Hubzero: Community Growth for Four Science Gateways Supporting Open Science

Sandra Gesing, *Discovery Partners Institute, Chicago, IL, 60606, USA*

Claire Stirm, *San Diego Supercomputer Center, La Jolla, CA, 92093, USA*

Gerhard Klimeck, Lynn Zentner, and Su Wang, *Purdue University, West Lafayette, IN, 47907, USA*

Braulio M. Villegas-Martinez and Hector M. Moya-Cessa, *National Institute of Astrophysics, Optics, and Electronics, Tonantzintla, Puebla, 72840, Mexico*

Carrie Diaz Eaton, *Bates College, Lewiston, ME, 04240, USA*

Sam Donovan, *BioQUEST Curriculum Consortium, West Lafayette, IN, 47907, USA*

Carol Song, Lan Zhao, I Luk Kim, and Alejandro Strachan, *Purdue University, West Lafayette, IN, 47907, USA*

Michael Zentner, *San Diego Supercomputer Center, La Jolla, CA, 92093, USA*

Rajesh Kalyanam, *Purdue University, West Lafayette, IN, 47907, USA*

The research landscape has become increasingly interdisciplinary and complex with novel hardware, software, data, and lab instruments. The reproducibility of research results, usability of tools, and sharing of methods are all crucial for timely collaboration for research and teaching. Hubzero is a widely used science gateway framework designed to support online communities with efficient sharing and publication processes. This article discusses the growth of communities for the four science gateways nanoHUB, MyGeoHub, QUBEShub, and HubICL using the Hubzero platform to foster open science and tackling education with a diverse set of approaches and target communities.

Addressing pressing global challenges, such as climate change and sustainability, relies on interdisciplinary collaboration and open science practices. Ensuring reproducibility, essential in these high-consequence topics, can be facilitated by giving researchers the right tools. Science gateways streamline data and computing infrastructure usage while reducing complexity as needed. Over the past two decades, various science gateway frameworks have emerged, including Hubzero,[1] Galaxy,[2] and the Open Science Framework (OSF),[3] each offering distinct advantages and services. Hubzero, in particular, excels at integrating diverse user environments, such as Jupyter Notebook and RStudio, while providing access to cloud and distributed infrastructures for simulations and tools. Users can customize data and tool sharing from fully open access to private access, accommodating various stages of data privacy and concerns, including predepersonalized health data.

Open science tools and science gateway frameworks have shifted from a system-centric approach to a user-centric approach, catering to user preferences and ease of use. For open science frameworks to be sustainable, their adoption and continued use by the user communities are crucial. The more users adopt a specific tool or framework, the more likely it will be maintained and available for long-term use. However, little is known about the strategies that successful science gateways use to grow their communities and foster open science and education.

This article aims to bridge this knowledge gap by analyzing four distinct science gateways based on the Hubzero platform, each serving communities from diverse domains, such as nanotechnology, geology, biology, mathematics, and education. Our purpose is

1521-9615 © 2023 IEEE
Digital Object Identifier 10.1109/MCSE.2023.3285092
Date of publication 16 June 2023; date of current version 21 July 2023.

to identify the key factors contributing to the growth and success of these science gateways in open science and education. We will examine their similarities, differences, and unique features, emphasizing how these gateways have effectively fostered community engagement and increased adoption. By doing so, we hope to provide valuable insights for the future development and sustainability of science gateways in the open science landscape. This article is extended from our publication at HICSS 2022.[3]

OPEN SCIENCE AND GATEWAYS

For this article, we adapt the definition by the European Commission[4]: "Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools."

Diffusing knowledge and new collaborative tools have many facets, from the publication of open-access manuscripts; to the reproducibility of results; to the findable, accessible, interoperable, and reusable data (FAIR) principle[2] and the sharing of research objects such as artifacts. This article focuses on the research computing tools and frameworks used for open science and open education. According to the Science Gateways Community Institute,[4] "Science gateways allow science and engineering communities to access shared data, software, computing services, instruments, educational materials, and other resources specific to their disciplines."

Frameworks fitting this definition play a substantial role in the open science and open education space. Hubzero, Galaxy, and OSF are representatives in the space. Galaxy is a widely used workflow-enabled science gateway for easily creating and managing workflows with drag-and-drop mechanisms. It enables sharing such workflows and data in one Galaxy instance between users and between different Galaxy instances. OSF provides seamless connectivity to a diversity of file systems, like Google folders, and preservation or version-control systems, such as Zenodo, GitHub, and GitLab. Preservation systems have a specific role in open science since they ensure the capability to access data, tools, and artifacts in the long term.

