2021

A Vision for Science Gateways: Bridging the Gap and Broadening the Outreach

Sandra Gesing
University of Notre Dame

Steven R. Brandt
Louisiana State University

Shannon Bradley
University of Illinois NCSA

Mark Potkewitz
University of Ulster

Kirk F. Kee
Texas Tech University

See next page for additional authors

Authors

Sandra Gesing, Steven R. Brandt, Shannon Bradley, Mark Potkewitz, Kirk F. Kee, Noreen Y. Whysel, Mark J. Perri, Sean B. Cleveland, Annalie Rugg, and Jack A. Smith

A Vision for Science Gateways: Bridging the Gap and Broadening the Outreach

Sandra Gesing
University of Notre Dame
Center for Research Computing
Notre Dame, IN, USA
sandra.gesing@nd.edu

Steven R. Brandt
Louisiana State University
CCT
Baton Rouge, LA, USA
sbrandt@cct.lsu.edu

Shannon Bradley
University of Illinois
NCSA
Urbana, IL, USA
sbrad77@illinois.edu

Mark Potkewitz
Ulster University
Legal Innovation Centre
Belfast, Co.Antrim, UK
m.potkewitz@ulster.ac.uk

Kerk F. Kee
Texas Tech University
College of Media and Communication
Lubbock, TX, USA
kerk.kee@ttu.edu

Noreen Whysel
City University of New York
New York, NY, USA
nwhysel@citytech.cuny.edu

Mark J Perri
Sonoma State University
Rohnert Park, CA, USA
perrim@sonoma.edu

Sean B Cleveland
University of Hawaii - System
Honolulu, HI, USA
seanbc@hawaii.edu

Annelie Rugg
University of California, Los Angeles
Los Angeles, CA, USA
annelierugg@ucla.edu

Jack A. Smith
Center for Environmental,
Geotechnical, and Applied Sciences
Marshall University
Austin, TX, USA
smith1106@marshall.edu

ABSTRACT

The future for science gateways warrants exploration as we consider the possibilities that extend well beyond 'science' and high-performance computing into new interfaces, applications and user communities. In this paper, we look retrospectively at the successes of representative gateways thus far. This serves to highlight existing gaps gateways need to overcome in areas such as accessibility, usability and interoperability, and in the need for broader outreach by drawing insights from technology adoption research. We explore two particularly promising opportunities for gateways - computational social sciences and virtual reality – and make the case for the gateway community to be more intentional in engaging with users to encourage adoption and implementation, especially in the area of educational usage. We conclude with a call for focused attention on legal hurdles in order to realize the full future potential of science gateways. This paper serves as a roadmap for a vision of science gateways in the next ten years.

PEARC '21, July 18–22, 2021, Boston, MA, USA

ACM ISBN 978-1-4503-8292-2/21/07.
https://doi.org/10.1145/3437359.3465579

CCS CONCEPTS

• **Social and professional topics → Software selection and adaptation**; • **Human-centered computing → Accessibility systems and tools**.

KEYWORDS

science gateways, usability, user community, interoperability, vision, impact

ACM Reference Format:
Sandra Gesing, Steven R. Brandt, Shannon Bradley, Mark Potkewitz, Kerk F. Kee, Noreen Whysel, Mark J Perri, Sean B Cleveland, Annelie Rugg, and Jack A. Smith. 2021. A Vision for Science Gateways: Bridging the Gap and Broadening the Outreach. In *Practice and Experience in Advanced Research Computing (PEARC '21), July 18–22, 2021, Boston, MA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3437359.3465579

1 INTRODUCTION

Science Gateways are one of many technologies experiencing exponential growth over the last several decades. They can be defined as an end-to-end solution through streamlined, user-friendly interfaces in support of a community-specific set of tools, applications, and data collections. Pierce et al. [37] identify the starting point of science gateways in the 1990's, notably when TeraGrid [11] leveraged gateways as part of their "wide strategy" to bring high-performance computing (HPC) to a broader set of researchers as new users by lowering the knowledge barrier needed to make use of these resources. By 2011, gateways already accounted for more than 40% of cycles on TeraGrid[1]. Since then, science gateways have been used for far more than just HPC. Examples include data sharing, as in the HUBzero-based [29] science gateway PURR (The Purdue University Research Repository) [38] or as workflow-enabler, as with the Galaxy science gateway [9] that was developed to support the biomedical community and is now used in a variety of disciplines.

With so much technical innovation and computing power, and so many devices and interfaces in the hands of new users, the future of science gateways is worth exploring. How much more can science gateways widen access to the newest computing and data infrastructures or lab instruments? How many more new communities can science gateways support? And what new capabilities are science gateways likely to have in the coming decade? Addressing these questions is not just interesting but also strategically necessary to ensure that we have the time to develop the technologies and interfaces needed to be ready for the kinds of solutions gateway users will want or need in ten years.

Perhaps the lowest common computing platform in the present year is the smartphone. Even in developing nations, more than half the population (53%) have access to such devices [42]. Science Gateways that run on such platforms can have truly global reach and the greatest capability to "widen" the usage of computing and data infrastructure and other science instruments. Possibly, these technologies will embrace more than just the web, providing access through chat services and related technologies.

Another area ripe for "widening" access with science gateways is in the burgeoning area of accessibility to support the computing needs of people with disabilities. In addition to democratizing access to computing and data infrastructures and gateway resources, the effort to reach out to this community has improved the standards for the overall design quality of gateways and other web-based applications. We expect this trend to continue, and to introduce legal hurdles, some of which we will touch upon.

But in addition to "widening" access, the next ten years is likely to see a "deepening" of access. Systems for visualizing scientific data in immersive, virtual reality (VR) environments have (like science gateways themselves) been around since the 1990's [5]. Until recently, VR systems have required substantial investment that put VR out of reach of the average researcher. In the past few years, however, technologies such as "Google Cardboard" and the Oculus Quest[2] have lowered the financial barrier to using these technologies. With more affordable VR, we can anticipate new applications of gateways, such as allowing users to immersively explore data, to explode over the coming decade.

