broad experience in this area of the leader and involved team members. The team has been developing the Apache Airavata framework with various services including connectors to cloud, grid, and batch resources; integrated workflow features; and easy-to-use, integrated user interfaces. One key aspect in the strategy of the team is to create a talent reservoir and to invest a lot of effort in recruiting students, teaching classes in the computer science department, and mentoring Google Summer of Code students. Early in the project, they focused on getting people to contribute to their framework and projects, and they continue to keep alumni involved. While the team has not grown in the past five years in the number of team members, it is highly sustainable with more active projects, more contributors, and a regular staff turnover. This strategy—the involvement of a large, diverse community from the beginning and reaching out via teaching and involving students—turned out well for the center and is an example of another roadmap to a sustainable campus-based group. Also crucial in this case is the evangelist leading the group.

### *D. Advanced Computing Interfaces Department at TACC at the University of Texas, Austin*

TACC has an Advanced Computing Interfaces Department with four sub-groups that handle the entire pipeline of science gateway development. From diverse backend infrastructure such as cloud and VM infrastructure to software-as-a-service products, APIs, and portal interfaces, the department can handle the entire stack of science gateway needs. The department supports multiple projects with various requirements and different platforms and serves a wide range of projects from cyberinfrastructure to specific scientific domains and humanities. The group was started in 2002 and grew to 14 full-time staff members plus undergraduate interns, and it collaborates closely with other departments. The group participated in more than 15 proposals in the past 12 months. The long-term investment in the group resulted in several additional benefits for the university:

- Earlier awareness of disruptive change
- Front-line reports about changing research technology needs
- Non-traditional partnership opportunities
- Built-in evangelism for standards, best practices, and software use
- Frequent multi-domain and multi-department collaborative activities

This strategy illustrates that centrally funded groups for developing science gateways are not only sustainable but a benefit the whole university and departments within. The concept was implemented by an evangelist to enable this structure and has led to an excellent growth.

## IV. CHALLENGES AND STRATEGIES

Prior research about successful collaborations taking place on campuses and between partners elucidated that being at one location, trusting each other's work, and finding a common vocabulary in multidisciplinary teams contribute significantly to the efficiency of a project and its chances of success [22]. Also the use cases presented here and our analysis (as well as related work) shows that centralized teams for science gateway development have advantages for researchers, science gateway creators, and institutions.

However, there are also challenges to building such teams. One major challenge is to engage an evangelist who can advocate and is enthusiastic about the creation of on-campus groups. The evangelist must reach out and convince key people and decision makers about the benefits of on-campus teams. This individual has to be dedicated to the idea of sustaining such a group. Another significant concern is to maintain the diversity of expertise and experience in the group so that projects can benefit from such a centralized team. Conversely, the leader must investigate not only in his or her home department but in other departments and domains whether there are existing science gateway projects on campus and then contact Principal Investigators (PIs) and/or developers for a collaboration. The research domain might be different, but often there is some overlap on requirements (e.g., access to high-performance computing, secure access via on-campus accounts, increased usability of the user interface). Specialists might not be needed for the full length of a project, but may be beneficial in certain phases and can be fully funded by serving on several projects, allowing projects and PIs to each contribute to part of the salary of such an expert.

Funding of experts is another one of the major challenges for building teams. Thus, it is crucial to look at diverse funding possibilities and in particular at options for free resources. Institutions often offer subsidized collaboration opportunities and maybe even free resources. For example, often digital librarians are funded via hard money at universities. They are generally serving multiple fields of research, and they may have knowledge about data preservation, data lifecycles, and programming. Additionally, the university might have employed data scientists with knowledge about machine learning, meta-data, ontologies, and statistics, for example. If there is an existing HPC center, there are employees/researchers knowledgeable about HPC resources, distributed data management, etc. It could be a good start to reach out to such experts and find out some of their specific knowledge that may be beneficial for science gateway creation. Besides such potential free resources, there are also the classic funding mechanisms for internal resources such as applying for funding via involvement of some person-months in grants, funding on some hard money from universities, or funding via re-charge.

SGCI is an external source for free services. It is funded by the NSF to offer its services to science gateway projects. Thus, to experience the benefits of such subsidized services, science gateway projects can apply to SGCI to receive consulting support in many stages of the science gateway lifecycle and from various experts. The Incubator service area offers short-term consultations for up to three months and provides support by diverse experts. The main service topics include but are not limited to:

- Technology advice: cybersecurity; technology planning, open-source licensing and selection; and development tools and processes
- Business planning: business and strategic planning; sustainability planning; and project management
- Usability and user engagement expertise: engaging user communities; impact measurement and evaluation; usability and user-centered design; and graphic and user-interface design
- Campus-based development groups: one-to-one consultations; webinars and on-campus visits

Through the Incubator, projects can apply for help in specific areas of expertise, which may not be available on their campuses and add additional, temporary staff without extra cost. The goal is that project leaders and members experience the benefits of including specific expertise in the lifecycle. This may lead universities to employ such experts themselves and/or plan to hire consultants serving on several projects in the future. Another service of the Incubator is twice-yearly bootcamps [23] focusing on sustainability through consideration of various aspects such as the customer relationship, usability, funding, and cybersecurity.