HUBZERO PLATFORM

The four science gateways we present have in common their support for sharing open research products and the underlying cyberinfrastructure. The Hubzero platform supports the research and educational communities through 20 science gateways, known as hubs. The concept for the platform originated from nanoHUB around the nanotechnology community. A hub (an instance of the Hubzero platform) offers research projects a space to host analytical tools, publish data, share resources, collaborate, and build communities in a single web-based ecosystem. Through a hub, research communities can do the following:

- Offer a reliable and easy-to-use web platform for researchers and students to connect applications, visualizations, and models to computing resources.
- Share research codes with peers and receive a persistent interoperable identifier—a digital object identifier (DOI).
- Engage with peers in interactive spaces to share knowledge and ideas.
- Host interactive virtual learning opportunities for students and professionals.
- Provide open access to research products, community resources, curated curriculum, and more.

Container technologies like Docker and Singularity allow for the packaging of whole environments with tools and data and shipping them to different locations to reuse tools and reproduce results. The advantage of using containers is that dependencies to operating systems, library versions, etc. are stored in the container. Hubzero works in the background with Docker containers to allow for the seamless operation of tools in its back end with varying computing infrastructures.

NANOHUB

nanoHUB is one of the world's leading scientific gateways; it served more than 22,000 simulation users and more than 1.8 million unique visitors in the year 2020.[5] In 1996, nanoHUB's predecessor, the Purdue University Network Computing Hub (known as PUNCH), was created to enable researchers to share their research codes via web interfaces without any code rewrites. The original goal was to share research software for semiconductor electron transport to be used by experimental groups for designs. It quickly became obvious that some faculty members adopted these web-form-based tools for education. In 2002, with about 500 annual users, the Network for Computational Nanotechnology (NCN) was funded by the U.S. National Science Foundation (NSF) to develop and operate nanoHUB as a national-level center. The goal was to make advanced scientific software useful to domain experts (researchers and instructors) without the need to become computational experts. This was done via easy-to-use online apps and tools. nanoHUB was the first such end-to-end portal, enabling tool development and online deployment. In its early years, nanoHUB demonstrated the following:

- Research codes can be reused for (good) research by noncomputational experts.
- Research codes can be transitioned into education.
- Tool developers can be empowered and enticed to deploy their codes/products via nanoHUB.
- A university project can operate and support a global infrastructure.
- Adoption can extend well beyond the small group of creators.

A key distinction between nanoHUB and most other early science gateways is its drive to go beyond the accessibility of simulation engines (portal concepts) and enable usability by many users beyond computational experts. Tool developers on nanoHUB created scientific end-to-end user apps before the iPhone came to the market, running those apps in a computing cloud before the "cloud" became a thing.

Having demonstrated adoption and impact in education and research around the world via advanced user analytics, in 2017, nanoHUB generalized its vision: *to accelerate innovation through user-centric science and engineering*.

Beyond providing single-point services, such as an online simulation or a lecture/tutorial, the goal was to enable users to consume simulation products in various modalities. The mission drives the continual development of nanoHUB: *to make science and engineering products usable, discoverable, reproducible, and easy to create for learners, educators, researchers, and business professionals*.

The gathering of stakeholder requirements continues to point to a series of infrastructural developments required to transform the nanoHUB vision into a reality. These ongoing improvements form the service foundation for future nanoHUB users and customers. The NCN cyberplatform team uses a formal customer discovery effort to support the goal of sustainability beyond the current funding stream. This process guides how the team packages these infrastructural enhancements into user capabilities that fulfill discovered value propositions.