Other ways for gateways to "deepen" access will be to provide access to more instruments, more kinds of instruments, and more applications across more areas of scientific research.

Another ongoing change in science gateways is a maturing of the technology. Standards for authentication, interoperability, etc. are improving and becoming easier to implement.

In the remainder of this paper, we will expand upon the above topics as well as several others.

2 BACKGROUND

The goal of science gateways is to make access to complex research infrastructures more user-friendly and to provide a platform for communities to share computational methods, data and knowledge. Access to HPC and the need to simplify the creation and usage of simulations and computational workflows were two major trends that led to the development of science gateway frameworks. Science gateway frameworks have been developed to support extensibility, scalability and flexibility, and to provide developers with building blocks for delivering end-to-end solutions.

Some of the more common science gateway building blocks include connecting services for batch and cloud systems, data management services, workflow management services and authentication and authorization. The provision of building blocks relieves developers from building the basic features to employ research infrastructures so they can focus on the specific needs of an envisioned science gateway.

Four main categories of science gateway technologies can be distinguished: 1) Complete frameworks, e.g., Galaxy, HUBzero, Open Science Framework [14], Taverna [51]; 2) RESTful APIs and services supporting multiple programming languages, e.g., Apache Airavata [36], TAPIS [43], Agave API [13]; 3) Re-used interface implementations of widely used science gateways such as CIPRES [31]; and 4) Science gateways as a service with the provision of hardware in the background, e.g., SciGap [35].

The lessons learned from the last two decades are that the frameworks that are most successful are those that are sustainable and widely used, technology agnostic, and use APIs and standard web technologies or deliver a complete solution. These factors for success are necessary but not sufficient. MoSGrid (Molecular Simulation Grid) [23] is an example for a successful science gateway that was turned off after eight years of operation. It was running out of funding despite an enthusiastic team and had an increasing need for refactoring and new developments. Thus, the team made the decision to turn the science gateway off and support users to find similar solutions.

On the user community side, physical and life sciences were the main drivers for the creation of science gateways. One of the first science gateways in the 1990's was nanoHUB [20] that has served the nanotechnology community and has been further developed for over 20 years. It was the basis for HUBzero that now services a diverse set of communities with more than 60 different hubs. Galaxy and Taverna also were developed almost 20 years ago, starting off as workflow-enabled science gateways tailored to biomedical applications. Though originally envisioned for specific communities, neither nanoHUB nor Galaxy nor Taverna is bound to a specific user domain; each can be re-used in other domains. Fast forward to today, where science gateways have found their way into new research domains in the social sciences, such as the Social Media Macroscope[3], and into audiences beyond research and teaching via citizen science. Zooniverse[4], for example, facilitates projects that require the active participation of human volunteers to complete research tasks. It has projects with over one million participants in areas from the arts to astronomy to biology to digital humanities.

---

[1] https://www.ideals.illinois.edu/bitstream/handle/2142/43874/TeraGrid_Final_Project_Report.pdf?sequence=2&isAllowed=y)

[2] https://www.oculus.com/quest/

[3] https://socialmediamacroscope.org/

[4] http://zooniverse.org/

Already, citizen scientists from around the globe and a wide range of domains are benefitting from science gateways.

Despite all the success stories and larger uptake of the concept of science gateways and its implementations, there is still much room for improvement in areas such as usability, in widening researcher access to expanded infrastructures and forms of data, and in expanding to more diverse user communities. Allow us to share some of the possibilities with you.

3 BRIDGING THE GAP

The growing availability of computational resources, such as HPC, is one of the motivations for science gateways. Another motivation is to overcome user resistance to exploring new resources — even when their research and teaching will benefit from their use. A big hurdle for many potential HPC users has been the need to use command line interfaces and/or to become acquainted with complex technologies. Science gateways provide access to these resources and digital applications by removing these barriers, making them accessible, usable and interoperable.

3.1 Accessibility

To use a science gateway, it must first be accessible. Unfortunately, a recent study presented at Gateways 2020 [46] revealed that none of the 50 randomly selected Life Sciences gateway sites from the Science Gateways Community Institute (SGCI) [16, 48] catalog[5] were found to meet either W3C[6] or ADA-recommended standards[7].

Accessibility is typically thought of as designing a physical space that is usable for the blind, hearing-impaired and those with mobility issues: accessible ramps, automatic doors, handrails, Braille, voice activated elevators and hearing loop systems, for example. Analogies in digital spaces include variable text size, color contrast, alt tags on images, and optimization for screen readers as well as alternatives to keyboards using voice commands or other input devices.

While we typically think of accessibility in terms of human disability, various user contexts can impair a user's ability to access systems and complete tasks. The sciences present contexts in which access to application controls, data input, extraction and interpretation may be impaired. For example, the scientist may need to focus attention on sensitive equipment or volatile substances while interacting with gateway interfaces. Use of personal protective equipment can make typical interaction with system controls difficult. Field researchers may need alternate interfaces and APIs for platforms including field data input devices, sensors and other machine-accessible formats. Flexible layouts and interoperability (discussed in the next section), can provide additional accessibility support in these situations.

According to the W3C's Web Content Accessibility Guidelines Version 2.1 [6], a site must operate according to the four key principles known as POUR: Perceptible, Operable, Understandable, and Robust.

3.1.1 Perceptible. Perceivability is a principle related to the Nielsen Norman Group's usability heuristics regarding the visibility of system status, match between the system and the real world (mental model), ability to recognize and recover from errors and aesthetic and minimalist design[8]. These heuristics are a refinement of Jakob Nielsen's earlier work in software interfaces[9]. Examples of accessible design include page structure that is easily readable and tab-able by both humans and machines, including a visible point of focus; Information and UI components that are configured and annotated so they are accessible via screen readers for the sight impaired; transcripts of audio content, visual signals such as color, haptic and other non-audio signals for the hearing impaired. Certain cognitive disabilities and tremors can also make the perceptibility of sites (i.e., reading text-based interfaces) difficult.

