Open Science via HUBzero: Exploring Five Science Gateways Supporting and Growing their Open Science Communities

Sandra Gesing
Discovery Partner Institute
sgesing@uillinois.edu

Claire Stirm
San Diego Supercomputer Center
cstirm@sdsc.edu

Gerhard Klimeck
Purdue University
gekco@purdue.edu

Lynn Zentner
Purdue University
lzentner@purdue.edu

Su Wang
Purdue University
wang2506@purdue.edu

Braulio M. Villegas Martinez
Instituto Nacional de Astrofísica, Óptica y Electrónica
bvillegas@inaoep.mx

Carrie Diaz Eaton
Bates College
cdeaton@bates.edu

Sam Donovan
BioQUEST Curriculum Consortium
sam.donovan@bioquest.org

Lan Zhao
Purdue University
lanzhao@purdue.edu

Carol Song
Purdue University
cxsong@purdue.edu

I Luk Kim
Purdue University
kim1634@purdue.edu

Alejandro Strachan
Purdue University
strachan@purdue.edu

Michael Zentner
San Diego Supercomputer Center
mzentner@ucsd.edu

Rajesh Kalyanam
Purdue University
rkalyana@purdue.edu

Abstract

The research landscape applying computational methods has become increasingly interdisciplinary and complex regarding the research computing ecosystem with novel hardware, software, data, and lab instruments. Reproducibility of research results, the 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. The paper discusses the growth of communities for the five science gateways nanoHUB, MyGeoHub, QUBEShub, CUE4CHNG, and HubICL using the HUBzero Platform to foster open science and tackling education with a diverse set of approaches and target communities. The presented methods and magnitude of the communities elucidate successful means for science gateways for fostering open science and open education.

1. Introduction

Tackling grand challenges such as climate change, global sustainability for water, food, land use, and lowering energy consumption requires interdisciplinary teams to access efficient collaboration methods and open science tools and frameworks. The worldwide COVID-19 pandemic has elucidated the strengths of science teams collaborating beyond borders and the importance of sharing data and reproducibility of research results [1]. Reproducibility is one aspect of open science and a cornerstone of science addressed by many tools and concepts for computational methods. So-called science gateways are end-to-end solutions enabling efficient use of data and computing infrastructures while hiding the computational complexity as far as desired. Quite a few mature science gateway frameworks have been developed in the last 20 years with different strengths and foci. HUBzero® [2], Galaxy [3], and the Open Science Framework (OSF) [4] are examples of widely used frameworks with frontend, backend, and middleware for different services. One of the strengths of HUBzero is the integration of varying user environments such as Jupyter notebook and RStudio while enabling access to Cloud and distributed infrastructures and submission of simulations and tools. Users can share their data and tools fine-grained from fully open access for a project team to only privateaccess for tackling different stages of data and

URI: https://hdl.handle.net/10125/79420
978-0-9981331-5-7

privacy concerns, e.g., health data before it is de-personalized.

Besides sharing and reproducibility, the FAIR principles (findability, accessibility, interoperability, reusability) [5] are essential concepts for open science and open education. Currently, researchers investigate maturity models and testing possibilities on how to fulfill these different aspects. One framework example is the Preservation Quality Tool (PresQT) [6], which connects preservation systems and science gateways to easily transfer and test the FAIRness of data and tools. These tests are suggestions for users and developers to fulfill different aspects of the FAIR principles. Additionally, users can add helpful metadata like keywords automatically via the PresQT services.

A further important aspect of open science is the usability of tools and science gateway frameworks. Users decide which novel technologies to use by asking a diverse set of questions [7].

- Is it easy to use?
- Easy to learn?
- Time-consuming or efficient?
- Can they do this?
- Do they have the knowledge, support, and resources?
- Does it fit in with their work style?

Thus, the computational environment has developed from a system-centric approach with users expected to learn the use of tools to a user-centric approach that considers their preferences.

