The Global Impact of Science Gateways, Virtual Research Environments and Virtual Laboratories

Michelle Barker[a,*], Silvia Delgado Olabarriaga[b], Nancy Wilkins-Diehr[c], Sandra Gesing[d], Daniel S. Katz[e], Shayan Shahand[f], Scott Henwood[g], Tristan Glatard[h], Keith Jeffery[i], Brian Corrie[j], Andrew Treloar[k], Helen Glaves[l], Lesley Wyborn[m], Neil Chue Hong[n], Alessandro Costa[o]

[a] *National eResearch Collaboration Tools and Resources, James Cook University, Australia*
[b] *Academic Medical Center of the University of Amsterdam, The Netherlands*
[c] *San Diego Supercomputer Center, University of California San Diego, USA*
[d] *Department of Computer Science and Engineering/Center for Research Computing, University of Notre Dame, USA*
[e] *National Center for Supercomputing Applications, Department of Computer Science, Department of Electrical and Computer Engineering, School of Information Sciences, University of Illinois at Urbana-Champaign, USA*
[f] *ServiceNow and Academic Medical Center of the University of Amsterdam, The Netherlands*
[g] *CANARIE, Canada*
[h] *Department of Computer Science and Software Engineering, Concordia University, Canada*
[i] *VRE4EIC and ERCIM, UK*
[j] *Simon Fraser University, Canada and New Zealand eScience Infrastructure, New Zealand*
[k] *Australian National Data Service, Australia*
[l] *British Geological Survey, UK*
[m] *National Computational Infrastructure Facility and The Research School of Earth Sciences, Australian National University, Australia*
[n] *University of Edinburgh, UK*
[o] *INAF, Italy*

Abstract

Science gateways, virtual laboratories and virtual research environments are all terms used to refer to community-developed digital environments that are designed to meet a set of needs for a research community. Specifically, they refer to integrated access to research community resources including software, data, collaboration tools, workflows, instrumentation and high-performance computing, usually via Web and mobile applications. Science gateways, virtual laboratories and virtual research environments are enabling significant contributions to many research domains, facilitating more efficient, open, reproducible research in bold

*Corresponding author
Email address: michelle.barker@nectar.org.au (Michelle Barker)

new ways. This paper explores the global effect of these programs in increasing research impact, demonstrates their value in the broader digital landscape and discusses future opportunities.

Keywords: science gateways, virtual research environments, virtual laboratories, open science, e-infrastructure, cyberinfrastructure

1. Introduction

Science gateways, virtual laboratories and virtual research environments (hereafter science gateways) refer to various kinds of community-developed digital interfaces to advanced technologies that support research. They are used in a wide variety of scientific domains, from high energy physics, astrophysics to humanities and the social sciences. By tailoring digital environments to community needs, science gateways perform a key role in integrating elements of the e-infrastructure landscape, providing online access to software, data, collaboration tools, instrumentation and high-performance computing, to facilitate increased research impacts.

Science gateways are enabling significant contributions in many research domains, with national and international initiatives to develop gateways further demonstrating their importance and value. This paper explores the global impact of these programs, highlighting their successes, values in the broader landscape and future focus. The paper begins with a discussion on the definition of terms, then documents national and international programs in this field. This investigation then highlights the role and value of science gateways in the digital research environment, and examines the impact of science gateways, to evidence how science gateways facilitate more efficient, open, reproducible research in bold new ways. A discussion of challenges and opportunities ahead concludes the study.

2. Definition of terms

A number of terms are often used in this field, including science gateways, virtual laboratories and virtual research environments (VREs). Different terms exist in large part for historical reasons; science gateways evolved in the USA, virtual laboratories in Australia, and VREs in Europe.

Shahand's analysis of science gateways research defines science gateways as "web-based enterprise information systems that provide scientists with customized and easy access to community-specific data collections, computational tools and collaborative services on e-Infrastructures." [1] This definition is similar to that used by the Science Gateways Community Institute, the USA's National Science Foundation-funded coordination project in this area, which also differentiates between science gateways and the generic cyberinfrastructure on which they build [2]. Australia's virtual laboratory community uses similar definitions, with an emphasis on access to integrated data, computational environments and tools [3].

Between 2004–2011, Jisc funded the development of a number of VREs in the UK, and defined VREs more broadly than science gateways and virtual laboratories: "The term VRE is now best thought of as shorthand for the tools and technologies needed by researchers to do their research, interact with other researchers ... and to make use of resources and technical infrastructures available both locally and nationally." [4] Horizon 2020, the European Commission's research and innovation framework programme, suggests that VREs "should integrate resources across all layers of the e-infrastructure (networking, computing, data, software, user interfaces), should foster cross-disciplinary data interoperability and should provide functions allowing data citation and promoting data sharing and trust." [5]

Carusi and Reimers work notes the relevance of alternative terms including collaborative e-research community, collaboratory and virtual research community [6] and identifies convergence on a set of characteristic features: "an electronic web-based environment for a) access to data, tools, resources; b) cooperation or collaboration with other researchers ...; c) cooperation at the intra- and inter-institutional levels; or d) preserving or taking care of data and other outputs." Candela, Castelli and Pagano's analysis of VREs also identifies five distinguishing features that are similar, but more focussed on serving the needs of a community of practice [7]: "(i) it is a web-based working environment; (ii) it is tailored to serve the needs of a community of practice; (iii) it is expected to provide a community of practice with the whole array of commodities needed to accomplish the community's goal(s); (iv) it is open and flexible with respect to the overall service offering and lifetime; and (v) it promotes fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance and attribution." Shahand also suggests that science gateways usually have five functional properties: usability, scalability, integration, automation and sharing and reuse [1].

It should be noted that science gateways can vary in scope depending on the problems they aim to address and the domains they support. In this paper, an inclusive definition of science gateways is used, covering all the aspects raised above.