3.1.2 Operable. The Operable principle is typically associated with disability, but must be considered in various research contexts in which interactive control may be minimized. Are there alternatives to mouse input such as keyboard input or voice control, for example? Does the display avoid flickers or flashes that can trigger seizures? Does the website have at least two ways of finding content such as navigation menu, search feature or a wayfinder site map or index? Are titles, link text and labels on controls meaningful, i.e., is it clear what response they will have if the user interacts with it?

3.1.3 Understandable. Understandability is a key principle of WCAG 2.1 and also Peter Moville's UX Honeycomb [32], which includes related concepts such as Findable, Usable and Accessible. Ensuring that all users can understand the content and interfaces of a site requires a number of areas to be addressed: defining the language of a site or document to support multilingual systems; consistent navigation that avoids automatic changes in mode or context without explanation; labeling in online forms; and accessible error and verification messages with instructions. Findability itself includes information architecture, navigation systems, and search engine optimization. Content that is optimized for findability will be more accessible to human users as well as machines like search engines.

3.1.4 Robust. The Robust principle brings accessibility guidance beyond the individual user to allowing for sustainable and interoperable access to digital systems. Content should be robust enough to be interpreted reliably by a variety of user agents and technologies including assistive devices, search engines and APIs. Compliance with this principle is achieved by using valid HTML[10] and ensuring that any rich media interfaces, such as modal windows, drop-down menus, slideshows, and carousels, include W3C's Accessible Rich Internet Applications (ARIA) markup[11].

3.2 Bringing Usability and Accessibility Into Gateway Design

It may seem that meeting the requirements to support the usability and accessibility of gateways would require having a way to detect

---

[5] https://catalog.sciencegateways.org/#/home

[6] https://www.w3.org/

[7] https://www.ada.gov/

[8] Nielsen, J. (April 24, 1994). 10 Usability Heuristics for User Interface Design. Nielsen Norman Group. https://www.nngroup.com/articles/ten-usability-heuristics/

[9] https://www.interaction-design.org/literature/topics/design-guidelines

[10] http://validator.w3.org/

[11] https://www.w3.org/TR/wai-aria-practices/

every user's impairment and addressing them with specialized tools such as transcripts of all audio content, voice and keyboard controls, haptic sensors and robotic agents. Luckily, that is not the case. Careful attention to proper HTML structure and other information architecture principles will ensure that gateways conform to WCAG guidlines. Creating accessibility guidelines for gateway systems, such as those provided at the University of Washington[12] will aid in bringing WCAG principles to the forefront of gateway design. Projects like the W3C's Web Accessibility Initiative (WAI)[13] and The Accessibility Project[14] offer frameworks and checkers to test compliance as gateways and new gateway features are released. We envision future gateways will recognize the significance of accessibility and usability in supporting users to expand adoption and equity for researchers.

3.3 Interoperability

As funding agencies look to gain more from their investments, gateways will offer demonstrable returns on investment if interoperability is kept at the forefront of gateway development to ensure re-use and integration. The creation and adoption of standards will be crucial to ensuring future gateway interoperability across frameworks, APIs and data sets. Already, efforts among several gateway frameworks HUBzero [30], Tapis [43], Open On-Demand [18], Airvata [27] and XSEDE [45] to create "resource" standards is allowing for easier adoption, migration and integration by reducing the management load on gateway administrators able to re-use standardized computing and storage resources. In addition, the potential for centralized repositories that can be maintained by resource administrators–keeping them current and relevant–is enticing.

3.3.1 APIs. Going forward, APIs developed according to web standards with tools such as OpenAPI and gRPC will advance gateway interoperability by enabling code generation, validation and documentation. These open APIs will enable gateways to interoperate with each other as standardized "services", allowing specialized gateways to focus on their domains or use cases and not have to recreate the foundational infrastructures. The emergence of these meta-gateways will deliver new interfaces and workflows that span multiple gateways.

3.3.2 Data Management. More gateways are enabling the management and dissemination of data products by supporting FAIR guiding principles[49] (Findability, Accessibility, Interoperability, Re-Use) in the movement towards Open Science[41]. Interoperability is the third tenet in FAIR and relates to the adoption of ontologies for creating and restricting metadata generation. This will also be crucial to the future of science gateways. Emerging tools such as the Preservation Quality Tool (PresQT[15]) and DataCite[15] will provide for improved reuse of preserved data and software in library repository systems such as Zenodo[16] and Figshare[17], and make research data more discoverable, as well as more interoperable across science gateways and research cyberinfrastructures.

3.3.3 Identity and Access Management. In order to enable much of this interoperability, trust and multi-institutional access must exist. The development and adoption of federated Identity and Access Management (IAM) services such as CI-Login, InCommon and SciTokens[50] through emerging frameworks, such as Custos[18], will solve the issue of trust across institutions and gateways. This will address complex challenges users currently face, such as with issues of multi-factor authentication in long-running or multi-resource workflows. With increased interoperability a new connected ecosystem of advanced science gateways cyberinfrastructure can arise to accelerate scientific discovery, reproducibility and integrity.

