Towards a Science Gateway Reference Architecture

Marlon E. Pierce
Science Gateways Research Center
Indiana University
Bloomington, Indiana USA

Mark A. Miller
San Diego Supercomputer Center
San Diego, California USA

Emre H. Brookes
University of Texas Health Science Center San Antonio
San Antonio, Texas USA

Mona Wong
San Diego Supercomputer Center
San Diego, California USA

Enis Afgan
Johns Hopkins University
Baltimore, Maryland USA

Yan Liu
University of Illinois
Urbana-Champaign, Illinois USA

Sandra Gesing
Center for Research Computing
University of Notre Dame
Notre Dame, IN, USA

Maytal Dahan
Texas Advanced Computing Center
Austin, Texas USA

Suresh Marru
Science Gateways Research Center
Indiana University
Bloomington, Indiana USA

Tony Walker
Research Technologies
Indiana University
Bloomington, Indiana USA

***Abstract*—Science gateways have been developed over the last twenty years and have grown into a large community of practice, as evidenced by international workshops and conferences. Because of the diversity of approaches to creating science gateways and the always changing landscape of technologies, the community lacks a common definition for the term "science gateway" itself and common terminology for describing the common components of a gateway architecture. Instead, a wide range of definitions and understandings exist and are used in different communities; this is evident, for example, in discussions whether science gateways are the same as virtual research environments. This paper attempts to address these issues by focusing on how science gateways support scientific research and considering the consequences on cyberinfrastructure.**

***Keywords*—science gateways, cyberinfrastructure**

I. Introduction

Science gateways are commonly described as user-centric environments that enable broader and deeper use of advanced computing resources, storage, data collections, and scientific applications. Gateways include graphical user interfaces (frequently Web browser-based), application programming interfaces (APIs), and middleware that provide access to software and data. Many modern gateways integrate diverse technologies, pulling together databases, messaging systems, content management systems, identity management systems, job submission systems, data catalogs, and other components into a unified working environment. Gateways are described in a significant body of literature [1-6]. There are many highly successful science gateways that support thousands of scientific users [7-12, 24]. An increasing number of gateway frameworks and platforms [13-17] exist to help create new gateways. Efforts like the XSEDE Gateway Cookbook [19, 20] have taken a cookbook approach to summarize the architecture and motivations of these getaways and provide an overview but did not normalize these disconnected recipes into a cohesive gateway definition.

We believe it is useful to step back from these operational descriptions of gateways to examine the reasons why gateways exist and how they may continue to thrive in the future, regardless of evolving technologies. Identifying these central propositions may lead to a stronger definition of science gateways, clarify terminology used by gateway practitioners, and clarify the relationships of science gateways to other types of cyberinfrastructure and distributed systems. We hope this will be useful to the community and also to those outside the field, including those interested in joining the field, operators of advanced computing and cloud resources, developers of non-gateway middleware such as workflow systems, and decision makers at universities and government agencies who recognize the need to provide science gateway capabilities to support their researchers but who are not yet familiar with the community. Additionally, we hope this paper will be useful to scientific researchers themselves since science gateways are created for them to help make their research processes more efficient across system, organizational, and national boundaries.

II. Science Gateways and Scientific Research

At their core, science gateways are created to support scientific research, either directly or indirectly through education and dissemination. The exact nature of this support for research depends on the specific gateway, but the general concept is useful as a starting point for describing a set of characteristic features of science gateways.

Scientific research consists of a) exploration of the current state of a research field, b) formulation of testable hypotheses or research questions, c) design and execution of experiments, d) management and sharing of data and metadata about experiments, e) analysis of experiments and development of conclusions, and f) communication of the conclusions and supporting methods through a broadening circle of colleagues, culminating in broadly available formal publications that are accessible to the community and reproducible, hopefully leading to the start of a new cycle of research. Strong or at least promising findings form the basis of future investigations. Important results of scientific research that are not typically included in the traditional sequential description include dead ends, ambiguous or opposing results, and accidental discoveries that happen during a research effort that alter its course. One may think of research as a mapping exercise of new terrain, with the unfortunate limitation that only the routes to specific destinations reach formal publication.