Such efforts over nearly two decades make nanoHUB a successful scientific portal that accelerates innovation in education and research via online simulations. This was recognized in 2020 with an R&D 100 award in the category of software/services. In 2020, nanoHUB served 22,612 simulation users. Figure 1 shows three major user classifications: 1) education use in structured and coordinated settings (classrooms);

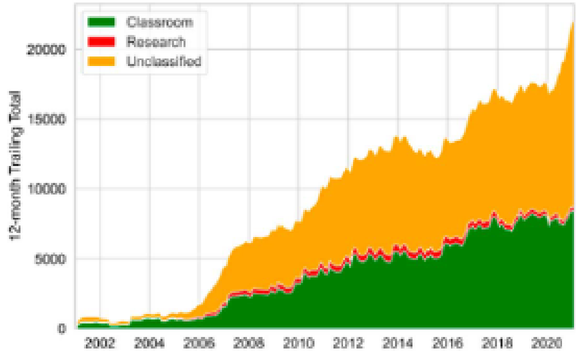

FIGURE 1. nanoHUB 12-month trailing simulation users categorized by education/classroom, research, and unclassified use.

2) use by individuals who have, in the past, cited nanoHUB (researchers); and 3) unclassified users. The impact of nanoHUB as a scientific knowledge exchange and simulation platform was validated and demonstrated in the year 2020 by the following:

- A total of 38% of users run simulations in structured education settings, as identified through coordinated user behavior.
- Figure 2 shows the seasonal education use in terms of users and institutions in a four-week average.
- More than 2500 cumulative citations with 54,329 secondary citations present a community body of work with an aggregate h-index of 105. These data show that *not only can nanoHUB used for research*, but the rate of secondary citations indicates the level of quality of the research.
- About 1%–2% of the active users have, in the past, cited nanoHUB in research publications.
- Efforts are underway to analyze and understand the behavior and goals of the 60% of users who are "unclassified."

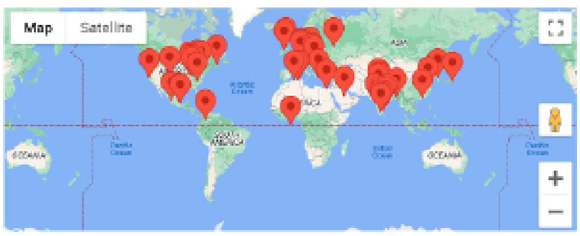

FIGURE 2. Users worldwide of the simulation "light propagation in photonic lattices."

nanoHUB online apps and tools are actual publications, and, since about 2005, nanoHUB has assigned DOIs to its simulation tools and compact models and preserves them via its own framework based on Hubzero. This effort put a stake in the ground that these online simulation tools are proper publications that enable the duplication of scientific results and use of authentic research codes by anyone in the world in an open-access forum. In 2017, this leadership was recognized by the Web of Science and Google Scholar, which now list nanoHUB tools as scientific publications that can reliably located.

The next fundamental challenge is to turn nanoHUB from a federally supported organization into a sustainable scientific knowledge exchange, delivery, and utility platform. The NCN team is increasingly focused on turning retrospective analytics into actionable analytics to drive nanoHUB toward sustainability.

The team also supports cutting-edge applications, such as simulations elucidating theories in quantum optics. An example is "light propagation in photonic lattices," a tool integrated in 2021 with the goal of presenting a theory that furthers a practical approach to quantum information and computation. While the tool has few users (55 in total), who originate from the relatively small theoretical quantum optics community, its worldwide uptake (see Figure 2) shows the potential of the growing community. Such an uptake in a bit more than a year would have been hardly achievable without an openly accessible science gateway.

MYGEOHUB

First released in 2014, the goal of MyGeoHub is to provide a cyberenvironment for geospatial data-driven research, education, and collaboration. MyGeoHub utilizes a shared hosting sustainability model whereby multiple open science projects are hosted as "supergroups" with a distinct look and feel while sharing the same underlying cyberinfrastructure. MyGeoHub builds on and extends the Hubzero platform with tools, software building blocks, services, and infrastructures that facilitate geospatial data access, processing, visualization, sharing, and publication. The Geospatial Data Analysis Building Blocks and GeoEDF (an extensible data framework designed to simplify data wrangling in geospatial research workflows) projects[6] developed and deployed easy-to-use libraries, tools, and services. They enable scientific users to connect large remote data repositories, data processing models and tools, and high-performance computing (HPC) resources in their workflows. More than 10 federally funded research projects are hosted on MyGeoHub. There are more than 9500 users who used MyGeoHub in the past year. Around 45 interactive online tools were published on MyGeoHub, most of which are open source.