3. Science Gateways Activities around the Globe

Activities involving science gateways are growing around the globe, with the establishment of programs, organizations, conferences and special issues in scientific journals.

3.1. Programs and Organizations

Whilst science gateways have historically been enabled through a wide variety of mechanisms, they are now increasingly facilitated through national and international programs that specifically facilitate their development and sustainability. International initiatives focused on science gateways have also evolved through the work of global consortiums, journals and conferences. National and international programs focusing on the development of science gateways include:

- *CANARIE*, a non-profit corporation, with the major investment in its programs and activities provided by the Government of Canada, funds the development of research software that enables Canadian researchers to more quickly and easily access research data, tools and collaborators. Since 2007, CANARIE has provided funding for 37 science gateway projects in disciplines such as high energy physics, astronomy, astrophysics, oceanography, human kinetics, robotics, bioinformatics, genomics, neurology, cartography, immunology, mechanical engineering, civil engineering, Arctic research, video analysis, animal biology, digital humanities, climatology, forestry, road traffic management, and e-Health [8].

- Science Gateways Community Institute (*SGCI*), funded in 2016 for USD$15 million over five years by USA's National Science Foundation, to act as a focal point to facilitate the development of a sustainable software ecosystem for science gateways [9]. SGCI's programs include a business incubator, extended developer support, scientific software collaborative, community engagement and exchange and workforce development. It is one of the two initial Scientific Software Innovation Institutes funded under NSF's Software Infrastructure for Sustained Innovation (SI2) program [10]. SI2 funds software projects of varying scales, from small research software groups to the large software institutes, including specific science gateways themselves as well as projects developing general software that can be used to build gateways.

- European Comission (*EC*) funding programs for research and innovation include the Seventh Programme Framework (*FP7*) and *Horizon 2020*. FP7 supported VRE projects from 2007-2013. For example, SCI-BUS explored new possibilities for European user communities to create custom science gateways through a generic-purpose gateway technology [11]. The project created a toolset to provide seamless access to major computing, data and networking infrastructures and services in Europe, including clusters, supercomputers, grids, desktop grids, academic and commercial clouds. Similarly, the Catania Science Gateway Framework [12] and its successor FutureGateways [13] provide application developers with tools to develop science gateways quickly and easily. Since 2014, Horizon 2020 has supported a number of European VRE projects including BlueBridge, EVER-EST, VRE4EIC, WEST-Life, VI-SEEM and MUG [14]. Most VREs are domain-specific, however there are also now initiatives creating toolsets for the creation of science gateways. For example, VRE4EIC, a Horizon 2020 research project totaling €4.37 million over 3 years, will provide a VRE reference model, a set of VRE components and a prototype Europe-wide interoperable VRE to empower multidisciplinary research communities [15]. Other Horizon 2020 projects include Sci-GalA (Energizing Scientific Endeavour through Science Gateways and meta-Infrastructures in Africa), a €1.4 million project that promotes the uptake of science gateways and strengthens and expands supporting e-infrastructures in Africa and beyond [16].

- National eResearch Collaboration Tools and Resources (*Nectar*), funded by the Australian Government, has distributed over AUD$20 million since 2011 specifically to facilitate software infrastructure programs that included the development of fourteen virtual laboratories. These virtual laboratories have received an additional AUD$20 million in co-investment [3]. By 2018, the virtual laboratories recorded over 23,000 users, and on average each virtual laboratory included users from over 20 international and 30 Australian organizations.

3.2. Collaborative Programs

A common observation in these national and international programs is that the development of science gateways is increasingly complex, therefore communities of practice have formed across international initiatives through global consortia. The very impetus for this paper comes from the *International Coalition on Science Gateways*, an international forum that brings together national, regional and international initiatives to provide leadership on future directions for science gateways, facilitate awareness and identify and share best practice in the field [17].

The Virtual Research Environment Interest Group (*VRE-IG*) within the Research Data Alliance (*RDA*) brings together initiatives actively developing VREs, virtual laboratories and science gateways, along with representatives of common infrastructure services and the researchers that seek to make use of these technologies. This group realized an effort to identify the necessary technical aspects, governance issues, and best practices required to support a more coordinated approaches [18]. The Virtual Research Environment Interest Group has been meeting at the twice-yearly RDA plenaries since March 2016 to discuss commonalities between science gateways, virtual research environments and virtual labs on intercontinental level. The goal of the interest group is to form a forum for discussions and support a common understanding of essential architectures as well as support a wider uptake of technologies via the gateways catalog of SGCI.

3.3. Conferences and Journal Special Issues

Conferences have been established by the science gateway community of practice to report on their advances, challenges, insights, and solutions.

The first International Workshop on the Gateway Computing Environments (*GCE*) took place within the Supercomputing conference in 2005. The CGE series successfully ran as half-day or full-day workshops hosted at Supercomputing, XSEDE, and other related conferences. From 2016 the *Gateway* conference series is organized yearly by the Science Gateways Community Institute as a two-day event that also includes tutorials and demos.

The International Workshop on Science Gateways (*IWSG*) series has been running in Europe since 2009 [19] as a three-day event with oral presentations and discussions, and that more recently has also included co-located satellite events. *IWSG-A*, the International Workshop on Science Gateways - Australia, occurred annually between 2015-17, in a one- to two-day format.