4 BROADENING THE OUTREACH

Science gateways have traditionally facilitated researchers in connecting them to HPC resources, in a "friendly" discipline-specific way, especially for "casual" users, typically by providing a web-based interface. HPC usually meant supercomputers, like those available through XSEDE, but it could also mean High-Throughput (HT) resources like Open Science Grid. HPC, or supercomputing, has often been defined as anything that exceeds the capacity of your "workstation", generally referring to:

- Compute (CPU) speed or number of shared-memory CPUs
- Memory size
- Storage capacity or access speed
- Network latency or bandwidth

Supercomputing typically involves large parallel computation across multiple high-speed CPUs with shared and distributed memory and shared distributed storage, both optimized for low-latency and high-bandwidth data communications. They generally do not cater to a specific discipline or research environment. Bridging that gap has been the primary role of science gateways. However, science gateways can also help alleviate other limitations - for example, the researchers' knowledge of and experience with the non-discipline-specific aspects of an advanced or otherwise desirable but unfamiliar resource, occasionally needed for both computationally and/or data-intensive tasks. Other barriers might include:

- Platform compatibility, where containerization might help,
- Access to data, where frameworks based on FAIR principles might help,
- Access to human expertise, where trained facilitators might help,
- Access to novel hardware, such as
  - Accelerators - e.g., GPUs, FPGAs, ASICs
  - AI processors - e.g., NPUs, TPUs, APUs
  - Quantum computers
  - Edge computing for IoT (Internet of Things)
  - Remote Instruments
    - Robots, for hazardous environments

---

[12] https://www.washington.edu/accessibility/checklist/
[13] https://www.w3.org/WAI/test-evaluate/
[14] http://a11yproject.org
[15] https://datacite.org
[16] https://zenodo.org
[17] https://figshare.org
[18] https://airavata.apache.org/custos/

  - Satellites, for Remote Sensing
  - Microscopes - e.g., Optical, Electron, Scanning Probe
  - Telescopes - e.g., Event Horizon Telescope
  - Medical Imaging - e.g., Xray, CT, MRI, PET

To further "widen" science gateways' impacts, it would be helpful for the community to draw from social science insights on technology adoption. One such area of study is the diffusion or adoption and spread of innovations [40]. Rogers documented that innovations that attract users and speed uptakes tend to possess five main attributes: relative advantage benefits over competing options, perceived compatibility (alignments with users' needs, situations, values, etc.), simplicity (low learning curves), trialability (opportunities to experiment with the technology before full adoption), and observability (increased visibility of the innovations in ecosystem). Rice [39] explains that potential adopters fall under 5 groups with distinct psycho-social profiles: innovators (2.5% of population) are venturesome, early adopters (13.5% of population) are visionary, early adopters (34% of population) are pragmatic, late majority (34% of population) are cautious, and laggards (16% of population) are suspicious. We maintain that how to design gateways with the 5 attributes and introduce gateways to the 5 adopter groups for the next 10 years require some thoughtful efforts guided by social science insights. Being thoughtful of such research insights will help developers and ambassadors be more strategic in promoting gateways to a broader audience.

4.1 Growth of the Diversity of Domains

While domains such as biology and geological sciences are already well served via science gateways, we expect the uptake of science gateways will further grow in these disciplines. We also assume that quite a few domains that still have much room for improvement in the uptake of science gateways, will use science gateways by large. Such areas include social sciences, arts, digital humanities, data science, business, law, medicine, mathematics and natural language. As examples of the potential that gateways offer to these communities, we share some insights into the areas of computational social sciences and virtual reality.

4.1.1 Computational Social Sciences. The advent of social media sparked the explosion of computational social science. In one particular area, social scientists are now harnessing large scale social media data as digital breadcrumbs [25] to study human social behaviors. Among recent studies, content analysis and network analysis are two commonly employed techniques [28]. However, limited attention has been given to the potential development of gateways for computational social sciences.

Researchers in this area often rely on fee-based software programs (such as LIWC for linguistic/content analysis; UCINet for network analysis) due to their robust analytical techniques, but these programs do not scale to big data effectively. Newer and some open source programs exist, but their analysis techniques/options are limited and superficial and sometimes suffer in being not robust enough to generate findings competitive for journal articles.

Furthermore, reproducibility is important in research, but social science data inherently have privacy and ethical concerns. Gateways may have the ability to protect data privacy without compromising the need for reproducibility. We propose computational social science as an opportunity for gateway developments in the next 10 years, as both the data and techniques are mature and ready.

4.1.2 Virtual Reality. Traditionally, both science and science gateways have required that one view the world through a screen. The emerging virtual reality technologies (virtual and augmented reality) make it increasingly possible to step through that screen and physically enter the world of data.

The concept of using virtual reality for science gateways has been around since at least 2009 [10]. At present, there are a handful of such projects that have been implemented [24, 47]. We are not alone in identifying the importance of virtual reality. Enhancing virtual reality was identified by the National Academy of Engineering as one of the Fourteen Grand Engineering Challenges of the 21st century.

Science Gateways using virtual reality would build upon experiences using the CAVE first developed in 1992 [12]. Not only were these prototypes much more expensive than comparable systems today, but they were far less capable (these early systems were limited to about 30MB of data). A more modern review of the CAVE technology may be found here [33].

Perhaps the easiest way for an application or gateway to get started in offering virtual reality content is by use of a standard markup language called X3D [7] or data formatted for Paraview, Immersive [2]. But virtual reality has more possibilities for gateways than simply visualizing datasets. In principle, it should be possible to create simulations, collaborate, to orchestrate and monitor scientific workflows inside the VR environment (e.g. Sua et al. [44]). The interactive capabilities of VR systems can be valuable. Kreylos et al. [22] simulate a geological environment so that geoscientists can do a virtual kind of field work without leaving their office.

Though some steps have been made, by and large, the science gateways community has yet to make the jump and make full use of modern virtual reality technologies. Although much of what was discussed in Section 4 is about broadening outreach to a diverse range of disciplines and domains, we note that similar efforts should be made to broaden the outreach to underrepresented groups and diverse institutions, following the guidance of the special report "NSF Includes"[19].

5 FUTURE USER COMMUNITIES AND THE FUTURE OF USER COMMUNITIES

While the vision of an increased variety of user communities and a much larger user community is exciting, we keep in mind that the gateway community needs to proactively do outreach in order to get the word out. It is a common pitfall that developers hold the mindset that 'if you build it, they will come'. This fallacy has recently been revised for gateway adoption to read 'if you build it, promote it, and they trust you, then they will come' [19] and is addressed in SGCI's Focus Week [17]. To expand beyond the early adopters and those already sold on computational resources, the gateway community needs to prioritize outreach, marketing, advertising, public relations, trust cultivation and relationship building with new communities of users. Such an effort will also open up opportunities for user feedback, so developers can build and refine gateways to meet

---

[19] https://www.nsf.gov/news/special_reports/nsfincludes/NSFIncludes_archivedSpecialReport.pdf

the needs of users (both researchers and beyond). With the right outreach and a sensitivity to addressing different users differently, science gateways are likely to see much wider uptake in the next 10 years in settings such as K-12 and higher ed, life-long learning, citizen science, and even outside research and education, e.g., for recreational applications.