MyGeoHub provides a cyberinfrastructure (CI) environment that promotes and enables FAIR-compliant practices. With automatic metadata extraction and documentation, data and tool publication with DOI assignment, open source online tool development and deployment, OAuth and CILogon authentication integration, and Representational State Transfer application programming interfaces for external programs to access project files and launch online tools, researchers are able to work on their data and research code across interoperable CI systems following the FAIR best practices. Furthermore, the latest addition of reusable and programmable data connector and processor modules and container-based workflow orchestration and submission to HPC resources has further reduced the time researchers spend wrangling large volumes of heterogeneous geospatial data. This enables the efficient creation of data-driven workflows that can execute in a variety of computational environments.

FIRST RELEASED IN 2014, THE GOAL OF MYGEOHUB IS TO PROVIDE A CYBERENVIRONMENT FOR GEOSPATIAL DATA-DRIVEN RESEARCH, EDUCATION, AND COLLABORATION.

The default Hubzero course platform on MyGeoHub was significantly enhanced to seamlessly connect scientific data and tool services, enabling interactive learning experiences for advanced training and workforce development. One of the main improvements was to integrate interactive coding environments, such as Jupyter Notebook and RStudio, with the course platform. This enables instructors to add coding exercises to their learning modules in which a Jupyter Notebook or RStudio session can be launched directly from the learning module with the example code automatically loaded for the students. There is no need for students to install any software or libraries either on their desktop or in their MyGeoHub environment, so they can focus on learning concepts, modeling, data processing, and visualization skills. In addition, all of the online models and tools published on MyGeoHub are directly accessible from the course platform, allowing students to get real-world modeling and simulation training by running research-grade online modeling tools using HPC resources.

The enhanced geospatial data and tool platform on MyGeoHub has attracted more and more education

FIGURE 3. Geospatial distribution of new education users registered on MyGeoHub in the past 12 months (May 2020–May 2021).

users in the past few years. The number of new education user registrations steadily increased in the past year, with two peaks in November 2020 and April 2021, corresponding to the academic calendars in higher education institutions. These new education users are distributed worldwide, covering six continents, as illustrated in Figure 3.

The education activities on MyGeoHub range from kindergarten through 12th grade (K–12) to postgraduate training in a variety of settings. This includes formal classrooms, summer camps, online tutorials, workshop training, and self-paced learning. As one example, the FAIR CyberTraining project developed two online courses on MyGeoHub to teach FAIR data practices in water and climate sciences. These online courses were used in teaching the "Data Mine I: Free and FAIR Climate Data" course in fall 2019 and the "FAIR CyberTraining for Water" and "Data Mine II: Free and FAIR Climate Data" courses in spring 2020, training four FAIR Cyber Training fellows in the summer of 2020 and delivering virtual tutorials at the 2021 FAIR workshop. During these events, participants received hands-on training on developing open source code for data access, processing, visualization, and publication following the FAIR principles; developed a new online course, "Python for Environmental Research"; published a new modeling tool for the California Food–Energy–Water System; and taught the developed materials at their home institutions. The integrated coding, data/tool publishing, and teaching platform on MyGeoHub have been a key success factor in the training activities of the CyberTraining project.

QUBESHUB AND THE RIOS INSTITUTE

Quantitative Undergraduate Biology Education and Synthesis (QUBES) was launched in 2014 with funding from multiple sources, including an NSF "IUSE Phase I Ideas Lab" convened to address the universal need for enhanced quantitative and computational expertise in the future biological sciences workforce.[7] The online hub was designed as a collaborative workspace where a consortium of diverse partners doing work at the interface of mathematics and biology education could share teaching and learning resources. Broadly, the project goals included building and supporting the use of a cyberinfrastructure to reduce barriers to interdisciplinary collaboration and shifting the community away from the inefficiencies of independently reinventing reform practices toward a coordinated, collaborative knowledge-building model. The Hubzero platform was chosen to host QUBES based on its capacity for managing parallel workspaces with integrated productivity and communications tools, a robust content publishing and access model, and an embedded cloud-based computational environment.

Beyond a focus on the functionality of the technical cyberinfrastructure, QUBES has adopted various strategies to engage and serve the needs of potential user communities. This parallel work on a social and professional infrastructure makes it possible to leverage the technical capacities of the cyberinfrastructure. It has been a key component of the QUBES community engagement. Strategies that support productive online collaboration have been refined over time and tested in many contexts. Faculty mentoring networks (FMNs) have proven to be a flexible and robust model for supporting distributed and diverse faculty communities. The project has run more than 70 FMNs with more than 1000 faculty participants. Additionally, it has developed models that use QUBES to support both face-to-face and online-only professional meetings. These experiences, working closely with the user communities, have informed the https://QUBESHub.org cyberinfrastructure as an educational gateway. An emphasis on accessing open data, using open source tools, sharing professional resources, and promoting a collaborative professional community have shaped the growth of QUBES as an open platform.