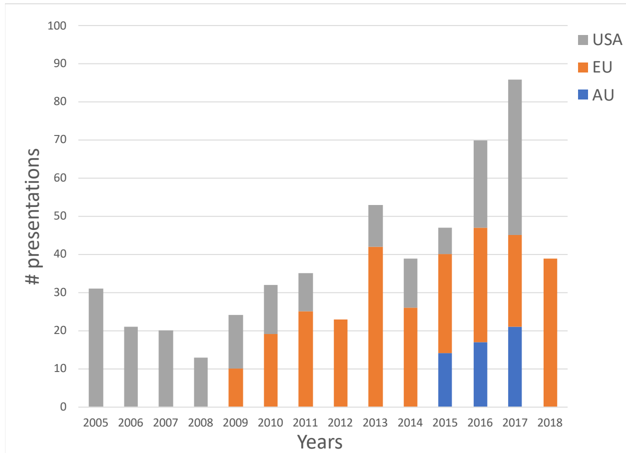

Figure 1: Number of talks and papers presented at Science Gateway events in the USA, Europe and Australia increases through time.

A summary of the events since 2005 is presented in table 1. Figure 1 illustrates the increasing number of publications and presentations in these conferences since their inception.

Initiated through the annual conferences, associated special issues on science gateways have been published by journals including the Journal of Grid Computing (JGC) [20, 21] and the Journal of Concurrency and Computation: Practice and Experience (CCPE) [22, 23, 24, 25, 26, 27, 28]. Currently the conference series in the USA, Europe, and Australia partner to organize a yearly special issue with the papers presented in all three events.

4. The value of science gateways in the e-infrastructures landscape

Science gateways are a key component of the emerging digital research environment. Researchers collaborate by using a global network of interacting digital platforms to access and share the leading-edge data and tools that are critical to their work. Gateways both facilitate, and are supported by, broader movements such as open research, open science, open source software and open data. Consequently, science gateways are valuable to a range of stakeholders:

students and educators, individual researchers, research communities, research organizations and institutions, industry, governments, infrastructure providers and funding agencies.

Defining science gateways in terms of common characteristics and functionality assists in identifying their value to their stakeholders. Science gateways lower barriers by hiding the complexity of the underlying digital research infrastructure and simplifying access to best-practice tools, data and resources, thereby democratizing their usage. An example is CBRAIN, a web-based collaborative research platform that offers transparent access to remote data sources, distributed computing sites, and an array of processing and visualization tools for neuroimaging research [29].

Science gateways can also enable collaboration and build communities, sharing data and analyses among multidisciplinary and geographically dispersed research groups, leading to increased openness. REMEDI illustrates well how successful collaboration was established through a science gateway: is a collaborative community of pharmacists, nurses, researchers, vendors and others working to improve patient safety and healthcare quality through the development and exchange of infusion pump medication administration knowledge and best practices [30].

Some gateways provide access to modelling and other software and hardware resources through a single portal. Researchers do not need to spend time downloading, installing and updating software on hardware that they also maintain. Now they can use the latest optimized software on powerful remote hardware completely through the web, of which nanoHUB is a good example [31].

Researchers no longer need to be physically co-located because resources can be globally distributed, with only an internet connection needed for participation. This also enables inclusion of less advantaged researchers/institutions. The Sci-GaIA project has demonstrated this through its tremendous success in deploying a vast array of applications available through the African Grid Science Gateway. For example, it supports a virtual collaborative community through the African Pharmacology Science Gateway and the Community Health Portal for health professionals and patients [16].

By sharing resources across multiple institutions, the costs of setting up and supporting research infrastructure is lowered, as each institution is no longer required to support a replica of data, compute and tools at their site. For gateways that are open source, their very building and evolution can be democratized with community members contributing in the development. Many gateway-building frameworks are available on github, for example Apache Apache Airavata [32], HUBzero [33], Galaxy [34], Drupal [35] and Django [36]

Science gateways provide these benefits to users by performing a key role in integrating e-infrastructure layers, in particular by linking together elements that can include data storage, tools, authentication, networks, cloud and high-performance computing, and access to data resources for reuse (sometimes called "data as infrastructure"). This integration tailors digital environments to community needs without the need for expertise in navigating the enabling information technology infrastructure that supports their work. They simplify linkage to other infrastructures, such as synchrotrons, ground-based telescopes, satellites, DNA sequencers, distributed archives and performance art studios. In some cases, science gateway architecture supports the whole research process from hypothesis generation to results analysis, including provenance information. One example is the VRE under construction in the EVER-EST project [37], which will support handling of research objects along the complete information lifecycle in earth science research. Science gateways also have a role in education, training researchers of the future and providing access to methods formerly only accessible to experts, for instance, CLEERhub [38] for Science, Technology, Engineering, and Mathematics (STEM) and STEMrelated disciplines, and Vortex Shedding, which provides a free online educational environment for high school and college level students to learn about physical phenomena [39].

Science gateways interact with the e-infrastructures landscape in multiple ways. At the broadest level, science gateways play a key role in driving standards and policy compliance, supporting initiatives including open research, open science, open source software, and open data. Zooniverse, for example, is a science gateway that promotes open citizen science, where anyone can be in the seat of a researcher (and define a project) or a volunteer (and perform some task in the project) [40]. Science gateways can also both drive standards and act as testbeds, as the increased user expectations encouraged by science gateways can drive requirements for harmonization. These standards often arise from sharing of best practice, with communities of practice addressing issues including reproducibility, sustainability, interfaces to cloud computing, workflows, integration of scientific instruments, success metrics, usability studies, scaling, mobile applications and security. An increasing number of international organizations address some of these issues, including the Software Sustainability Institute, the US Research Software Sustainability Institute (URSSI) conceptualization project "Working toward Sustainable Software for Science: Practice and Experiences" (WSSSPE [41]), the FORCE11 Software Citation Working Group [42] and the Centre for Open Science [43]. A one-week bootcamp offered by the Science Gateways Community Institute helps developers articulate the value of their work to key stakeholders and to create a strong development, operations, and sustainability plan. Working in teams, participants have the opportunity to network and establish relationships with people who are engaging in similar activities. An abridged version will be offered internationally for the first time in 2018. With diverse and constantly changing technologies available, collaboration among practitioners continues to be essential to share best practice and to avoid reinventing the wheel, helping developers to easily develop science gateways for specific user communities.