5.1 Educational Settings

Science Gateways are perfect for use in an educational setting because they remove the barriers to computing and data infrastructure beyond one's own workstation. Gateways have been an important source of remote instruction during the COVID-19 pandemic. To increase participation in an educational setting, gateways need to be designed for all students and marketed to all faculty. Not all faculty are familiar with computational work as many are experimentalists. For example, an experimental Physical Chemist may have little or no experience with computational chemistry, but may be the only faculty teaching the quantum section of Physical Chemistry in a small department. These faculty need easy access to computational resources as educational tools. While they are well versed with the theory of their subject, they might lack the practical expertise required to select the correct computational packages, parameters, and settings required to produce reasonably useful data. In order to broaden participation, gateways should automate selection of parameters based on the input system and provide working examples with instructions. Educational gateways particularly benefit from rapid support and a guided experience due to the inexperience of the users. Over the next ten years AI could play a role in helping select reasonable parameters based on a system under study, using past job submissions as a training set.

Educational gateways benefit students directly by reinforcing class material. Indirect benefits may be even more important though: increased confidence with computers and increased awareness of computer-based jobs (e.g. data-science) helps to prepare students for STEM-based careers, where data analysis and visualization are required. Jupyter Notebooks [21] are commonly used for data science. Over the next ten years notebooks will likely continue to increase in popularity due to their ease of access and versatility. Free online services such as Binder and Google Colab serve to broaden access, but many computational notebooks require custom libraries or binaries that are difficult to install on such services and dedicated CPU / GPU to run interesting calculations. Such custom installs can be deployed locally at an institution with effort through, e.g. Zero to JupyterHub [8]. However, container services like Kubernetes are not easy to deploy and access to compute resources are required for the deployment. XSEDE resources now include direct Jupyter access, but security concerns and readily available cryptocurrency mining software makes it problematic to open access to one's entire class. Over the next ten years gateways could be tasked to support custom Jupyter installations for science education without requiring a deep background in containerization or IT security. One extreme of security is voilà[20] an alternate Jupyter notebook server that doesn't allow users to alter the notebook code or run arbitrary commands. Voilà can be used to create UI elements that harness the power of Julia, Python, or R from a Jupyter notebook. These can be used to create simulations [34] or for data analysis on a gateway, allowing designers to program in languages that are comfortable to them and effective in data analysis, rather than in Javascript.

5.2 Science Gateways for Everyone

One example for science gateways for everyone is the opportunity to travel in VR. The lessons learned from COVID-19 have drawn attention to the need for better ways for people to enjoy tourist sites virtually. This use case would allow people who are unable to physically travel to travel virtually and experience places immersively. Even visiting on site could be extended with augmented reality provided by means of gateways. Over the next ten years, imagine utilizing gateways to take large amounts of different types of data (i.e. photos, scans, architectural drawings, Infrared, lidar. video, audio recordings, etc.) and making it into a system able to create a multi-media site visualization which can show the space as it is seen by the naked eye and allowing drill down into spaces not normally seen even on an in-person tour. The Real-time Immersive Virtual Environments for Education & Learning (RIVEEL 3D) is a study of Mediterranean Medieval Graffiti by Mia Trentin and Colter Wehmeier. They have taken the gateway, Clowder, [26] and are working on an application that will ingest diverse sets of media and organize it into a spatialized, searchable archive. This can then be presented via a curated interface which allows a user to dig deeper into a historical site view and see more about the history and graffiti there[21]. Fig. 1 shows on the left side pictures of graffitis. On the right side is information for the context and location such as primary documentation, e.g. a map, a laser scan, a photogrammetry scan and a panoramic photo. They can be combined to creating a VR component for the related graffiti. Another example is research from the Cyprus Institute Virtual Environments Lab which is working on a virtual environment to allow urban planners to work with local stakeholders, international experts, authorities and inhabitants of the medieval city of Nicosia on developing the Paphos Gate. Looking forward, virtual walk-throughs like this will allow each user to sketch their own preferred path through the site (or other sites). With this information future site planners will be able to steer the construction of the tour site in such a way that all citizens can "grow links with a place which will contribute to a feeling of belonging".

6 OUTLOOK

There exists a wide range of legal issues that touch and concern any endeavor involving the Internet and thus science gateways. Bridging the gap and broadening the outreach for science gateways will be important for accelerating science and education and exciting for areas beyond these areas as described above. Rules and laws will also have to keep the pace with the developments. A number of federal laws regulate privacy around health data, educational records, children's privacy, banking and financial information, and, when coupled with various state laws, they create a series of byzantine legal obligations and regulatory challenges.

---

[20] https://blog.jupyter.org/and-voil%C3%A0-f6a2c08a4a93

[21] RIVEEL 3D is a long standing research activity developed in the context of the CyI's collaboration with the NCSA and the University of Illinois at Urbana-Champaign, with the support of the Cyprus Department of Antiquities and the Municipality of Nicosia, 2016-21 [1, 3, 4]. Recently, project results were disseminated and reused by the European Digital Research Infrastructure for the Arts and Humanities (DARIAH)

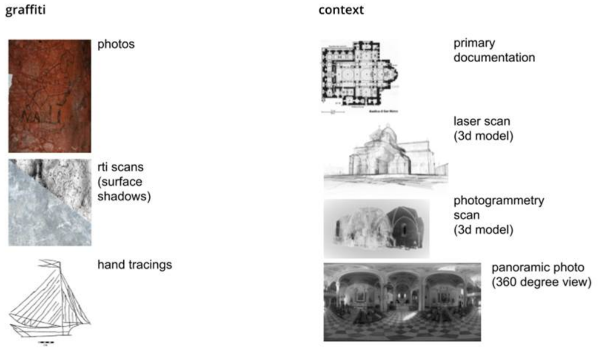

Figure 1: Example for Mediterranean Medieval Graffiti and their context.