We now wish to make a specific connection between scientific research and the capabilities offered by science gateways. Publication, in its broadest sense, is the keystone activity of scientific research, in which conclusions (including negative ones) enter the community for larger inspection, debate, verification, and extension; see Fig 1. A science gateway is a software implementation that creates a specific set of capabilities, based in large part on access to externally managed resources, to support the creation, sharing, publication, and broad distribution of scientific research results. “Publication” may mean the traditional processes used by peer-reviewed journals and the custom of pre-publication used in many scientific fields to circulate results quickly, stake a claim to a particular finding, and solicit initial feedback. It may also mean the broader dissemination of findings through non-traditional means that are associated with research altmetrics [18].

Many science gateways measure their success by the scientific publications that they support; nanoHUB, for example, calculates its own h-index (https://nanohub.org/citations). We believe it is useful to examine more directly how science gateways can broadly support the scientific publication process and how this should be reflected in a reference architecture.

We see two important roles for science gateways on the publication process: they can support research as it is currently practiced, including training in practices through education, and they can help the research process expand beyond some of the limits imposed by conventional journal publication practices. Gateways can provide both restricted and public access to all experiments and analysis, supporting “actionable” or “living” publications [21-23], support replication of results, and generally promote wider, interactive discussion of new findings among researchers.

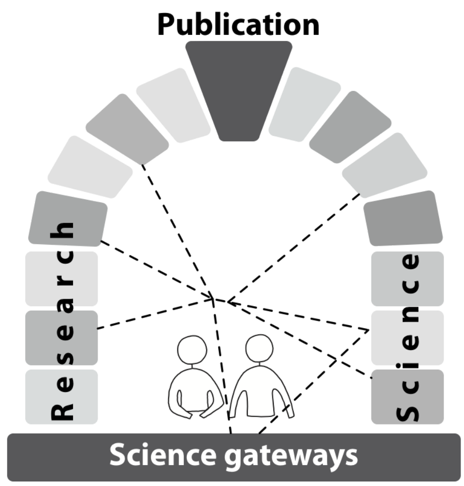

Fig. 1. A symbolic representation of the underpinnings science gateways provide in supporting individual steps of a research activity and collection of results for science to deliver a more accessible environment for the researchers.

We will now work through some of the implications of this assumption on key features that science gateways may choose to implement. We note that some gateways may seek to encapsulate the entire publication process in their implementations, but we do not view this as the end goal for all gateways. A gateway may, for example, share data within a community or publicly, and thus support the “publication” of data sets that are easily referenceable. A gateway may also provide a software-as-a-service reference implementation as an essential component of a publication, assisting both reviewers and readers.

Recognition of Users: Gateways provide authentication, authorization, and identity management. This is conventionally presented as necessary to protect valuable backend resources like supercomputers. By thinking instead of science gateways as researcher-centric cyberinfrastructure, we see an expanded purpose for identity management. Recognition of the user is a prerequisite for the capabilities that follow, all of which support the research process. This recognition may be simple authentication or it may incorporate the notion of user roles that distinguishes different categories of users and the functions and information each role is allowed to access. Additionally, gateways may allow project-based user groups to support relationships among users.

Integration of Services: Gateways act as agents that integrate scientific and other services for their users. These services may be implemented by the gateway itself (such as access to supercomputers), or they may be provided by external entities, including other gateways. For example, a gateway may help a user search and retrieve a data set from an external catalog service provider and then take action on it, using it as input to a simulation.

Organization of User Interactions: Gateways help scientific users search and explore data sets and conduct computational experiments. The latter include both input and output data and metadata that others may explore. It is thus useful to organize these interactions into "sessions". Sessions capture the state of a user's interactions with the system. Sessions may be organized into hierarchies or other structures, and they can vary in their level of prescriptive detail: a session could be implemented as a set of arbitrary or predefined key-value pairs or structures. Sessions may also be annotatable; that is, the researcher may annotate a session to explain the purpose, insights, etc of the session. These annotations may be part of the session (a comment field), or they may be external. The latter option requires the session and perhaps its elements to have pointers; this is simply the hypertext concept, in which the annotation hypertext uses URLs to refer to supporting or related digital entities.