There are four primary ways in which QUBES has facilitated open science practices:

- QUBEShub serves as a repository for curriculum products tied to research publications.
- Some projects, such as the Network for Integrating Bioinformatics into Life Sciences Education,[5] use QUBES as an incubator where drafts of curriculum material are shared and collaboratively revised. FMNs and summer workshops provide similar functionality. While these groups are often

private to members until the final resource publication, even the final resource publication is considered to be a snapshot of a living resource that can continue to have new versions, and these publications offer the opportunity for any registered user to comment or fork. FMNs are also multi-institutional and, therefore, help to break barriers between labs and institutions. Some research and interest groups are fully open and share pedagogical discussions and products openly as well, though this is a smaller fraction of the overall number of active groups.

- The project partnered with several projects, such as the National Ecological Observatory Network,[6] that have open data repositories and want to broaden their educational impact and outreach—i.e., get their open data to be used in the classroom. By helping bring open data into the classroom, professional development helps faculty facilitate discussions about open science.
- Finally, there are several parallels between doing education in the open and doing science in the open. Evaluation research showed that individuals like to have opportunities to try things within a small community before feeling confident enough to post more publicly and that FMN experiences help instructors build that agency and confidence.

In 2019, the QUBES leadership helped cofound the Sustainability Challenges for Open Resources to promote an Equitable Undergraduate Biology Education (SCORE-UBE) Network sustainability challenges for open resources to promote equitable science, technology, engineering, and mathematics (STEM) education, later expanding under funding from the Hewlett Foundation to the Institute for a Racially Just, Inclusive, and Open STEM Education (the RIOS Institute). The impetus was QUBES'S work with partners that were struggling to fund the invisible labor and technological costs of open education while balancing a commitment to providing zero cost to users who submit or download curricular materials. At the center of this struggle is the question, "Open for whom?"—which asks us to center equity and inclusion in our community engagement and sustainability decisions.

The RIOS Institute primarily supports project leaders in STEM education, open educational resources, and related policy/administration to work through these difficult challenges that are at the intersection of open and inclusive.[8] With QUBEShub as a cyberinfrastructure partner, the RIOS Institute sponsors virtual learning communities, seminars, research working groups, and other collaborative opportunities for working on equity and inclusion in open science education.

RIOS asks how we can re-engineer our gateways and processes to center a commitment to inclusivity. We say "re-engineer" because, often, we modify systems to achieve goals, but systems that are designed around a goal operate more efficiently toward that goal. For gateways, the conversation on inclusivity has historically focused on accessibility—for example, mobile accessibility or accessibility for those who have vision impairments. We have also been thinking about how our choices around data—which data we collect and which types of tag frameworks we offer for our collections and products—are all signals of inclusion/exclusion.

We have also been thinking critically about our demographic data collection practices. The RIOS Institute has been experimenting with giving demographic surveys outside of the platform, e-mailed directly to users, so that data privacy levels are independently maintained with external evaluators. This gives us the opportunity to ask for qualitative descriptions of identity instead of just checkboxes. We hope this qualitative option will help us improve our user data collection questions in the future as well as support individuals in declaring their own identities.

INTERCULTURAL LEARNING HUB (HUBICL)

Intercultural learning is the process of "acquiring increased awareness of subjective cultural context (world view), including one's own, and developing greater ability to interact sensitively and competently across cultural contexts as both an immediate and long-term effect of change."[9] Intercultural learning is an applied knowledge designed from research and applied through experiential tools. Often, intercultural learning is used in the classroom for students traveling abroad and with international students. Usually, these experiential tools were stored in multiple open source repositories, on professional websites, and shared between peers. Practitioners or educators lacked a way to explore experiential tools without knowing what they were looking for or guidance from experienced intercultural learning professionals. HubICL[9] was launched in 2018 out of the Center for Intercultural Learning, Mentorship, Assessment, and Research (CILMAR), a unit within the Office of the Dean of International Programs at Purdue University.