Machine Learning technologies create new legal considerations around areas like privacy and intellectual property. Laws like the Americans with Disabilities Act place obligations on website operators to ensure ease of access. Further, the international nature of the Internet requires consideration of laws and rules overseas, such as Europe's General Data Protection Regulation as well as varying approaches to free speech, safe harbour provisions, and defamation. In addition, varying jurisdictions require different de minimus cybersecurity standards and data breach notification horizons. The Brussels Effect refers to the impact of EU-wide regulations on the global regulatory marketplace. Passage of the GDPR saw many countries outside the European Economic Area pass GDPR-style privacy rules, and many firms adopted global policies aimed at unified compliance strategy. Over the next 10 years, countries will likely continue to pass legislation aimed at giving individual computer users/consumers more control over their personal data and create regulations forcing greater transparency and accountability in algorithmic decision-making. As autonomous systems gain wider adoption in areas such as transportation and medicine, novel legal considerations regarding liability and causation will emerge when systems err or fail and cause harm to people and property. Disharmony between national approaches to fundamental rights will continue a fractured approach to liability regimes with the United States taking a caveat emptor approach and the European Union-style states placing stronger obligations on manufacturers and technology firms. Harmonization of legal and ethical rules and increased transparency through independent audit will allow for greater international collaboration between researchers. More unified compliance regimes will reduce administrative overhead and encourage a wider range of applications of computing technology across research areas. Transparency and familiarization results in greater trust and public confidence in research and technology.

Going forward, if we are mindful of these important legal considerations, foster a standards-based culture and a commitment to accessibility, and reach out to new communities, the future of gateways is expansive and impactful. In truth, it is a matter of the gateway communities staying true to their original focus on simplifying interfaces for high-end computation with the added step of thinking intentionally about WHO we are simplifying for: users of all abilities, from within and outside the sciences, and reaching beyond the academy to our engaged citizenry.

ACKNOWLEDGMENTS

We would like to acknowledge NSF OAC 1547611 (SGCI) and NSF OAC #1939067 & #1453864. Clowder has been funded by grants: NSF #BCS-0941268, #EAR- 331906, #ACI-1261582, #ACI-1443013, #OCI-0940824, #OCI-0525308; NIH #1P01AI089556-01A1. Illinois - Indiana Sea Grant #DW92329201, European Commission #RI-261600, XSEDE #OCI-1053575, ARPA-E #DE-AR0000594, Collaborative Research: CSSI: Framework: Data: Clowder Open Source Customizable Research Data Management, Plus-Plus, 2018-2023 (#OAC-1835834, 1835543, 1835877),the Tapis Framework:(#OAC-1931439 and #OAC-1931575)

REFERENCES

[1] 2019. *ICAT EVGE.* https://youtu.be/RUzNFDPV1_o

[2] 2021. *Paraview Immersive.* https://www.paraview.org/immersive/

[3] 2021. *Research from Cyl's Virtual Environments Lab Selected Among the Best International Practices of Public Space Co-Creation.* https://www.cyi.ac.cy/index.php/in-focus/cyi%E2%80%99s-virtual-environments-lab-research-selected-among-the-best-international-practices-of-public-space-co-creation.html

[4] 2021. *RIVEEL 3D.* http://avl.ncsa.illinois.edu/realworld-software/riveel-3d

[5] 2021. *Visbox - The CAVE Environment.* Retrieved March 8, 2021 from http://www.visbox.com/prajlich/T/node7.html

[6] 2021. *WCAG 2.1.* https://www.w3.org/TR/2018/REC-WCAG21-20180605/

[7] 2021. *X3D.* https://en.wikipedia.org/wiki/X3D

[8] 2021. Zero To Jupyterhub. https://zero-to-jupyterhub.readthedocs.io/en/latest/index.html

[9] Enis Afgan, Dannon Baker, Bérénice Batut, Marius van den Beek, Dave Bouvier, Martin Cech, John Chilton, Dave Clements, Nate Coraor, Björn A. Grüning, Aysam Guerler, Jennifer Hillman-Jackson, Saskia D. Hiltemann, Vahid Jalili, Helena Rasche, Nicola Soranzo, Jeremy Goecks, James Taylor, Anton Nekrutenko, and Daniel J. Blankenberg. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, Webserver-Issue (2018), W537–W544. https://doi.org/10.1093/nar/gky379

[10] Robert N Allan. 2009. *Virtual research environments: From portals to science gateways.* Elsevier.

[11] Charles E. Catlett. 2005. TeraGrid: A Foundation for US Cyberinfrastructure. In *Network and Parallel Computing*, Hai Jin, Daniel Reed, and Wenbin Jiang (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–1.

[12] Carolina Cruz-Neira, Jason Leigh, Michael Papka, Craig Barnes, Steven M Cohen, Sumit Das, Roger Engelmann, Randy Hudson, Trina Roy, Lewis Siegel, et al. 1993. Scientists in wonderland: A report on visualization applications in the CAVE virtual reality environment. In *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium.* IEEE, 59–66.