The Incubator supports campus-based, centralized, gateway-development teams by providing direct consultation to campuses interested in implementing such teams. They can request an initial presentation by an SGCI consultant about the use and value of campus teams as well as more extensive advice and support for navigating the challenges associated with building such teams. With time, the goal is that these campuses can become resources for each other and for other campuses interested in adding a development team. In the future, online learning resources and best-practices guides developed by the Incubator, EDS, and SGCI's Community Engagement and Exchange service area will also make the Institute's specialized expertise and experience more broadly available for self-service implementation at campuses

Additionally, SGCI offers hands-on, development support for up to one year to successful applicants via Extended Developer Support (EDS). Projects can be supported with expertise in diverse programming languages and environments, different types of infrastructure, including data and high-performance computing systems and third-party tools for building gateways. EDS can support many stages of the science gateway lifecycle and can fill a gap on the development side until a first version of the science gateway or a specific feature is accomplished. The goal is that after the period of hands-on support, the project is more sustainable and/or has added the anticipated features and ready for maintenance. Further support beyond a year might still be possible but would be not subsidized any more through SGCI funds. Another option with external funding would be to temporarily hire contractors. These are potential measures for creating and/or maintaining science gateways successfully while working on building up internal resources at an institution.

## V. OUTLOOK

Sustainability of science gateways is an important topic for serving communities effectively, efficiently, and reliably. On-campus teams for the science-gateway creation process are a means of tackling this challenge, and SGCI provides services to support campuses to develop their own roadmap to such centralized teams. We are conducting interviews at several institutions to capture more successful use cases to explore common patterns and gather data about the growth of projects and/or about success metrics for such teams. We also are reaching out to researchers and evangelists to offer webinars and/or campus visits to present the benefits, and we will collaborate with diverse projects to address challenges faced by developers in academia. Via these actions, we aim at initiating

the creation of more on-campus groups. A long-term goal is to influence the academic landscape for science gateway creators and developers with improved career paths and incentives to stay in academia.

## ACKNOWLEDGMENT

This work of SGCI is supported by the National Science Foundation Award ACI-1547611.

## REFERENCES

[1] Lawrence, KA, Zentner, M, Wilkins-Diehr, N, Wernert, JA, Pierce, M, Marru, S, Michael, S. 2015. "Science gateways today and tomorrow: Positive perspectives of nearly 5,000 members of the research community," Concurrency and Computation: Practice and Experience 2015, DOI: 10.1002/cpe.3526.

[2] McLennan, Michael, and Rick Kennell. HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering. Computing in Science & Engineering 12, no. 2 (2010): 48-53.

[3] Marru, Suresh, Lahiru Gunathilake, Chathura Herath, Patanachai Tangchaisin, Marlon Pierce, Chris Mattmann, and Raminder Singh. Apache Airavata: A Framework for Distributed Applications and Computational Workflows. In Proceedings of the 2011 ACM Workshop on Gateway Computing Environments, pp. 21-28. ACM, 2011.

[4] http://www.nsf.org/

[5] https://www.nih.gov/

[6] https://energy.gov/

[7] Gesing, S., Wilkins-Diehr, N., Dahan, M., Lawrence, K., Zentner, M., Pierce, M., Hayden, L.B., and Marru, S. Science Gateways: The Long Road to the Birth of an Institute, Proc. of HICSS-50 (50th Hawaii International Conference on System Sciences), 4-7 January 2017, Hilton Waikoloa, HI, USA, http://hdl.handle.net/10125/41919

[8] Katz, D. S., Niemeyer, K. E., Gesing, S., Hwang, L., Bangerth, W., Hettrick, S., Idaszak, R., Salac, J., Chue Hong, N., Corrales, S. N., Allen, A., Geiger, R. S., Miller, J., Chen, E., Dubey, A., Lago, P. Report on the Fourth Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE4), arXiv:1602.02296 [cs.SE], 2017.

[9] https://www.software.ac.uk/

[10] J.E. Hannay et al., "How Do Scientists Develop and Use Scientific Software?", Proc. ICSE Workshop Software Eng. for Computational Science and Eng., pp. 1-8, 2009.

[11] http://xsede.org/

[12] https://aci-ref.org/

[13] http://www.internet2.edu/

[14] https://www.nsf.gov/awardsearch/showAward?AWD_ID=1546711

[15] https://www.software.ac.uk/blog/2017-04-10-state-nation-report-research-software-engineers-released

[16] https://crc.nd.edu/

[17] https://hubzero.org/

[18] https://sgrc.iu.edu/

[19] https://www.tacc.utexas.edu/

[20] FORTES, J. A. B., LUNDSTROM, M. S., AND KAPADIA, N. H. 1997. The semiconductor simulation hub: A network-based microelectronics simulation laboratory. In Proceedings of the 12th Biennial Symposium on University Government Industry Microelectronics. 72-77.

[21] https://nanohub.org/

[22] Judith S. Olson, Erik C. Hofer, Nathan Bos, Ann Zimmerman, Gary M. Olson, Daniel Cooney, and Ixchel Faniel. A Theory of Remote Scientific Collaboration. In Scientific Collaboration on the Internet. Edited by Gary M. Olson, Ann Zimmerman and Nathan Bos. MIT Press. 2008.

[23] Gesing, S., Zentner, M., Casavan, J., Hillery, B., Vorvoreanu, M., Heiland, R., Marru, S., Pierce, M., Mullinix, N., Maron, N. Science Gateways Incubator: Software Sustainability Meets Community Needs. WSSSPE 5.2 (Working Towards Sustainable Software for Science: Practice and Experiences), 24 October 2017, Auckland, New Zealand (accepted)