HubICL caters to intercultural specialists, teachers, students, and professionals and enables these audiences to discover experiential tools, contribute an experiential tool, join virtual communities, and publish other open-access materials in a research repository. HubICL community members can access these open-access research and educational products by logging in to the HubICL platform. The most accessed feature of HubICL is the toolbox, a searchable collection of experiential tools.

Interculturalists can explore the 800 tools by searching for specific queries or explore via identifying materials, including what the practitioner or educator hopes to achieve from the activity.

- *Subgroup size*: The size of the participating groups.
- *External cost*: If the activity will have external costs related to purchasing or obtaining materials.
- *Duration*: The minimum and maximum of the activity time.
- *Tool type*: The type of activity, including experimental tools, assessments, media and texts, debriefing and reflection tools, and courses and training programs.
- *Kinesthetic*: If a physical activity.
- *Association of American Colleges and Universities rubric outcomes*: Learning outcomes, standardized by the Association of American Colleges and Universities.

These tools are sourced from submitted contributions, books on intercultural learning, and other resources. In addition, these tools offer group activities, curriculum design, and assessment materials. Just as any HubICL member can contribute a tool, any member can also leave a review on a published tool and advise other community members on best utilizing a tool. By sharing community insight, the members of HubICL can save fellow practitioners and educators time and resources.

HubICL is an expanding project. New features are being developed in collaboration with the Hubzero and HubICL teams to enable HubICL members to earn credentials as they learn about intercultural learning practices from the platform.

RESULTS AND CONCLUSION

"If you build it, they will come"[10] is rarely sufficient to attract a large community for a science gateway. Providers of science gateways need well-planned outreach measures, documentation, a usable and accessible platform, and the trust of users in the technology. The four presented science gateways are at different stages of maturity in this process. The project nanoHUB, with 25 years of operation, has paved the way with its vision on usability and adopting novel technologies while also thoroughly analyzing the needs of its community and usage patterns. MyGeoHub's unique feature is the building of "supergroups," tackling the needs of the user community for distinct features for projects with geological aspects. QUBEShub focuses on sharing teaching resources to support educators in undergraduate biology and mathematics, and the extensive use of discussion groups is one of its unique features. Since its inception, QUBES has expanded to support projects across STEM and K–12. The domain of HubICL is also education, and this is a newer project that started in 2018. Its approach is similar to those of the first three science gateways, providing a well-defined toolbox on topical areas shared in groups.

The lessons learned from the different science gateways include that they need a mixture of different outreach measures and that each community may prefer different formats. For example, publications, presentations, and online tutorials attracted and grew a large community for nanoHUB, while newsletters and e-mail campaigns were necessary to attract a wider community for further growth after the initial years. It is not always clear which features and outreach measures attract a wider community, though. Some communities pick up fast on Slack, while others barely use it for discussions. There is not one unique outreach approach that fits all. The outreach strength of MyGeoHub, for example, is the provision of tutorials with content about geospatial challenges paired with FAIR solutions. Addressing the educational needs of widespread concepts, such as the FAIR principles, is promising for attracting a wider community.

Another lesson learned is that educators and researchers in academia expect open access to data and tools. A recommendation is to have a basic set of features and tools for open access. Specific features for a community might be successful behind a paywall; however, the uptake is improved when part is openly available via so-called freemium models.

It is important to analyze the broader impacts by gathering consistent measures—for instance, the key user numbers for each science gateway: nanoHUB data from 1996 to 2020 show a total of 22,000 users and 1.8 million visitors; MyGeoHub from 2014 to 2020 counts more than 9500 users in the same period; QUBEShub reports 1000 faculty users; and HubICL, for the 2020–2022 period, reports a total of 3938 active members.

This article presents the growth of four communities in the open science and open education ecosystem by analyzing the features of four science gateways: nanoHUB, MyGeoHub, QUBEShub, and HubICL. Each science gateway tackles different community challenges by using the framework to share various data, simulations, and collaborative workflows. The presented science gateways are part of the open science ecosystem, with continued community growth and the use of shared products.