[13] Rion Dooley, Steven R. Brandt, and John Fonner. 2018. The Agave Platform: An Open, Science-as-a-Service Platform for Digital Science. In *Proceedings of the Practice and Experience on Advanced Research Computing* (Pittsburgh, PA, USA) *(PEARC '18)*. Association for Computing Machinery, New York, NY, USA, Article 28, 8 pages. https://doi.org/10.1145/3219104.3219129

[14] Erin Foster and Ariel Deardorff. 2017. Open Science Framework (OSF). *Journal of the Medical Library Association* 105, 2 (2017), 203–206. https://doi.org/10.5195/jmla.2017.88

[15] S. Gesing, N. Meyers, R. Johnson, and J Wang. 2020. Improving the Publication and Re-Use of Data via PresQT Tools and Services.. In *Proceedings of the AGU Conference December 2020.* https://agu.confex.com/agu/fm20/webprogram/Paper775209.htm

[16] Sandra Gesing, Nancy Wilkins-Diehr, Maytal Dahan, Katherine A. Lawrence, Michael G. Zentner, Marlon E. Pierce, Linda Hayden, and Suresh Marru. 2017. Science Gateways: The Long Road to the Birth of an Institute. In *50th Hawaii International Conference on System Sciences, HICSS 2017, Hilton Waikoloa Village, Hawaii, USA, January 4-7, 2017*, Tung Bui (Ed.). ScholarSpace / AIS Electronic Library (AISeL), 1–10. http://hdl.handle.net/10125/41919

[17] S. Gesing, M. Zentner, J. Casavan, B. Hillery, M. Vorvoreanu, R. Heiland, S. Marru, M. Pierce, N. Mullinix, and N. Maron. 2017. Science Gateways Incubator: Software Sustainability Meets Community Needs. In *2017 IEEE 13th International Conference on e-Science (e-Science).* 477–485. https://doi.org/10.1109/eScience.2017.77

[18] Dave Hudak, Doug Johnson, Alan Chalker, Jeremy Nicklas, Eric Franz, Trey Dockendorf, and Brian L. McMichael. 2018. Open OnDemand: A web-based client portal for HPC centers. *Journal of Open Source Software* 3, 25 (2018), 622.

https://doi.org/10.21105/joss.00622

[19] Kerk F. Kee, Bethanie Le, and Kulsawasd Jitkajornwanich. [n.d.]. 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 and Computation: Practice and Experience* n/a, n/a ([n. d.]), e6192. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.6192

[20] G. Klimeck, M. McLennan, S. P. Brophy, G. B. Adams III, and M. S. Lundstrom. 2008. nanoHUB.org: Advancing Education and Research in Nanotechnology. *Computing in Science Engineering* 10, 5 (2008), 17–23. https://doi.org/10.1109/MCSE.2008.120

[21] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, and Sylvain Corlay. [n.d.]. Jupyter Notebooks-a publishing format for reproducible computational workflows. In *ELPUB*. 87–90.

[22] Oliver Kreylos, Gerald W Bawden, Tony Bernardin, Magali I Billen, Eric Cowgill, R D Gold, Bernd Hamann, Margarete Jadamec, Louise H Kellogg, Oliver Dr Staadt, and Dawn Y Sumner. 2006. Enabling scientific workflows in virtual reality. In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications*, Stephen N Spencer (Ed.). Association for Computing Machinery (ACM), United States of America, 155 – 162. International Conference on Virtual Reality Continuum and its Applications ; Conference date: 01-01-2006.

[23] Jens Krüger, Richard Grunzke, Sandra Gesing, Sebastian Breuers, Andre Brinkmann, Luis de la Garza, Oliver Kohlbacher, Martin Kruse, Wolfgang E. Nagel, Lars Packschies, Ralph Müller-Pfefferkorn, Patrick Schäfer, Charlotta Schärfe, Thomas Steinke, Tobias Schlemmer, Klaus Dieter Warzecha, Andreas Zink, and Sonja Herres-Pawlis. 2014. The MoSGrid Science Gateway - A Complete Solution for Molecular Simulations. *Journal of Chemical Theory and Computation* 10, 6 (2014), 2232–2245.

[24] Antonio Laganà and Osvaldo Gervasi. 2010. A priori molecular virtual reality on EGEE grid. *International Journal of Quantum Chemistry* 110, 2 (2010), 446–453.

[25] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. Computational Social Science. *Science* 323, 5915 (2009), 721–723. https://doi.org/10.1126/science.1167742 arXiv:https://science.sciencemag.org/content/323/5915/721.full.pdf

[26] Luigi Marini, Indira Gutierrez-Polo, Rob Kooper, Sandeep Puthanveetil Satheesan, Maxwell Burnette, Jong Lee, Todd Nicholson, Yan Zhao, and Kenton McHenry. 2018. Clowder: Open Source Data Management for Long Tail Data. In *Proceedings of the Practice and Experience on Advanced Research Computing* (Pittsburgh, PA, USA) *(PEARC '18)*. Association for Computing Machinery, New York, NY, USA, Article 40, 8 pages.

[27] Suresh Marru, Lahiru Gunathilake, Chathura Herath, Patanachai Tangchaisin, Marlon Pierce, Chris Mattmann, Raminder Singh, Thilina Gunarathne, Eran Chinthaka, Ross Gardler, Aleksander Slominski, Ate Douma, Srinath Perera, and Sanjiva Weerawarana. 2011. Apache Airavata: A Framework for Distributed Applications and Computational Workflows. In *Proceedings of the 2011 ACM Workshop on Gateway Computing Environments* (Seattle, Washington, USA) *(GCE '11)*. Association for Computing Machinery, New York, NY, USA, 21–28.

[28] Sorin A. Matei and Kerk F. Kee. 2019. Computational communication research. *WIREs Data Mining and Knowledge Discovery* 9, 4 (2019), e1304. https://doi.org/10.1002/widm.1304 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1304

[29] M. McLennan and R. Kennell. 2010. HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering. *Computing in Science and Engg.* 12, 2 (March 2010), 48–53.

[30] M. McLennan and R. Kennell. 2010. HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering. *Computing in Science Engineering* 12, 2 (2010), 48–53. https://doi.org/10.1109/MCSE.2010.41

[31] M. A. Miller, W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*. 1–8. https://doi.org/10.1109/GCE.2010.5676129

[32] Peter Morville and Paula Sullenger. 2010. Ambient Findability: Libraries, Serials, and the Internet of Things. *The Serials Librarian* 58, 1-4 (2010), 33–38. https://doi.org/10.1080/03615261003622999 arXiv:https://doi.org/10.1080/03615261003622999

[33] Muhanna A Muhanna. 2015. Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions. *Journal of King Saud University-Computer and Information Sciences* 27, 3 (2015), 344–361.