The sustainability of open science frameworks is dependent on their use and uptake by the user communities. The more users adopt a specific tool or framework, the more likely it will be further maintained and available for long-term use.

This paper details five different science gateways and their measures to achieve growth of the communities by offering open science and open educational resources. All five science gateways are based on the HUBzero Platform. The science gateways have been chosen for the analysis because of their different stages of maturity and serving communities from a diverse set of domains, e.g., nanotechnology, geology, biology, mathematics, education. All five have the goal to foster open science and education, they use different concepts and features of HUBzero though. The discussion will highlight the differences, commonalities, and key findings for offering science gateways in open science and education.

2. Background

There are several definitions for open science and open education that are adapted in the academic community [8, 9]. They are distinguished from each other in regard of focus: whether it is on collaboration or novel tools and at the level of openness. For this paper we adapt the definition by the European Commission "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" [10].

Diffusing knowledge and new collaborative tools have many facets, from the publication of open-access manuscripts to the reproducibility of results to FAIR principles and sharing research objects such as artifacts. This paper focuses on tools and frameworks used for open science and open education in research computing. The Science Gateways Community Institute (SGCI) [11] defines that "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 rightfully a substantial role in the open science and open education space. HUBzero, Galaxy, and OSF are representatives in the space and we go into detail for five communities for HUBzero. Galaxy is a widely used workflow-enabled science gateway with its strength in easily creating and managing workflows with drag-and-drop mechanisms. Furthermore, it enables sharing such workflows and data in one Galaxy instance between users and between different Galaxy instances and other workflow-enabled science gateways such as Taverna [12]. Strengths of OSF include the seamless connectivity to a diversity of file systems like Google folders and preservation systems such as Zenodo [13], GitHub [14], and Gitlab [15]. Preservation systems have a specific role in open science since they assure the capability to access data, tools, and artifacts in the long term.

Container technologies like Docker [16] and Singularity [17] 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 seamless operation of tools in its backend with varying computing infrastructures.

3. HUBzero Platform

The commonality between these five science gateways is their support to sustain open research products and using the same cyberinfrastructure. The HUBzero Platform supports the research and educational communities through 20 science gateways, known as hubs. The original concept for the platform originated from nanoHUB around the nanotechnology community [18, 19]. 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:

- Offer a reliable and straightforward 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, 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.

4. nanoHUB

nanoHUB is one of the world's leading scientific gateways and served over 22,000 simulation users and over 1.8 million unique visitors in the year 2020 [20]. In 1996, nanoHUB's predecessor PUNCH (Purdue University Network Computing Hub) was created to enable researchers to share their research codes via web interfaces without any code rewrites [21]. 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 US National Science Foundation 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 that:

- Research codes can be reused for (good) research by non-computational 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 online simulation or a lecture/tutorial, the goal was to enable users to consume simulation products in various modalities. For example, evolving the original offerings within the site to embedded apps, to desktop and mobile apps, using nanoHUB web services. In addition, nanoHUB seeks to be part of users' day-to-day research, tool development, or education workflows with real-time user behavior analytics shaping the way users experience nanoHUB. 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*.

Continual stakeholder requirements gathering continues to point to a series of infrastructural developments required to transform the nanoHUB vision into a reality. These ongoing implementations form the scalable service foundation for future nanoHUB users and customers. The NCN cyber-platform 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 in the category of Software/Services [20]. In 2020 nanoHUB served 22,612 simulation users. Figure 1 shows three major user classifications: 1) education use in structured and coordinated settings (classrooms), 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 has been validated and demonstrated in the year 2020 by:

- 38% of the users run simulations in structured education settings as identified through coordinated user behavior [21].
- Figure 2 shows the seasonal education use in terms of users and institutions in a 4-week average.
- In 2012, nanoHUB measured adoption time from tool publication to first time use in a classroom with a median time of less than 6 months [20].
- Over 2,500 cumulative citations with 54,329 secondary citations present a community body of work with an aggregate h-index of 105. This data shows that nanoHUB *not only can be used for research*, but the rate of secondary citations indicates the level of quality of 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% “unclassified” users.

nanoHUB online apps and tools are actual publications, and since about 2005, nanoHUB has assigned DOIs to its simulation tools and compact models. This effort put a stake in the ground that these online simulation tools are proper publications that enable 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 proper scientific publications.