Science gateways also provide valuable resources for cross-disciplinary research, and increased interoperability across science gateways will enable more multidisciplinary research. The adoption of common interfaces and formats to build a global network of science gateways will further promote open and reproducible science, and will increase the availability and usage of existing scientific tools and data. This will lead to the emergence of a new class of scientific services such as application stores, search engines and continuous integration services. Science gateways are beginning to access the services of other gateways, allowing gateway developers to design interfaces and implement functionalities specific to their communities, yet use already built infrastructure as it exists elsewhere. For example, the Characterisation Virtual Laboratory produces and supports software that is used internationally [44], and their MyTardis software is being deployed by Euro-Bioimaging in partnership with ELIXIR Finland at the Global Bioimaging head node in Turku, Finland. Another example is the CIPRES science gateway [45], which provides an API interface to its software-as-a-service offerings, allowing others developing gateways to use those services from within their own frameworks.

Whilst some gateways already cross a number of disciplines to answer research questions, a global, decentralized network of science gateways may emerge. In this network, platforms would expose a consistent front through open specifications offering common interfaces, formats and protocols, allowing for the exchange of data, processing tools and experiments. In such a network, common web APIs such as Agave [46] or CARMIN [47] will expose methods to query and manipulate data, to run data processing tools and to share experiments. Description formats such as the Common Workflow Language [48] and Boutiques [49], which leverage the now-mature virtual containerization systems, will represent and install processing tools consistently in multiple science gateways from a single description. At the data level, domain-specific description formats such as the Neuroimaging Data Model [50], the Brain Imaging Data Structure [51], the Minimal Standard for Adaptive Immune Receptor Repertoires [52, 53], or the data models provided by the International Virtual Observatory Alliance (IVOA) [54], will facilitate the exchange of datasets and the improvement of existing data models for new categories of scientific experiments. The adoption of common interfaces and formats to build a global network of science gateways will further promote open and reproducible science, it will increase the availability and usage of existing scientific tools and data, and it will lead to the emergence of a new class of scientific services such as application stores, search engines and continuous integration services.

An important requirement for interoperability is a common vision about how to provide the research communities with federated access to a VRE. A great effort has been put in this direction by the EC-funded project AARC [55] (and by the recently approved AARC2) towards an interoperable architectural design, policy harmonization and community-driven piloting activity. Some examples of AARC-compliant e-infrastructures are the EGI CheckIn Service [56], the INDIGO-Datacloud [57] Authentication and Authorization Infrastructure (AAI) and the INAF Cherenkov Telescope Array (CTA) AAI which includes the INAF-CTA Science Gateway [58]. The H2020 VRE4EIC project is also dedicated to definition of an interoperability framework that will enable exchange of resources among science gateways more easily [15].

Related to the need for science gateway interoperability is a need for an effective discovery mechanism to assist researchers in identifying existing software that might meet their needs. Registries of science gateways and other software for research do exist, but there is no single authority for these resources at an international level. The current ecosystem is a combination of registries for individual, reusable gateways [59, 33] that do not necessarily inter-operate, general software registries that include scientific components [60, 61], funder-specific registries [62], and registries that are limited to one, or a handful of related disciplines [63, 64]. Since there is already a proliferation of registries as described above, a federated approach is more appropriate than the creation of yet another registry. Such a federation would not only support search and discovery, but in the longer term it opens the door for dynamic creation of workflows based on publicly available components.

The majority of analyses of both specific science gateways and large e-infrastructure programs emphasize the importance of appropriate skills and training, and many of the organizations mentioned here include a focus on this crucial need. For example the Science Gateways Community Institute features a Workforce Development component that includes a coding institute, workshops and summer internships where students are paired with gateway developers working on real world problems. Also, Indiana University offers a graduate level course on Science Gateway Architectures [65]. A key question is what skills do all researchers need, versus what will remain as specialist knowledge, particularly with regard to informatics. Where specialist skills are needed, career paths, recognition mechanisms and training opportunities are critical, as common issues emerge in integrating tools, applications, and data collections through a tailored web-based environment. It is also essential that scientists, researchers and students are able to learn and adopt a new set of software-related skills and methodologies, as well as learning to collaborate virtually amongst teams that are widely distributed. Many research communities or science gateways also provide their own programs, The Biodiversity and Climate Change Virtual Laboratory's EcoEd program provides training in the use of virtual laboratories and data repositories available to ecosystem scientists and lecturers [66].

5. The impact of science gateways

A range of ways exist to quantitatively provide evidence for the impact of individual science gateways:

- number of users and individual researchers,
- number of laboratories and groups served,
- number of organizations,
- computing infrastructure activity (number of jobs, computing time and storage),
- number of citations (to Science Gateways),
- number of (enabled) publications,
- value of access to software,
- value of access to data,
- contingent valuation
- efficiency savings, and

- return on investment

Traditional metrics such as user numbers are still actively used and some groups used more impact-focused studies to demonstrate contingent valuation. These are often used alongside emerging measures such as software citation [67]. However, different science gateways (programs) utilize different combinations of measures. It would also be useful to be able to analyse the sustainability of science gateways (beyond initial grant funding) as another measure of success.

It is difficult to make comparisons across science gateway programs due to their different structures and ways of measuring impact. For example, Nectar-funded virtual laboratories identify over 23,000 users; however, the methods used by each virtual laboratory to measure users can vary widely. In contrast, CANARIE defines users as referring to research teams or groups, rather than individual researchers. While the US-based XSEDE program does not fund gateways, dozens of gateways use its compute resources. In an Interim Project Report from 2018 [68], Table 12-1 shows gateway users varying between 10,000 and 12,000 in calendar 2017, about four times higher than active users at the command line. There are also many successful gateways that do not need high-end computing, for example, the vast majority of the more than a million nanoHUB users [31], for which such metrics would not be appropriate.