[34] M. J. Perri. 2020. Online Data Generation in Quantitative Analysis: Excel Spreadsheets and an Online HPLC Simulator Using a Jupyter Notebook on the Chem Compute Web site. *Journal of Chemical Education* 97, 9 (2020), 2950–2954. https://doi.org/10.1021/acs.jchemed.0c00565

[35] Marlon Pierce, Suresh Marru, Eroma Abeysinghe, Sudhakar Pamidighantam, Marcus Christie, and Dimuthu Wannipurage. 2018. Supporting Science Gateways Using Apache Airavata and SciGaP Services *(PEARC '18)*. Association for Computing Machinery, New York, NY, USA, Article 99, 4 pages. https://doi.org/10.1145/3219104.3229240

[36] M. Pierce, S. Marru, L. Gunathilake, T. A. Kanewala, R. Singh, S. Wijeratne, C. Wimalasena, C. Herath, E. Chinthaka, C. Mattmann, A. Slominski, and P. Tangchaisin. 2014. Apache Airavata: Design and Directions of a Science Gateway Framework. In *2014 6th International Workshop on Science Gateways*. 48–54. https://doi.org/10.1109/IWSG.2014.15

[37] M. Pierce, M. Miller, Emre H. Brookes, M. Wong, Y. Liu, E. Afgan, Sandra Gesing, M. Dahan, S. Marru, and T. Walker. 2018. Towards a Science Gateway Reference Architecture. In *Proc. of IWSG 2018 (10th International Workshop on Science Gateways), 13-15 June 2018, Edinburgh, Scotland*. http://ceur-ws.org/Vol-2357/paper6.pdf

[38] PURR. 2021. — PURR. https://purr.purdue.edu/wiki/?version= [Online; accessed 9-3-2021].

[39] Rice, R. E. 2009. *Diffusion of innovations: Theoretical extensions.* (In R. Nabi and M. B. Oliver (Eds.), Handbook of media effects ed.). Thousand Oaks, CA: Sage. 489–503 pages.

[40] Everett M. Rogers. 2003. *Diffusion of innovations* (5th ed.). Free Press, New York, NY [u.a.]. 576 pages.

[41] Gianmaria Silvello. 2018. Theory and practice of data citation. *Journal of the Association for Information Science and Technology* 69, 1 (2018), 6–20. https://doi.org/10.1002/asi.23917 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23917

[42] Laura Silver, Aaron Smith, Courtney Johnson, Kyle Taylor, Jingjing Jiang, Monica Anderson, and Lee Rainie. 2019. Mobile connectivity in emerging economies. *Pew Research Center* 7 (2019).

[43] Joe Stubbs, Richard Cardone, Mike Packard, Anagha Jamthe, Smruti Padhy, Steve Terry, Julia Looney, Joseph Meiring, Steve Black, Maytal Dahan, Sean Cleveland, and Gwen Jacobs. 2021. Tapis: An API Platform for Reproducible, Distributed Computational Research.. In *Proceedings of the Future of Information and Communication Conference (FICC) 2021.* https://agu.confex.com/agu/fm20/webprogram/Paper775209.htm

[44] S. Sua, A. Chaudhary, P. O'Leary, B. Geveci, W. Sherman, H. Nieto, and L. Francisco-Revilla. 2015. Virtual reality enabled scientific visualization workflow. In *2015 IEEE 1st Workshop on Everyday Virtual Reality (WEVR)*. 29–32. https://doi.org/10.1109/WEVR.2015.7151692

[45] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr. 2014. XSEDE: Accelerating Scientific Discovery. *Computing in Science Engineering* 16, 5 (2014), 62–74. https://doi.org/10.1109/MCSE.2014.80

[46] N. Whysel, S. Thurow, and B. Corwin. [n.d.]. Accessibility Compliance and Assessments for Gateway Websites in Life Sciences: Toward Inclusive Design.

[47] Nancy Wilkins-Diehr, Chaitan Baru, Dennis Gannon, Kate Keahey, John McGee, Marlon Pierce, Rich Wolski, and Wenjun Wu. 2010. Science gateways: Harnessing clouds and software services for science. *Cloud Computing and Software Services* (2010), 17.

[48] Nancy Wilkins-Diehr, Michael Zentner, Marlon Pierce, Maytal Dahan, Katherine Lawrence, Linda Hayden, and Nayiri Mullinix. 2018. The science gateways community institute at two years. In *Proceedings of the Practice and Experience on Advanced Research Computing*. 1–8.

[49] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1 (March 2016), 160018. https://doi.org/10.1038/sdata.2016.18

[50] Alex Withers, Brian Bockelman, Derek Weitzel, Duncan Brown, Jeff Gaynor, Jim Basney, Todd Tannenbaum, and Zach Miller. 2018. SciTokens: Capability-Based Secure Access to Remote Scientific Data. In *Proceedings of the Practice and Experience on Advanced Research Computing* (Pittsburgh, PA, USA) *(PEARC '18)*. Association for Computing Machinery, New York, NY, USA, Article 24, 8 pages. https://doi.org/10.1145/3219104.3219135

[51] Katherine Wolstencroft, Robert Haines, Donal Fellows, Alan Williams, David Withers, Stuart Owen, Stian Soiland-Reyes, Ian Dunlop, Aleksandra Nenadic, Paul Fisher, Jiten Bhagat, Khalid Belhajjame, Finn Bacall, Alex Hardisty, Abraham Nieva de la Hidalga, Maria P. Balcazar Vargas, Shoaib Sufi, and Carole Goble. 2013. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic Acids Research* 41, W1 (05 2013), W557–W561. https://doi.org/10.1093/nar/gkt328 arXiv:https://academic.oup.com/nar/article-pdf/41/W1/W557/3822473/gkt328.pdf