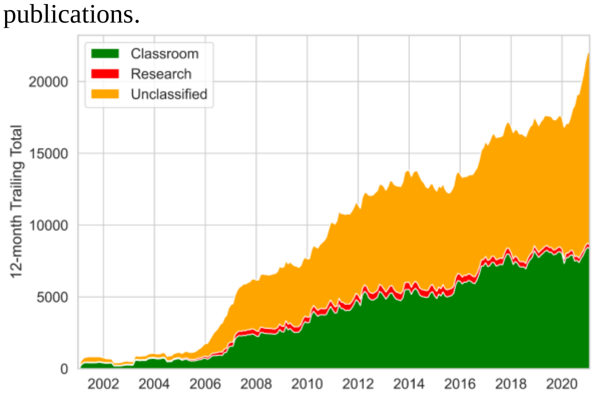


Figure 1. nanoHUB 12-month trailing simulation users categorized by education/classroom, research, and unclassified use.

The fundamental next challenge is to turn nanoHUB from a federally supported organization into a sustainable scientific knowledge exchange, delivery, and utility platform. Conceptually nanoHUB is similar to Uber or Airbnb in their early phases. Such platforms depend on a deep understanding of all users and providers to create a viable and sustainable market. Likewise, the NCN team is increasingly focused on turning retrospective analytics into actionable analytics to drive nanoHUB towards sustainability.

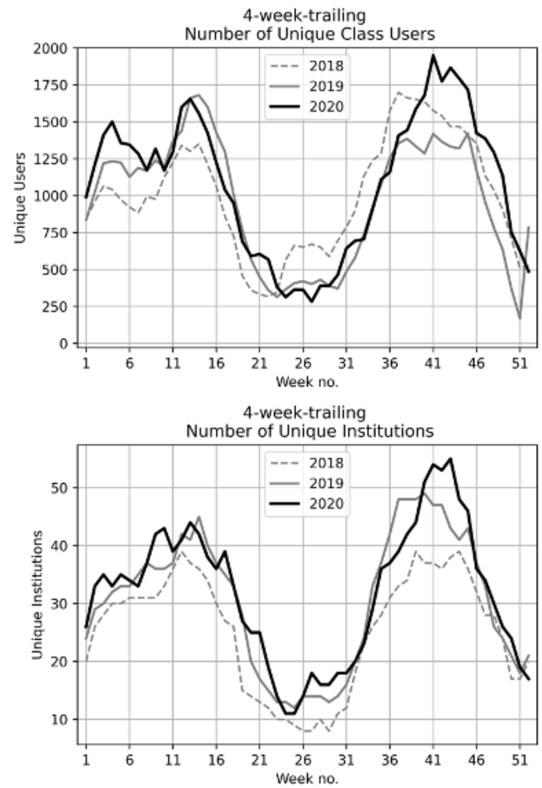


Figure 2. 4-week trailing simulation user numbers in formal classroom settings (top) and associated institutions (bottom) for the years 2018-20.

5. MyGeoHub

First released in 2014, the goal of MyGeoHub is to provide a value-added cyber environment for geospatial data driven research, education, and collaboration [22-23]. MyGeoHub utilizes a shared hosting sustainability model whereby multiple open science projects are hosted as “supergroups” with distinct look-and-feel while sharing the same underlying cyberinfrastructure.. MyGeoHub builds on and extends the out-of-box HUBzero platform with tools, software building blocks, services, and infrastructures that facilitate geospatial data access, processing, visualization, sharing, and publication. Funded by the NSF DIBBS and CSSI programs, the GABBs and GeoEDF projects [24-25] developed and deployed easy-to-use libraries, tools, and services that enable scientific users to connect large remote data repositories, data processing models and tools, and HPC resources in their workflows on

MyGeoHub. More than 10 federally funded research projects, all with synergies in geospatial data processing and management, 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 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 REST APIs 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 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 computation environments.