Persistence of User Interactions: Science gateways allow researchers to recover previous sessions after initial interaction. Gateways support this feature to provide reproducibility or repeatability, help users organize their results, and avoid unnecessary repetition. Persistent sessions allow users to check their work and assist gateway operators in diagnosing user reported issues. Persistence is also needed to support annotation. In practice, full persistence is frequently limited because of large file storage limitations and other practical considerations. However, it is still possible to retain the metadata of a session, and the metadata can be used to replicate the experiment. We note the relationship here to the concept of provenance [25, 26], although we wish to avoid specific implementation considerations.

Sharing and Publication of User Interactions: Sessions are a core implementation concept for many science gateways. A specific scientific publication may be supported by many different sessions, perhaps from many different researchers. Sessions and their constituent elements should therefore be sharable. This may be done in widening circles: graduate students may review a set of computational experiments with their advisors and colleagues before depositing preprints in public archives and sending papers to journals for peer review. Gateways should therefore provide sharing mechanisms for results that map to these access levels. The publication may directly or in supplementary material reference the supporting sessions in the science gateways, exposing them to the community at large. A research paper itself may or may not be written using tools provided by the science gateway (such as a electronic notebook [10]), but the gateway should provide a way for the results that are used in the paper (and its drafting process) to be accessed and reviewed by other researchers.

In summary, the publication process, in the general sense of communicating scientific results, interpretations, and evidence in a convincing manner to a professionally skeptical audience, is central to scientific research. Science gateways can help with the publication process by supporting the management and organization of experimental results, the review process, and replicability. They may further help expand the publication process beyond its conventional limitations by supporting the dissemination of unpublished results. The latter may include experiments that are deemed failures or preliminary results that are later discarded after the researcher focused on a different aspect of the problem.

III. Considerations for a Science Gateway Reference Architecture

The conceptual features defining a science gateway need to be embodied into a framework implementation and delivered as a functional service to its users. The implementation of these features largely depends on the science domain a gateway serves so that it can accommodate the research workflow of its domain community. In this section, we look at a cross-section of such implementations to extract a consensus set of gateway attributes that help define general elements needed by a science gateway. These attributes are often interconnected and depend on each other to offer the comprehensive user experience science gateways target. However, a science gateway does not need to possess all the attributes, particularly not at the outset, and they can instead be built or cultivated as the complexity and size of the community a gateway serves grows (see Fig. 2).

- **Services** are the internal components a gateway requires to operate. For example, they handle authentication and authorization or manage retrieval, caching, and persistence of datasets in cases of federated storage configurations. Services also include session management and support for sharing. Overall, services represent the "glue" that make up the gateway framework and allow higher-level features to be exposed to the researchers.
- **Workspace** represents the main interface to the gateway, allowing the researcher to interact with the gateway's services. The workspace focuses on improving accessibility of the tools and services exposed by the gateway in a way most suitable to the specific domain and different roles of researchers. Ideally, these role-specific views offered by the dashboard are customizable and take into consideration the researcher's vocabulary and workflow preferences. By offering an easy-to-use interface the workspace lowers the barrier of entry by abstracting complicated tasks into easy to understand interfaces.
- **Integration** is characterized by the ability of a gateway to connect multiple disparate elements (e.g., tools, resources) into a unified interface. This interface can be utilized internally by the workspace and services as well as by other gateways or applications through its API. The integration layer focuses on translating gateway-specific representation of data, inputs, etc. to the format required by the actual tool or resources performing the requested action while exposing a well-defined, documented, and consistent interface to its users.
- **Infrastructure** is typically the workhorse of a gateway. While the workspace exposes the available functionality in an accessible format, the gateway must interface to a specific and often highly specialized set of compute and storage resources by performing the necessary configurations, data transfers, authentication, etc. As the complexity or popularity of a gateway increases, the gateway may also need to accommodate scaling by implementing or more robust or diverse infrastructure.
- **Community** is a pivotal element of a gateway that drives its success. The gateway needs to facilitate collaboration of its members as well as provide means of offering support. With time, if successful at forming a devoted community, the gateway can also benefit from community contributions, which help fuel future direction and sustainability.