Part of the evidence for the value of science gateways comes from work expanding recognition of the importance of e-infrastructures, such as Mayernik, Hart, Maull and Weber's work [69]. They note the increasing recognition that "traditional assessments of research impact have missed broad swaths of important activities, including the benefits associated with the collection, management and preservation of digital resources, such as data and software, and the provision of research facilities and services, such as computational facilities and observational platforms". Metrics for quantitatively measuring the impacts of analytical tools over data are now beginning to emerge, and can contribute to the valuation of science gateways. Beagrie and Houghton's work on the European Molecular Biology Laboratory and European Bioinformatics Institute

(EMBL-EBI) assessed the value and impact of the EMBL-EBI by identifying four valuation levels: access (use) value, contingent valuation, efficiency savings, and return on investment [70]. This was applied to a range of EMBL-EBI services, including both data access and analytical services over the data one of very few studies examining the latter. In 2017, Nectar commissioned Victoria University to apply Beagrie and Houghton's methodology to evaluate the economic impact of three of virtual laboratories. The report measures the economic benefits created in five different ways. For all three of the virtual laboratories, each measure shows that the economic benefit is greater than the investment required. Taking a long term perspective, the research enabled by the virtual laboratories generates substantial returns compared to their costs [71].

The need for science gateways is also being demonstrated through increasing acknowledgment of the critical role of software in research. A 2009 survey by Hannay, MacLeod, Singer, Langtangen, Pfahl and Wilson with 2,000 responses showed that 84 percent of researchers view the development of software as "important or very important for their own research" [72]. The USA's National Science Foundation's research software vision identifies software as "directly responsible for increased scientific productivity and significant enhancement of researchers' capabilities" [10]. Further, a 2014 National Science Foundation-funded survey sent to NSF-funded principal investigators and Chief Information Officers and Chief Technology Officers at US academic institutions resulted in 5,000 respondents. In total 88% indicated to rely on science gateway-like interfaces to conduct their work and 57% were themselves involved in some capacity in the creation of these [73].

A recent study applied a similar methodology to the Industrial Ecology Virtual Laboratory (IELab), a high-performance computing lab used for compiling large-scale, high- resolution, enviro-socio-economic accounts for the purpose of conducting integrated sustainability assessment project [74]. Wiedmann's analysis of 30 IELab publications that were published in either peer-reviewed journal papers or in the form of conference proceedings, concluded that two-thirds of the studies would not have been possible without the Industrial Ecology Virtual

Laboratory, and a further 16% would have required considerable extra resources to complete. This type of contingent valuation could also be inferred from other metrics, such as the emerging emphasis on software citations, an area where organisations such as the FORCE11 Software Citation Implementation Working Group [42] is undertaking significant work. For example, the CIPRES Science Gateway (for phylogenetic research) has enabled 3,000 publications since 2010. Without this science gateway, many users would not have undertaken this type of research, instead needing to set up their own clusters, and install, maintain and optimize the many pieces of software offered via CIPRES [45].

6. Conclusion: opportunities for science gateways

Science gateways have been a valuable addition to the digital infrastructure landscape, facilitating more efficient, open, reproducible research. The many science gateway initiatives available provide abundant opportunities for reflection, identification of best practice and analysis of beneficial ways forward. Some of the key areas in which continued collaboration may advance the field include:

- Technical solutions for the development of science gateways, including interoperability, standards, software registries, and data management.
- Best practices and policies for the valuation of science gateways, including incentives for open science, reproducibility, data and software citation.
- Sustainability models for the maintenance, development, and exploitation of science gateways, including development of skills, training, career paths and funding.

For example, developing interoperability across science gateways is key to a successful conduct of collaborative data- and compute-intensive research, to enable open data and reuse of methods across domains and applications. The adoption of common interfaces and formats to build a global network of science gateways will further promote open and reproducible science, it will increase the availability and usage of existing scientific tools and data, and it will lead to the emergence of a new class of scientific services such as application stores, search engines and continuous integration services. Science gateways are beginning to access the services of other gateways, allowing gateway developers to design interfaces specific to their communities, yet use already built infrastructure as it exists elsewhere.

The majority of analyses of both specific science gateways and large e-infrastructure programs also emphasize the importance of appropriate skills and training, a crucial need for the future. A key question is what skills do all researchers need, versus what will remain as specialist knowledge, particularly with regard to informatics. Where specialist skills are needed, career paths, recognition mechanisms, and training opportunities are critical. Another issue is the degree to which domain-specific skills remain necessary for gateways developers. While most science gateways are developed for a specific research domain, common issues emerge in integrating tools, applications, and data collections through a tailored web-based environment. Web technologies such as HTML5, WebGL, and JavaScript frameworks have never been so agile and fast developing as in the last five years, leveraging possibilities to utilize applications more efficiently and more effectively with increased positive user experience. In addition, it is essential that scientists, researchers, and students are able to learn and adopt a new set of software-related skills and methodologies, as well as learning to collaborate virtually amongst teams that are widely distributed, often globally.

In conclusion, it is important that the field of science gateways continues to evolve, increasing interoperability to enable more multidisciplinary research, increasing collaboration and sharing mechanisms, to facilitate more efficient, open, reproducible research. Appropriately skilled users and developers also need to be trained in tandem with this software infrastructure, to ensure the maximum value of the infrastructure is realized, to further facilitate increased research impacts. The ongoing investment in national and international programs, in tandem with community and disciplinary initiatives, are facilitating the development of many communities of practice to address these issues, including ways to demonstrate the value of contributions of individuals, science gateways, and national and international programs to this field. Increasing coordination across these varied initiatives will continue to improve identification of best practice and development of policies and standards, enhancing the ability of science gateways to increase impact of research.