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 [26]. One of the main improvements was to integrate interactive coding environments such as Jupyter Notebook and RStudio with the course platform. Enabling instructors to add coding exercises to their learning modules in which a Jupyter Notebook or RStudio 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 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 high performance computing resources.

The enhanced geospatial data and tool platform on MyGeoHub has attracted more and more education users in the past few years. As shown in Figure 3, 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 4.

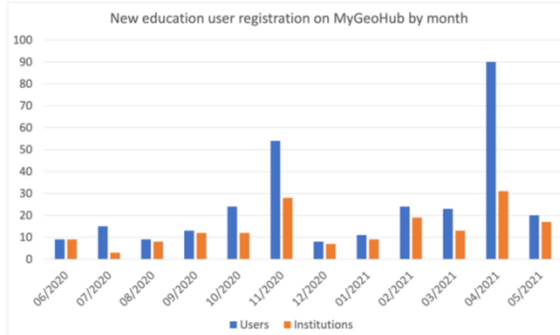


Figure 3. New education user registration on MyGeoHub and the number of institutions they came from in the past 12 months.

Figure 4. Geospatial distribution of new education users registered on MyGeoHub in the past 12 months (May 2020 - May 2021).

The education activities on MyGeoHub range from K-12 to post graduate training in a variety of settings. This includes formal classroom, summer camp, online tutorial, 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 [27]. These online courses were used in teaching the “Data Mine I: Free & FAIR Climate Data” course in fall 2019, the “FAIR CyberTraining for Water” and “Data Mine II: Free & FAIR Climate Data” courses in spring 2020, training four FAIR Cyber Training (FACT) fellows in the summer of 2020 and delivering virtual tutorials at the 2021 FAIR workshop [28]. 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 titled “Python for Environmental Research”, published a new modeling tool for California Food-Energy-Water System (CALFEWS), and taught the developed materials in 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.

6. QUBEShub & the RIOS Institute

QUBES (Quantitative Undergraduate Biology Education and Synthesis) 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 [29]. 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 over 70 FMNs with over 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 QUBESHub.org cyberinfrastructure as an educational gateway. An emphasis on the open education lifecycle, 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. 1) QUBESHub serves as a repository for educational curriculum products tied to research publications. For example, in a special issue of the mathematics teaching and learning journal PRIMUS, authors provided information about how their research article informed their teaching practices shared in supplementary materials or linked to a published resource on QUBESHub [29]. 2) Some projects, such as NIBLSE, use QUBES as an incubator where drafts of curriculum material are shared and collaboratively revised [30]. FMNs and Summer Workshop provide similar functionality. While these groups are often private to members until final resource publication, even the final resource publication is considered a snapshot of a living resource that can continue to version, 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. 3) The project partnered with several projects such as NEON [31] 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. 4) Lastly, there are several parallels between doing education in the open and doing science in the open. Evaluation research showed that individuals liked to have opportunities to try things within a small community before feeling confident enough to post more publicly and that FMN experiences helped instructors build that agency and confidence (unpublished).

In 2019, QUBES Leadership helped co-found the SCORE-UBE Network, Sustainability Challenges for Open Resources to promote and Equitable STEM Education [32], later expanding under funding from the Hewlett Foundation to the Institute for a Racially just, Inclusive, and Open STEM education (RIOS Institute). The impetus was QUBES' 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. The RIOS Institute primarily supports project leaders in STEM education, OER, and related policy/administration.

The mission of the RIOS institute is to support leaders to achieve their sustainability and broader impact goals by working together to amplify the value and reach of open education in STEM and to align our resources and practices with the principles of anti-racism, equity, social justice, and inclusion. In particular, we see open education as an approach and mindset to transform teaching and learning to center the needs of underrepresented and marginalized learners and instructors who have been systematically excluded from the benefits of traditional educational systems.