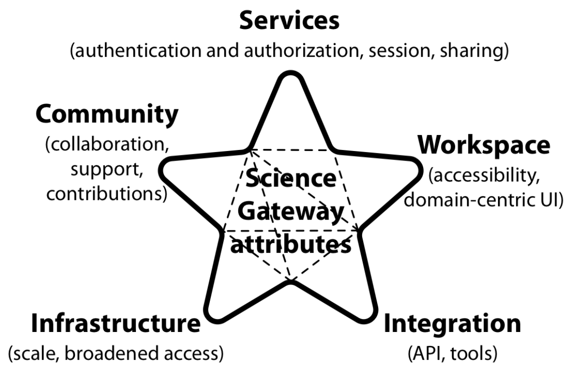

Fig. 2. Interconnected attributes of a science gateway.

IV. Evaluation

To illustrate the concepts in the previous sections, we apply them to the very well-known science gateway, nanoHUB. nanoHUB is a science and engineering gateway for enabling and broadening nanotechnology research and education. As a well-established gateway with over one million visitors using hundreds of simulation tools and running millions of simulation jobs a year, the architecture of nanoHUB has been evolving for years to incorporate a rich set of community requirements and gateway capabilities. However, the initial reference architecture concepts proposed in this paper can be well applied to capture major components in this gateway.

In nanoHUB community, a few users contribute content and tools while the majority of the users rely on nanoHUB content for research and education purposes. User-oriented design and development principles determine the advancement of nanoHUB. The gateway provides a comprehensive framework to monitor user activities, cite and publish contents, and make content reproducible through cached jobs and Jupyter notebooks. Gateway development takes the "adopt-and-adapt" strategy based on user requirements. Broad community usage is a major success metric of this gateway.

A user's workspace in nanoHUB is customized based on user preference and usage history. As a strategy to promote user engagement in content creation, the citation and publication framework keeps track of content usage and publishes

DOI-indexed citations on nanoHUB, which is also indexed by Web of Science (WOS) and Google Scholar. A recommendation system has been developed to learn user patterns in order to better serve nanoHUB content.

The nanoHUB user environment is backed by an effective integration process executed by a sizeable developer team. The integration workload for nanoHUB is enormous, including technology integration, a well-engineered content and tool development process, cyberinfrastructure (computing venues) integration, gateway operation, etc. As a result, HUBZero captures nanoHUB middleware components and is released as a platform approach for supporting general gateway development and hosting for other science domains.

A major goal of nanoHUB infrastructure provisioning is to provide sufficient computing power for nano simulations. Various computing venues are continuously incorporated for that purpose, including local clusters, high-performance (e.g., XSEDE), high-throughput (e.g., OSG), volunteer computing (BOINC), and cloud computing resources.

nanoHUB simulation tools and contents are mainly accessed by end users on nanoHUB portal. To accommodate future growth of nanoHUB, their gateway team is developing a universal service-oriented architecture based on REST web service protocols. The service computing approach will scale nanoHUB content/tool access and make the gateway programmable.

V. Discussion, Conclusions, and Future Directions

It is the intention of this paper to consider some defining features of science gateways and promote greater discussion within the community on what exactly a science gateway is and does. Instead of basing these on operational considerations, we instead posit that science gateways are user-centric cyberinfrastructure that support scientific research by federating access to diverse remote resources. Focusing on support for a broadly defined publication process, we identify the following key characteristics: recognition of users, integration of services on behalf of the user, creation of sessions to enable the user to track work, persistent archiving of sessions, and sharing and publication of sessions and their content to a broader community.