References

[1] S. Shahand, Science gateways for biomedical big data analysis, University of Amsterdam, Amsterdam, 2015. URL: http://hdl.handle.net/11245/1.490613.

[2] Wikipedia, Science gateway, 2018. URL: https://en.wikipedia.org/wiki/Science_gateway, [Accessed: 20 May 2018].

[3] Nectar, Nectar impact, 2016. URL: https://nectar.org.au/nectar-impact/, [Accessed: 20 May 2018].

[4] Jisc, Virtual research environment programme, 2014. URL: http://webarchive.nationalarchives.gov.uk/20140702163345/http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx, [Accessed: 20 May 2018].

[5] European Comission, EINFRA-9-2015 - e-Infrastructures for virtual research environments (VRE), 2015. URL: https://cordis.europa.eu/programme/rcn/664625_en.html, [Accessed: 20 May 2018].

[6] A. Carusi, T. Reimer, Virtual research environment collaborative landscape study, 2010. URL: https://spiral.imperial.ac.uk/bitstream/10044/1/18568/2/vrelandscapereport.pdf.

[7] L. Candela, D. Castelli, P. Pagano, Virtual research environments: An overview and a research agenda, Data Science Journal 12 (2013) GRDI75–GRDI81.

[8] CANARIE, Research Platforms, 2017. URL: https://www.canarie.ca/software/platforms/, [Accessed: 20 May 2018].

[9] SGCI, Science Gateway Community Institute, 2015. URL: http://sciencegateways.org/about/, [Accessed: 20 May 2018].

[10] National Science Foundation, Software Infrastructure for Sustained Innovation, 2016. URL: https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17526, [Accessed: 20 May 2018].

[11] P. Kacsuk (Ed.), Science Gateways for Distributed Computing Infrastructures: Development framework and exploitation by scientific user communities, Springer, Amsterdam, 2014. doi:10.1007/978-3-319-11268-8.

[12] V. Ardizzone, R. Barbera, A. Calanducci, et al., The DECIDE Science Gateway, J Grid Computing 10 (2012) 689707.

[13] FutureGateway Application Programming Interface, n.d. URL: https://www.indigo-datacloud.eu/future-gateways-programmable-scientific-portal, [Accessed: 7 June 2018].

[14] European Comission, Six new projects on e-Infrastructures for virtual research environments, 2015. URL: https://ec.europa.eu/programmes/horizon2020/en/news/six-new-projects-e-infrastructures-virtual-research-environments, [Accessed: 20 May 2018].

[15] VRE4EIC, Press release, 2016. URL: http://www.vre4eic.eu/publications/press-releases/69-a-4-37m-european-investment-towards-next-generation-virtual-research-environments-fo [Accessed: 20 May 2018].

[16] Sci-GaIA, About The Sci-GaIA project Energising Scientific Endeavour through Science Gateways and e-Infrastructures in Africa, 2015. URL: http://www.sci-gaia.eu/about-sci-gaia/, [Accessed: 20 May 2018].

[17] ICSG, International Coalition on Science Gateways, n.d. URL: http://www.icsciencegateways.org/, [Accessed: 20 May 2018].

[18] Research Data Alliance, Virtual Research Environment Interest Group, n.d. URL: https://rd-alliance.org/groups/vre-ig.html, [Accessed: 20 May 2018].

[19] International Workshop on Science Gateways, About IWSG, n.d. URL: http://iwsg-life.org/site/iwsglife/, [Accessed: 20 May 2018].

[20] Journal of Grid Computing, Special Issue: Science Gateways, volume 10, 2012. doi:10.1007/s10723-012-9245-0.

[21] N. Gesing, S.and Wilkins-Diehr, M. Barker, G. Pierantoni, Science gateway workshops 2015 special issue conference publications, Journal of Grid Computing 14 (2016) 495–498.

[22] Concurrency and Computation: Practice and Experience, Special Issue on Workshop on Grid Computing Portals (GCE 2005), volume 19, 2007. doi:10.1002/cpe.1258.

[23] Concurrency and Computation: Practice and Experience, Special Issue: Science Gateways-Common Community Interfaces to Grid Resources, volume 19, 2007. doi:10.1002/cpe.1098.

[24] S. Gesing, J. van Hemert, P. Kacsuk, O. Kohlbacher, Special issue: portals for life sciences – providing intuitive access to bioinformatic tools, Concurrency and Computation: Practice and Experience 23 (2011).

[25] Concurrency and Computation: Practice and Experience, Special Issue: Science Gateway Workshop 2013 (SGW 2013), volume 27, 2015. URL: https://onlinelibrary.wiley.com/toc/15320634/2015/27/2.

[26] N. Wilkins-Diehr, S. Gesing, T. Kiss, Science gateway workshops 2013 special issue conference publications, Concurrency and Computation: Practice and Experience 27 (2014) 253–257.

[27] N. Wilkins-Diehr, S. Gesing, T. Kiss, Science gateway workshops 2014 special issue conference publications, Concurrency and Computation: Practice and Experience 27 (2015) 4247–4251.

[28] Concurrency and Computation: Practice and Experience, GCE'15 Special Issue Conference Publications, volume 28, 2015. doi:`10.1002/cpe.3743`.

[29] T. Sherif, P. Rioux, M.-E. Rousseau, N. Kassis, N. Beck, R. Adalat, S. Das, T. Glatard, A. C. Evans, CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research, Frontiers in Neuroinformatics 8 (2014) 54.

[30] M. Zentner, R. Zink, HUBzero and CatalyzeCare: A community-driven platform for data sharing and collaboration in medical informatics research, PeerJ Preprints 5:e2819v1 (2017).