With QUBESHub as a cyberinfrastructure partner, the RIOS Institute sponsors virtual learning communities, seminars, research working groups, and other collaborative opportunities for helping curate conversations about equity and inclusion and open science education. For example, this past Spring (2021) RIOS led a learning community for network members on open science and open education which integrated discussions of equity and justice. The curriculum for the learning community was also published on QUBESHub so that others can run their own discussion groups within their organizations [33].

RIOS Institute activities are open to all, and we continue to grow in both membership and as an organization. Since its founding, RIOS's focus has shifted from its original founding to simply support sustainable and open biology education projects to more broadly support organizations to move towards social justice orientations in undergraduate STEM education, while emphasizing open science education practices [34] as a key lever for this transformation. All activities are meant to support leaders in their work toward broad organizational change, from sharing information across multiple communities in open education and STEM education and creating peer support communities to sponsoring and facilitating synthesis research and offering relevant professional development.

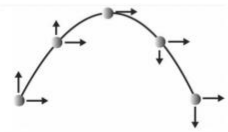
7. CUE4CHNG

Diversity, equity, and inclusion (DEI) is a crucial topic for society and academia. The project Coordinating Curricula and User Preferences to Increase the Participation of Women and Students of Color in Engineering (CUE4CHNG) investigates the hypothesis that a factor in the low representation of women and students of color in STEM results from the lack of accessibility of STEM content and curricula, their format and presentation [35]. The project started 2018 and explores user preferences of students, e.g., for different representational forms, such as equations, images, narratives, simulations, and videos. The characteristics and distinctive typology of STEM curricula, syllabi, course content, and public spaces for STEM content may lead to underrepresentation. Understanding the intersection between learner preferences and such content has the potential to improve engineering education and broaden participation in the field of engineering.

CUE4CHNG applies HUBzero as an open science platform for two main goals: one is to inform about the project, its research, methodology, results, and as a single point of entry to student and teacher/instructors surveys. The second goal is to provide a science gateway that allows the community to upload and share their syllabi, curricula, and teaching material for public access. By now, the project has organized focus groups with 102 students at three universities. The students were surveyed regarding their preferences for presentations of contents. The survey presented five topics in STEM in various ways, including text, equation, two- and three-dimensional illustrations with and without color, animations using the same variations as the illustrations, and an interactive simulation. Figure 5 illustrates different presentations of content in the survey. Because of COVID-19 the project stopped in person focus groups and surveyed over 800 students online. To increase participation, the project used outreach campaigns via targeted emails to student departments, presentations at online conferences and blog posts.

$$x = x_0 + v_{x0}t$$

$$y = y_0 + v_{y0}t - \frac{1}{2}gt^2$$



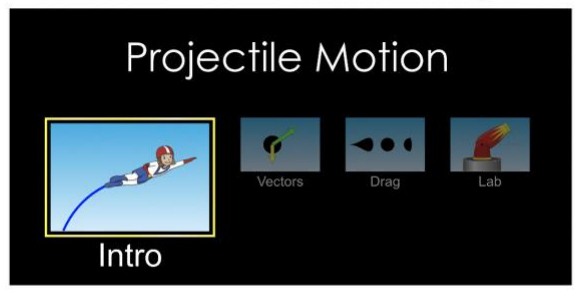


Figure 5. Presentation of a problem in different ways. The problems to be solved include determining how far the projectile flies in the air, subject to gravity, based on the angle and magnitude of the impulse. The presentations here show the formula, a static 2-D visualization, and an interactive simulation [36].

The project has developed two web toolkits: one automated web scrape toolkit to collect syllabi and classes available online and one for extracting pedagogical keywords from the corpus of syllabi and providing various statistical analyses. Data on over 3,000 syllabi and classes have been collected and integrated into the science gateway. The toolkits will be available for researchers that would like to perform their analyses. The plan is to leverage natural language processing algorithms to identify formulations and infer usage of the pedagogic keywords in student preferences.