This definition may help science gateways to clarify their mission to themselves and to stakeholders in science, engineering, and scholarship. This may also help determine areas that need further development within specific projects and across the community. It may provide guidance to gateways on how to measure success, and how they want to present this measurement to stakeholders ranging from users to funding agencies. In particular, many science gateways measure success through supported scientific publications, but gateways should consider ways to make this support richer and more direct, such as through the use of persistent identifiers, stronger guarantees on persistence, integration with publication mechanisms like FigShare, richer metadata, and more powerful sharing mechanisms.

A criticism of this paper's thesis is that its emphasis on supporting scientific publication does not address all of the potential uses of a gateway, such as the support for education. We use the term "publication" in the broad sense of making scientific data, results, and conclusions available to an increasingly broad circle, and science education is a form of practice publication through problem solving. We thus believe the current discussion can accommodate education-centric gateways.

Another objection is that "recognition of users" excludes many science gateways and gateway-like services that do not require authentication or have notions of sessions. Gateways support scientific research through a reproducible sequence of steps to produce a particular state (such as a particular simulation result). Sessions represent this state, and some form of identification allows the user to manage the session. Gateways that do not support identities and sessions explicitly can still support scientific research, but the steps in the creation of the state must be communicated through some means outside the gateway, such as a written description. Using sessions associated with identities is a more straightforward mechanism.

Another interesting variation may be a science gateway that helps manage other cyberinfrastructure. A gateway may allow users, for example, to dynamically create and manage resources, which are in turn used for scientific research. These gateways may not directly support scientific publications, but they may still implement some of the basic abstractions described here.

Future work is to consider a comprehensive survey of the larger community to map the concepts of this paper to specific gateways. The basic ideas described in this paper may also serve as the basis of a reference architecture for gateways that can be developed using the mechanisms of The Open Group Architecture Framework [27].

References

1. Wilkins-Diehr, N., Gannon, D., Klimeck, G., Oster, S. and Pamidighantam, S., 2008. TeraGrid science gateways and their impact on science. *Computer*, *41*(11).