ACKNOWLEDGMENTS

The NSF has supported nanoHUB development through Award EEC-0228390, Award EEC-0634750, and Award EEC-1227110 as well as Award OCI-0438246 and Award OCI-0721680. (Office of Cyberinfrastructure is now Office of Advanced Cyberinfrastructure.) QUBES is currently supported by NSF Division of Undergraduate Education (DUE) Award 1446269, Award 1446284, and Award 1446258 as well as a number of partner project contracts. The RIOS Institute and former Sustainability Challenges for Open Resources Network have been supported by NSF Division of Biological Infrastructure (DBI) 1919879 and grants from the William and Flora Hewlett Foundation (Grant 2020-1363 and Grant 021-2923. The NSF has supported MyGeoHub development and projects on MyGeoHub through Award OISE-2020635, Award OAC-1261727, Award OAC-1829764, and Award CBET/ENG-1805741. The U.S. Department of Agriculture and the National Institute of Food and Agriculture have funded projects hosted on MyGeoHub. HubICL is managed by CILMAR and funded by the Office of the Dean of International Programs at Purdue University. We would like to acknowledge Kris Acheson-Clair and Annette Benson for the work described on HubICL. Hubzero is a registered trademark of the Regents of the University of California, San Diego. B.M. Villegas Martínez thanks Consejo Nacional de Ciencia y Tecnología (Mexico) for the postdoctoral scholarship.

REFERENCES

1. M. McLennan and R. Kennell, "HUBzero: A platform for dissemination and collaboration in computational science and engineering," *Comput. Sci. Eng.*, vol. 12, no. 2, pp. 48–53, Mar./Apr. 2010, doi: 10.1109/MCSE.2010.41.
2. M. Wilkinson et al., "The FAIR guiding principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, Mar. 2016, Art. no. 160018, doi: 10.1038/sdata.2016.18.
3. S. Gesing et al., "Open science via HUBzero: Exploring five science gateways supporting and growing their open science communities," in *Proc. 55th Hawaii Int. Conf. Syst. Sci. (HICSS), Open Sci. Pract. Inf. Syst. Res.*, 2022, doi: 10.24251/HICSS.2022.090. [Online]. Available: http://hdl.handle.net/10125/79420
4. S. Gesing et al., "Science gateways: The long road to the birth of an institute," in *Proc. 50th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Hilton Waikoloa, HI, USA, Jan. 2017, doi: 10.24251/HICSS.2017.755. [Online]. Available: http://hdl.handle.net/10125/41919
5. K. Madhavan, M. Zentner, and G. Klimeck, "Learning and research in the cloud," *Nature Nanotechnol.*, vol. 8, no. 11, pp. 786–789, Nov. 2013, doi: 10.1038/nnano.2013.231.
6. R. Kalyanam et al. *GeoEDF: A Plug-and-Play Workflow Framework for Earth Science Research.* (Oct. 2020). [Online Video]. Available: https://www.youtube.com/watch?v=0wflyfznv_Y
7. S. Donovan et al., "QUBES: A community focused on supporting teaching and learning in quantitative biology," *Lett. Biomathematics*, vol. 2, no. 1, pp. 46–55, Dec. 2015, doi: 10.30707/LiB2.1Donovan.
8. C. Diaz Eaton et al., "Sustainability and justice: Challenges and opportunities for an open STEM education," *CBE–Life Sci. Educ.*, vol. 21, no. 3, Sep. 2022, Art. no. es4, doi: 10.1187/cbe.20-08-0180.
9. M. J. Bennett, "Defining, measuring, and facilitating intercultural learning: A conceptual introduction to the Intercultural Education double supplement," *Intercultural Educ.*, vol. 20, no. sup1, pp. S1–S13, Jan. 2009, doi: 10.1080/14675980903370763.
10. K. F. Kee, B. Le, and K. Jitkajornwanich, "If you build it, promote it, and they trust you, then they will come: Diffusion strategies for science gateways and cyberinfrastructure adoption to harness big data in the science, technology, engineering, and mathematics (STEM) community," *Concurrency Comput., Pract. Experience*, vol. 33, no. 19, Oct. 2021, Art. no. e6192, doi: 10.1002/cpe.6192.

SANDRA GESING is a senior researcher and the scientific outreach and DEI lead at the Discovery Partners Institute, University of Illinois, Chicago, IL, 60606, USA. Her research interests include science gateways, computational workflows, and sustainability. Gesing received her Ph.D. degree in computer science from the University of Tuebingen, Germany. Contact her at sgesing@uillinois.edu.

CLAIRE STIRM is a project manager at the San Diego Supercomputer Center, La Jolla, CA, 92093, USA. Her research interests include science gateways, community building, and sustainability. Stirm received her M.S. degree in communications from Purdue University. Contact her at cstirm@ucsd.edu.