8. 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" [37]. 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. There was not an excellent way for practitioners or educators to explore experiential tools without knowing what they were looking for or guidance from experienced intercultural learning professionals. The Intercultural Learning Hub (HubICL) [38] 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 into the HubICL platform. The most accessed feature of HubICL is the Toolbox, a searchable collection of experiential tools.

Interculturalists can explore the 700 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
- AAC&U rubric outcomes: Learning outcomes, standardized by the Association of American Colleges and Universities [39]

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. For example, the tool "My Emotional Hot Buttons" is an affirmation introspection. The activity enables participants to examine what behaviors challenge them and manage their reactions [40]. Community members offer advice in the form of reviews, such as encouraging study abroad program leaders to use this activity to preemptively prepare students that will be sharing a room during the program or introducing concepts of intercultural self-awareness, empathy, and practice communication. Other reviews for the "My Emotional Hot Buttons" tool encourage world language classrooms to apply this activity to their curriculum as it lends itself to include statements that pertain to the desired target culture. 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. These badges will incentivize members to grow their skills and encourage further contribution [41].

9. Results and Discussion

"If you build it, they will come" [42] is rarely sufficient to attract a large community for a science gateway. Providers of science gateways need well-planned outreach measures, documentation, a well usable and accessible platform and the trust of users in the technology. The five 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 their community and usage patterns. The unique characteristics of MyGeoHub is the building of "supergroups" tackling the needs of the user community for distinct features for projects with geospatial aspects. QUBEShub focus especially on sharing teaching resources and material to support educators in biology and mathematics and the extensive use of chatrooms and discussion groups is one of the unique features in this science gateway. CUE4CHNG also aims at supporting educators but on college-level and not beginning with K-12 education like QUBEShub and is a research project itself in education. In contrast to the three science gateways before, the project has received only a small seed funding yet and is at an early stage to research which data, material and tools about user preferences should be shared in the science gateway to be beneficial for educators. The domain of HubICL is also education and is a quite young project with start in 2018. Its approach in the science gateway is more comparable to the first three science gateways with providing a well-defined toolbox on topical areas shared in groups. The lessons learned from the different science gateways include that a mixture of outreach measures from publications, presentations to tutorials to newsletters and email campaigns are necessary to attract a wider community and continue to grow and that the science gateway need to follow closely the trends on technologies and concepts such as integrating Jupyter in nanoHUB or FAIR concepts in MyGeoHub. It is not always clear in advance which features and outreach measures attract a wider community. The important aspect is to analyze the impact of activities and to start consistent measures. A Twitter account, for example, without regular tweets might give an inactive impression and the project is better off to have no Twitter account. The table below includes key user numbers for each science gateway.

	Nano Hub	MyGeo Hub	QUBES hub	CUE4 CHNG	Hub ICL
Start	1996	2014	2014	2018	2018
2020	22k users 1.8 Mio visitors	Over 9,500 users	Over 1,000 faculty	Over 800 students	2,937 mem-bers

Table 1: The science gateways, their start data and user numbers in 2020.

10. Conclusion

This paper presents HUBzero and its capabilities in the open science and open education ecosystem by analyzing the features of the five science gateways nanoHUB, MyGeoHub, QUBEShub, CUE4CHNG, and HubICL. The projects tackle different challenges for different communities and use the framework for sharing various data, simulations, and collaborative workflows. The array reflects on the adaptability of HUBzero and its scalability. The presented features are highly beneficial for open science reflected in the continued growth of the communities sharing data and research outcomes. HUBzero will be continuously adapted for further user environments favored by researchers and educators from various communities. The project aims at further improving the sharing and publication processes via feedback from users and developers of hubs. For the near future the HUBzero team envisions to add replication structures and peer-to-peer protocols for situations when servers are temporarily not accessible. This additional feature would make users less dependent on connectivity and enable them to continue working on their data.