2. Wilkins-Diehr, N., 2007. Special issue: science gateways—common community interfaces to grid resources. *Concurrency and Computation: Practice and Experience*, *19*(6), pp.743-749.
3. Lawrence, K.A., Zentner, M., Wilkins-Diehr, N., Wernert, J.A., Pierce, M., Marru, S. and Michael, S., 2015. Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community. *Concurrency and Computation: Practice and Experience*, *27*(16), pp.4252-4268.
4. Gesing, S. and Wilkins-Diehr, N., 2015. Science gateway workshops 2014 special issue conference publications. *Concurrency and Computation: Practice and Experience*, *27*(16), pp.4247-4251.
5. Wilkins-Diehr, N., Gesing, S. and Kiss, T., 2015. Science gateway workshops 2013 special issue conference publications. *Concurrency and Computation: Practice and Experience*, *27*(2), pp.253-257.
6. Gesing, S., Wilkins-Diehr, N., Barker, M. and Pierantoni, G., 2016. Science Gateway Workshops 2015 Special Issue Conference Publications. *Journal of Grid Computing*, *14*(4), pp.495-498.
7. Afgan, E., Baker, D., Van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C. and Grüning, B., 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research*, *44*(W1), pp.W3-W10.
8. Blankenberg, D., Kuster, G.V., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A. and Taylor, J., 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Current protocols in molecular biology*, pp.19-10.
9. Miller, M.A., Pfeiffer, W. and Schwartz, T., 2010, November. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Gateway Computing Environments Workshop (GCE), 2010* (pp. 1-8). Ieee.
10. Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B.E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J.B., Grout, J., Corlay, S. and Ivanov, P., 2016, May. Jupyter Notebooks-a publishing format for reproducible computational workflows. In *ELPUB* (pp. 87-90).
11. Klimeck, Gerhard, Michael McLennan, Sean P. Brophy, George B. Adams III, and Mark S. Lundstrom. "nanohub. org: Advancing education and research in nanotechnology." Computing in Science & Engineering 10, no. 5 (2008): 17-23.
12. Nakandala, S., Pamidighantam, S., Yodage, S., Doshi, N., Abeysinghe, E., Kankanamalage, C.P., Marru, S. and Pierce, M., 2016, July. Anatomy of the SEAGrid Science Gateway. In Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale (p. 40). ACM.
13. Savelyev, A., Brookes, E., 2017, GenApp: Extensible tool for rapid generation of web and native GUI applications, *Future Generation Computer Systems*, DOI: 10.1016/j.future.2017.09.069.
14. Pierce, M.E., Marru, S., Gunathilake, L., Wijeratne, D.K., Singh, R., Wimalasena, C., Ratnayaka, S. and Pamidighantam, S., 2015. Apache Airavata: design and directions of a science gateway framework. *Concurrency and Computation: Practice and Experience*, *27*(16), pp.4282-4291.
15. Kacsuk, P., Farkas, Z., Kozlovszky, M., Hermann, G., Balasko, A., Karoczkai, K. and Marton, I., 2012. WS-PGRADE/gUSE generic DCI gateway framework for a large variety of user communities. *Journal of Grid Computing*, *10*(4), pp.601-630.
16. McLennan, M. and Kennell, R., 2010. HUBzero: a platform for dissemination and collaboration in computational science and engineering. *Computing in Science & Engineering*, *12*(2).
17. Goff, S.A., Vaughn, M., McKay, S., Lyons, E., Stapleton, A.E., Gessler, D., Matasci, N., Wang, L., Hanlon, M., Lenards, A. and Muir, A., 2011. The iPlant collaborative: cyberinfrastructure for plant biology. *Frontiers in plant science*, *2*, p.34.
18. Priem, J., Taraborelli, D., Groth, P. and Neylon, C., 2010. Altmetrics: A manifesto.
19. Marru, Suresh, Rion Dooley, Nancy Wilkins-Diehr, Marlon Pierce, Mark Miller, Sudhakar Pamidighantam, and Julie Wernert. "Authoring a Science Gateway Cookbook." In Cluster Computing (CLUSTER), 2013 IEEE International Conference on, pp. 1-3. IEEE, 2013.
20. https://www.xsede.org/web/gateways/gateways-cookbook
21. Broggini, F., Dellinger, J., Fomel, S. and Liu, Y., 2017. Reproducible research: Geophysics papers of the future—Introduction.
22. Brinckman, A., Chard, K., Gaffney, N., Hategan, M., Jones, M.B., Kowalik, K., Kulasekaran, S., Ludäscher, B., Mecum, B.D., Nabrzyski, J. and Stodden, V., 2018. Computing environments for reproducibility: Capturing the “Whole Tale”. *Future Generation Computer Systems*.
23. Bánáti, A., Kacsuk, P. and Kozlovszky, M., 2017. Reproducibility analysis of scientific workflows. *Acta Polytechnica Hungarica*, *14*(2), pp.201-217.
24. Quinn, P.J., Barnes, D.G., Csabai, I., Cui, C., Genova, F., Hanisch, B., Kembhavi, A., Kim, S.C., Lawrence, A., Malkov, O. and Ohishi, M., 2004, September. The International Virtual Observatory Alliance: recent technical developments and the road ahead. In *Optimizing Scientific Return for Astronomy through Information Technologies* (Vol. 5493, pp. 137-146). International Society for Optics and Photonics.
25. Simmhan, Y.L., Plale, B. and Gannon, D., 2005. A survey of data provenance in e-science. *ACM Sigmod Record*, *34*(3), pp.31-36.
26. Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J. and Plale, B., 2011. The open provenance model core specification (v1. 1). *Future generation computer systems*, *27*(6), pp.743-756.
27. Haren, V., 2011. *TOGAF Version 9.1*. Van Haren Publishing.