[31] K. Madhavan, L. Zentner, V. Farnsworth, S. Shivarajapura, M. Zentner, N. Denny, G. Klimeck, nanoHUB.org: Cloud-based Services for Nanoscale Modelling, Simulation, and Education, Nanotechnology Reviews 2 (2013) 107–117.

[32] Apache Airavata, n.d. URL: `https://airavata.apache.org/`, [Accessed: 7 June 2018].

[33] HUBzero website, n.d. URL: `http://hubzero.org`, [Accessed: 1 June 2018].

[34] Galaxy Project website, n.d. URL: `https://galaxyproject.org/`, [Accessed: 7 June 2018].

[35] Drupal website, n.d. URL: `https://www.drupal.org/`, [Accessed: 7 June 2018].

[36] Django Project website, n.d. URL: `https://www.djangoproject.com/`, [Accessed: 7 June 2018].

[37] The EVER-EST project, n.d. URL: https://ever-est.eu/about, [Accessed: 5 June 2018].

[38] CLEERhub, n.d. URL: https://stemedhub.org/groups/cleerhub, [Accessed: 5 June 2018].

[39] Vortex-Induced Vibration Simulation Website, n.d. URL: http://js-156-75.jetstream-cloud.org/vortexshedding/, [Accessed: 5 June 2018].

[40] Zooniverse webiste, n.d. URL: https://www.zooniverse.org, [Accessed: 5 June 2018].

[41] Working toward Sustainable Software for Science: Practice and Experiences, n.d. URL: http://wssspe.researchcomputing.org.uk/, [Accessed: 7 June 2018].

[42] FORCE11, Software Citation Implementation Working Group, 2017. URL: https://www.force11.org/group/software-citation-implementation-working-group, [Accessed: 1 June 2018].

[43] Center for Open Science website, n.d. URL: https://cos.io/, [Accessed: 7 June 2018].

[44] Characterisation Virtual Lab code repository, n.d. URL: https://github.com/CVL-dev, [Accessed: 7 June 2018].

[45] M. A. Miller, W. Pfeiffer, T. Schwartz, The CIPRES Science Gateway: Enabling High-impact Science for Phylogenetics Researchers with Limited Resources, in: Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the Campus and Beyond, XSEDE'12, ACM, New York, NY, USA, 2012, pp. 39:1–39:8. doi:10.1145/2335755.2335836.

[46] Texas Advanced Computing Centre, Agave Platform, 2016. URL: http://agaveapi.co/, [Accessed: 20 May 2018].

[47] CARMIN, Common API for Research Medical Imaging Network, n.d. URL: https://github.com/CARMIN-org, [Accessed: 20 May 2018].

[48] CWL, Common Workflow Language, n.d. URL: http://www.commonwl.org/, [Accessed: 20 May 2018].

[49] T. Glatard, G. Kiar, T. Aumentado-Armstrong, N. Beck, P. Bellec, R. Bernard, A. Bonnet, S. Camarasu-Pop, F. Cervenansky, S. Das, R. F. da Silva, G. Flandin, P. Girard, K. J. Gorgolewski, C. R. G. Guttmann, V. Hayot-Sasson, P. Quirion, P. Rioux, M. Rousseau, A. C. Evans, Boutiques: a flexible framework for automated application integration in computing platforms, 2017. URL: http://arxiv.org/abs/1711.09713. arXiv:abs/1711.09713.

[50] NIDM Working Group, Neuroimaging Data Model, n.d. URL: http://nidm.nidash.org/, [Accessed: 20 May 2018].

[51] K. J. Gorgolewski, et al., The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments, Scientific Data 3 (2016) 175–191.

[52] F. Rubelt, et al., Adaptive immune receptor repertoire community recommendations for sharing immune-repertoire sequencing data, Nat Immunol 18 (2017) 1274–1278.

[53] F. Breden, E. T. Luning Prak, B. Peters, F. Rubelt, C. A. Schramm, C. E. Busse, J. A. Vander Heiden, S. Christley, S. A. C. Bukhari, A. Thorogood, F. A. Matsen IV, Y. Wine, U. Laserson, D. Klatzmann, D. C. Douek, M.-P. Lefranc, A. M. Collins, T. Bubela, S. H. Kleinstein, C. T. Watson, L. G. Cowell, J. K. Scott, T. B. Kepler, Reproducibility and reuse of adaptive immune receptor repertoire data, Frontiers in Immunology 8 (2017) 1418.

[54] IVOA, International Virtual Observatory Alliance, n.d. URL: http://www.ivoa.net/, [Accessed: 20 May 2018].

[55] AARC, Authentication and Authorisation for Research and Collaboration project website, n.d. URL: https://aarc-project.eu/about/, [Accessed: 1 June 2018].

[56] EGI, EGI AAI Check-in Service, n.d. URL: https://wiki.egi.eu/wiki/AAI, [Accessed: 1 June 2018].

[57] INDIGO-DataCloud project website (INtegrating Distributed data Infrastructures for Global ExplOitation), n.d. URL: https://www.indigo-datacloud.eu/the_project, [Accessed: 1 June 2018].

[58] A. Costa, P. Massimino, M. Bandieramonte, U. Becciani, M. Krokos, C. Pistagna, S. Riggi, E. Sciacca, F. Vitello, An innovative science gateway for the cherenkov telescope array, Journal of Grid Computing 13 (2015) 547–559.

[59] Alces Flight Compute website, n.d. URL: https://alces-flight.com/, [Accessed: 1 June 2018].

[60] GitHub website, n.d. URL: http://github.com, [Accessed: 5 June 2018].

[61] OpenHatch website, n.d. URL: http://openhatch.org/, [Accessed: 5 June 2018].

[62] CANARIE research software portal, n.d. URL: http://science.canarie.ca, [Accessed: 5 June 2018].

[63] iPlant Collaborative, n.d. URL: https://en.wikipedia.org/wiki/IPlant_Collaborative, [Accessed: 5 June 2018].