10. Acknowledgements

The US National Science Foundation (NSF) has supported nanoHUB development through awards EEC-0228390, EEC-0634750, EEC-1227110, as well as OCI-0438246 and OCI-0721680 (OCI is now OAC). QUBES is currently supported by NSF DUE awards 1446269, 1446284, and 1446258 as well as a number of partner project contracts. NSF has supported MyGeoHub development and projects on MyGeoHub through awards OISE-2020635, OAC-1261727, OAC-1829764, and CBET/ENG-1805741. The United States Department of Agriculture (USDA) and the National Institute of Food and Agriculture (NIFA) have funded projects hosted on MyGeoHub. NSF is supporting CUE4CHNG under the grant agreement 1826632. We would like to acknowledge Sharon Tettegah, Rafael Ferreira da Silva, Yingtao Jiang, David Wallet and Alan Craig for the work described above on CUE4CHNG. The Intercultural Learning Hub (HubICL) is managed by the Center for Intercultural Learning, Mentorship, Assessment, and Research (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 above on HubICL. HUBzero® is a registered trademark of the Regents of the University of California, San Diego.

11. References

[1] Austin CC et al. Research Data Alliance. Fostering global data sharing: highlighting the recommendations of the Research Data Alliance COVID-19 working group. Wellcome Open Res. 2020 Nov 9;5:267. doi: 10.12688/wellcomeopenres.16378.1. PMID: 33501381; PMCID: PMC7808050.

[2] McLennan, M., and Kennell, R. (2010). HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering. Computing in Science & Engineering, vol. 12, no. 2, pp. 48-53, DOI: 10.1109/MCSE.2010.41.

[3] Enis Afgan, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update, Nucleic Acids Research, Volume 46, Issue W1, 2 July 2018, Pages W537–W544, doi:10.1093/nar/gky379

[4] Open Science Framework (OSF). (n.d.). https://osf.io/.

[5] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[6] Gesing, S., et al. Improving the Publication and Re-Use of Data via PresQT Tools and Services. AGU Conference, December 2020.

[7] Venkatesh, V.; Morris, M. G.; Davis, G. B.; Davis, F. D. (2003), "User acceptance of information technology: Toward a unified view" MIS Quarterly, 27 (3): 425–478, doi:10.2307/30036540

[8] Vicente-Saez, Ruben & Martinez-Fuentes, Clara, 2018. "Open Science now: A systematic literature review for an

integrated definition," Journal of Business Research, Elsevier, vol. 88(C), pages 428-436.

[9] https://www.yearofopen.org/open-education-definitions/

[10] https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en

[11] Gesing, S., et al. 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

[12] Wolstencroft K, et al. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. Nucleic Acids Res. 2013 Jul;41(Web Server issue) W557-61. doi:10.1093/nar/gkt328.

[13] Zenodo. (n.d.). https://zenodo.org/.

[14] GitHub. (n.d.). http://github.com/.

[15] GitLab. (n.d.). https://about.gitlab.com/.

[16] Docker. (n.d.). https://www.docker.com/.

[17] Singularity. (n.d.). https://sylabs.io/singularity/.

[18] Klimeck, G., et al. (2008). nanoHUB.org: Advancing Education and Research in Nanotechnology. Computing in Science & Eng., vol. 10, no. 5, pp. 17–23.

[19] Kapadia, N.H., Fortes, J.A.B., Lundstrom, M.S. (1997). The Semiconductor Simulation Hub: A network-based microelectronics simulation laboratory. Proceedings of the Twelfth Biennial Conference: University/Government/Industry Microelectronics Symposium, IEEE Xplore, DOI: 10.1109/UGIM.1997.616686.

[20] Research and Development World. (2020). https://www.rdworldonline.com/rd-100-archive/?YEAR=2020.