GERHARD KLIMECK is the Elmore Chaired Professor at the Elmore Family School of electrical and computer engineering and Reilly Chaired Director of the Center for Predictive Materials and Devices and the Network for Computational Nanotechnology (NCN) at Purdue University, West Lafayette,

IN, 47907, USA. His research interests include science gateways, nanotechnology, and sustainability. Klimeck received his Ph.D. degree in electrical engineering from Purdue University. He is a Fellow of IEEE. Contact him at gekco@purdue.edu.

LYNN ZENTNER is the executive director of the NCN at Purdue University, West Lafayette, IN, 47907, USA. Her research interests include cyberinfrastructure, sustainability, and community engagement. Zentner received her Ph.D. degree in mechanical engineering from Purdue University. Contact her at lzentner@purdue.edu.

SU WANG is a fourth-year Ph.D. candidate in electrical and computer engineering at Purdue University, West Lafayette, IN, 47907, USA. His research interests include federated learning, distributed machine learning, and web-based engineering education. Contact him at wang2506@purdue.edu.

BRAULIO M. VILLEGAS-MARTINEZ is a postdoctoral researcher at the National Institute of Astrophysics, Optics, and Electronics (INAOE), Tonantzintla, Puebla, 72840, Mexico. His research interests include quantum optics, PT-symmetry, and waveguides. Villegas-Martinez received his Ph.D. degree in quantum optics from INAOE. Contact him at bvillegas@inaoep.mx.

HECTOR M. MOYA-CESSA is a professor at INAOE, Puebla, 72840, Mexico. His research interests include quantum optics, Ion-laser interactions, and field propagation. Moya-Cessa received his Ph.D. degree in physics from Imperial College. Contact him at hmmc@inaoep.mx.

CARRIE DIAZ EATON is an associate professor of digital and computational studies at Bates College, Lewiston, ME, 04240, USA, and the director of strategy and advancement at the RIOS Institute. Her research interests include undergraduate STEM education, modeling, and data analysis. Eaton received her Ph.D. degree in evolution, ecology, and mathematics from the University of Tennessee, Knoxville. Contact her at cdeaton@bates.edu.

SAM DONOVAN is the director of outreach and strategic engagement with the BioQUEST Curriculum Consortium, West Lafayette, IN, 47907, USA. His research interests include data science education, ecology, and evolution. Donovan received his Ph.D. degree in science education from the University of Wisconsin, Madison. Contact him at sam.donovan@bioquest.org.

CAROL SONG is the chief scientist at the Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN, 47907, USA. Her research interests include science gateways, cyberinfrastructure, and geospatial research. Song received her Ph.D. degree in computer science from the University of Illinois Urbana-Champaign. Contact her at cxsong@purdue.edu.

LAN ZHAO is a research scientist at Purdue University, West Lafayette, IN, 47907, USA. Her research interests include cyberinfrastructure, streaming data management, Zhao received her M.S. degree in computer science and her M.S. degree in biology from the University of Illinois Urbana-Champaign. Contact her at zhao4@purdue.edu.

I LUK KIM is a Ph.D. candidate in computer science at Purdue University, West Lafayette, IN, 47907, USA. His research interests include cyberinfrastructure, web security, and program analysis. Kim received his Ph.D. degree in computer science from Purdue University. Contact him at kim1634@purdue.edu.

ALEJANDRO STRACHAN is a professor of materials engineering at Purdue University, West Lafayette, IN, 47907, USA. His research interests include materials engineering, nanotechnology, and reusability. Strachan received his Ph.D. degree in physics from the University of Buenos Aires. Contact him at strachan@purdue.edu.

MICHAEL ZENTNER is the director of sustainable scientific software at the San Diego Supercomputer Center, La Jolla, CA, 92093, USA. His research interests include science gateways, sustainability, and business models. Zentner received his Ph.D. degree in chemical engineering from Purdue University. Contact him at mzentner@ucsd.edu.

RAJESH KALYANAM is a research scientist at Purdue University, West Lafayette, IN, 47907, USA. His research interests include cyberinfrastructure, automated reasoning, and cybersecurity. Kalyanam received his Ph.D. degree in computer engineering from Purdue University. Contact him at rkalyana@purdue.edu.