[64] CYVERSE website, n.d. URL: http://www.cyverse.org/, [Accessed: 5 June 2018].

[65] Indiana University – School of Informatics, Computing, and Engineering, Course: Science Gateway Architectures, Fall 2017. URL: https://www.sice.indiana.edu/graduate/courses/index.html?number=b649&department=CSCI, [Accessed: 5 June 2018].

[66] C. Huijbers, EcoEd training for first-rate science education, n.d. URL: http://www.bccvl.org.au/ecoed-training-for-first-rate-science-education/, [Accessed: 5 June 2018].

[67] A. Smith, D. Katz, K. Niemeyer, FORCE11 Software Citation Working Group, Software Citation Principles, PeerJ Computer Science 2:e86 (2016).

[68] XSEDE: The Extreme Science and Engineering Discovery Environment, Interim project report 5, 2018. URL: http://hdl.handle.net/2142/99773.

[69] M. S. Mayernik, D. L. Hart, K. Maull, N. M. Weber, Assessing and tracing the outcomes and impact of research infrastructures, Journal of the Association for Information Science and Technology 68 (2016) 1341–1359.

[70] N. Beagrie, J. Houghton, The value and impact of the european bioinformatics institute, 2016. URL: https://beagrie.com/static/resource/EBI-impact-report.pdf.

[71] NECTAR, NECTAR virtual labs have huge impact, n.d. URL: https://nectar.org.au/vls-have-huge-economic-research-impact/, [Accessed: 5 June 2018].

[72] J. Hannay, C. MacLeod, J. Singer, H. Langtangen, D. Pfahl, W. G, How do scientists develop and use scientific software, in: Proceedings of 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering, ACM Computer Science, 2009, pp. 1–8. doi:10.1109/SECSE.2009.5069155.

[73] K. Lawrence, M. Zentner, N. Wilkins-Diehr, W. JA, M. Pierce, S. Marru, Science gateways today and tomorrow: Positive perspectives of nearly 5,000

members of the research community, Concurrency and Computation: Practice and Experience 27 (2015) 4252–5268.

[74] T. Wiedmann, An input-output virtual laboratory in practice - survey of uptake, usage and applications of the first operational IELab, Economic Systems Research 29 (2017) 296–312.

Table 1: Overview of Science Gateways events showing year, number of presentations (talks and papers), event name, location, and links to the proceedings and/or program.

Year	#	Event	Location	Proceedings and Agendas
2005	15	Science Gateways[1]	Chicago, US-IL	https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.1098
2005	16	GCE	Seattle, US-WA	http://onlinelibrary.wiley.com/doi/10.1002/cpe.1258/full
2006	21	GCE	Tampa, US-FL	http://www.cogkit.org/GCE06
2007	20	GCE	Reno, US-NV	https://www.researchgate.net/publication/ 259366865_International_Workshop_on_Grid_Computing_Environments_2007_in_Conjunction_with_SC07
2008	13	GCE	Austin, US-TX	https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4729055
2009	14	GCE	Portland, US-OR	http://dblp.uni-trier.de/db/conf/sc/gce2009.html
2009	10	IWPLS[2]	Edinburg, UK	http://ceur-ws.org/Vol-513/
2010	13	GCE	New Orleans, US-LA	http://www.proceedings.com/10226.html
2010	19	IWSG	Catania, IT	http://agenda.ct.infn.it/event/347/
2011	10	GCE	Seattle, US-WA	https://dl.acm.org/citation.cfm?id=2110486
2011	25	ISWG-Life	London, UK	https://sites.google.com/a/staff.westminster.ac.uk/iwsg-life2011 http://ceur-ws.org/Vol-819/

[1] with Global Grid Forum

[2] International Workshop on Portals for Life Sciences

Table 1 – continued from previous page

2012	23	IWSG -Life	Amsterdam, NL	https://sites.google.com/site/iwsglife2012
				http://ebooks.iospress.nl/volume/healthgrid-applications-and-technologies-meet-science-gateways-for-life-sciences
2012	n.a.	CGE		not held this year
2013	11	SGCI Workshop[3]	Indianapolis, US-IN	https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6689497
2013	42	IWSG	Zurich, CH	https://en.xing-events.com/iwsg2013.html
				http://ceur-ws.org/Vol-993/
2014	13	GCE	New Orleans, US-LA	https://dl.acm.org/citation.cfm?id=2690887
2014	26	IWSG	Dublin, IE	https://sites.google.com/a/my.westminster.ac.uk/iwsg2014/home/dates
				https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6881322
2015	7	GCE	Boulder, US-CO	https://onlinelibrary.wiley.com/doi/epdf/10.1002/cpe.3743
2015	26	IWSG	Budapest, HU	https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7217893
2015	14	IWSG-A	Brisbane, AU	https://sites.google.com/site/iwsglife/about-iwsg-a/iwsg-a-2015
2016	23	Gateways	San Diego, US-CA	https://sciencegateways.org/gateways2016/program
				https://gateways2016.figshare.com
2016	30	IWSG	Rome, IT	https://sites.google.com/a/nd.edu/iwsg2016/homehttp://ceur-ws.org/Vol-1871
2016	17	IWSG-A	Melbourne, AU	https://sites.google.com/site/iwsglife/about-iwsg-a/iwsg-a-2016

[3] in conceptualization phase

Table 1 – continued from previous page

2017	41	Gateways	Ann Arbor, US-MI	https://sciencegateways.org/web/gateways2017/program
				https://gateways2017.figshare.com
2017	24	IWSG	Polzan, PO	http://iwsg2017.psnc.pl/programme
2017	21	IWSG-A	Brisbane, AU	http://iwsg-life.org/site/iwsglife/about-iwsg-a
2018	39	ISWG	Edinburg, UK	https://sites.google.com/a/nd.edu/iwsg2018