[21] Madhavan, K., Zentner, M., and Klimeck, G. (2013). Learning and research in the cloud. Nature Nanotechnology 8, 786–789; DOI:10.1038/nnano.2013.231.

[22] Kalyanam, R., et al. (2018). MyGeoHub - A sustainable and evolving geospatial science gateway, Future Generation Computer Systems, doi.org/10.1016/j.future.2018.02.005.

[23] Zhao, L., et al. (2016). MyGeoHub Science Gateway for Spatial Data and a Model for Sustainability. Gateways 2016, San Diego, CA, November 2-3, 2016.

[24] R. Kalyanam, et al.. GeoEDF: A Plug-and-play Workflow Framework for Earth Science Research. Demonstration track. Gateways 2020 conference, USA, October 19-21, 2020.

[25] Kalyanam, R., et al. Cloud-enabling a Collaborative Research Platform: The GABBs Story. Practice and Experience in Advanced Research Computing (PEARC17), New Orleans, LA, July 9-13, 2017.

[26] Zhao, L., et al. (2020). Building a Gateway Infrastructure for Interactive Cyber Training and Workforce Development. In Practice and Experience in Advanced Research Computing (PEARC '20), July 27–31, 2020. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3311790.3396639.

[27] Courses: FAIR (Findable, Accessible Interoperable and Reusable) Climate and Water Science. MyGeohub. (n.d.). https://mygeohub.org/courses/fair_climate.

[28] Zhao, L. (n.d.). 2021 FAIR Workshop. MyGeohub. https://mygeohub.org/cybertraining/2021workshop.

[29] Donovan, S., et al. (2015). QUBES: a community focused on supporting teaching and learning in quantitative biology. *Letters in Biomathematics* 2, no. 1, 46-55.

[30] Rebova R., Comar T., Diaz Eaton, C. (Ed.) (in press), Special Issue on Mathematics and the Life Sciences: Initiatives, programs, curricula, PRIMUS.

[31] Sierk, M., et al. (2019). NIBLSE Incubators: A community-based model for the development of bioinformatics learning resources. GLBIO2019 Special Session on Bioinformatics Education, (Version 2.0). QUBES Educational Resources. doi:10.25334/Q4QX7T.

[32] Farrell, K. J., et al. (2021). Training macrosystems scientists requires both interpersonal and technical skills. Frontiers in Ecology and the Environment, 19(1), 39-46.

[33] SCORE-UBE Network, Sustainability Challenges for Open Resources to promote and Equitable STEM Education. (n.d.). http://score.qubeshub.org/.

[34] Cangialosi, K. (2021). Open Practices Learning Community Curriculum. Sustainability Challenges for Open Resources to promote an Equitable Undergraduate Biology Education (SCORE-UBE), QUBES Educational Resources. doi:10.25334/G7V1-Y715.

[35] Coordinating Curricula and User Preferences to Increase the Participation of Women and Students of Color in Engineering (CUE4CHNG). (n.d.). https://cue4chng.org/.

[36] https://phet.colorado.edu/sims/html/projectile-motion/latest/projectile-motion_en.html.

[37] Bennett, M.J. (2009). Defining, measuring, and facilitating intercultural learning: A conceptual introduction to the Intercultural Education double supplement. Intercultural Education, 20, S1-S13.

[38] The Intercultural Learning Hub (HubICL). (n.d.). http://hubicl.org/.

[39] Association of American Colleges & Universities (AAC&U). (2009). Intercultural knowledge and competence VALUE rubric. AAC&U.org.

[40] Cherbosque, J., Gardenswartz, L., & Rowe, A. (2006). My emotional hot buttons! In Affirmative Introspection (pp. 35-36). Los Angeles: Emotional Intelligence and Diversity Institute.

[41] Benson, A., et al. (2020). Intercultural Learning Hub (HubICL). Retrieved from osf.io/tc7u3.

[42] Kee, K.F., Le~, B., Jitkajornwanich, K. (2021). 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 STEM community. Concurrency and Computation: Practice and Experience.