

# Science Gateway Implementation at the University of South Dakota: Applications in Research and Education

Joseph D. Madison  
Research Computing Group,  
University of South Dakota,  
Vermillion, SD, 57069, USA,  
joseph.madison@coyotes.usd.edu

Eroma Abeysinghe  
Science Gateways Research  
Center, Pervasive Technology  
Institute, Indiana University,  
Bloomington, IN, 47408, USA,  
eabeysin@iu.edu

Sudhakar Pamidighantam  
Science Gateways Research  
Center, Pervasive Technology  
Institute, Indiana University,  
Bloomington, IN, 47408, USA,  
pamadigs@iu.edu

Suresh Marru  
Science Gateways Research  
Center, Pervasive Technology  
Institute, Indiana University,  
Bloomington, IN, 47408, USA,  
smarru@iu.edu

Marcus Christie  
Science Gateways Research  
Center, Pervasive Technology  
Institute, Indiana University,  
Bloomington, IN, 47408, USA,  
machrist@iu.edu

Douglas M. Jennewein  
Research Computing Group,  
University of South Dakota,  
Vermillion, SD, 57069, USA,  
doug.jennewein@usd.edu

Marlon Pierce  
Science Gateways Research Center, Pervasive Technology Institute, Indiana University, Bloomington, IN, 47408, USA,  
mapierc@iu.edu

## ABSTRACT

Science Gateways are virtual environments that accelerate scientific discovery by enabling scientific communities to more easily and effectively utilize distributed computing and data resources. Successful Science Gateways provide access to sophisticated and powerful resources, while shielding their users from the underlying complexities. Here we present work completed by the University of South Dakota (USD) Research Computing Group in conjunction with the Science Gateways Community Institute (SGCI) [1] and Indiana University on setting up a Science Gateway to access USD's high-performance computing resources. These resources are now available to both faculty and students and allow ease of access and use of USD's distributed computing and data resources. The implementation of this gateway project has been multifaceted and has included placement of federated user login, user facilitation and outreach, and integration of USD's cyberinfrastructure resources. We present this project as an example for other research computing groups so that they may learn from our successes and the challenges that we have overcome in providing this user

resource. Additionally, this project serves to exemplify the importance of creating a broad user base of research computing infrastructure through the development of alternative user interfaces such as Science Gateways.

## CSS CONCEPTS

Computing methodologies→Simulation evaluation;  
•Software and its engineering→Software design engineering;

## KEYWORDS

Gateway, Apache Airavata, Keycloak, SciGaP

## ACM Reference format:

Joseph D. Madison<sup>1</sup>, Eroma Abeysinghe<sup>2</sup>, Sudhakar Pamidighantam<sup>3</sup>, Suresh Marru<sup>4</sup>, Marcus Christie<sup>5</sup>, Douglas M. Jennewein<sup>6</sup> and Marlon Pierce<sup>7</sup>. 2018. Science Gateway Implementation at the University of South Dakota: Applications in Research and Education. In *Proceedings of Practice and Experience in Advanced Research Computing (PEARC '18)*. ACM, New York, NY, USA, 6 pages.  
<https://doi.org/10.1145/3219104.3229265>

## 1 INTRODUCTION

The University of South Dakota Research Computing Group (USDRCG), in conjunction with the Science Gateways Community Institute, has developed a gateway, <https://sciencegateway.usd.edu/>, to facilitate access and use of

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

PEARC '18, July 22-26 2018, Pittsburgh, PA, USA

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6446-1/18/07.

<https://doi.org/10.1145/3219104.3229265>

the two high-performance computing clusters at the University of South Dakota (Legacy and Lawrence).

This gateway will be useful for introducing both students and research faculty to the capabilities of cluster computing without the need to know the intricacies of command-line task execution, allowing them instead to focus on the software that they want to run and to manage their interactions with system resources persistently.

## 2 GATEWAY DESIGN

### 2.1 USD Gateway with SciGaP & Apache Airavata

The USD gateway is hosted and maintained as a SciGaP [2] gateway. SciGaP is the hosted gateway platform that provides gateways as a service with Apache Airavata [3] middleware as the underlying framework. The SciGaP platform is multi-tenanted, supporting multiple gateways at the same time. SciGaP services include but are not limited to hosting the USD gateway, maintaining it, and providing storage for its user data. The USD gateway is a web based gateway with a graphical interface for HPC job submissions. The gateway has a flexible framework that enables development of customizable and accessible interfaces to the traditional command line applications prominent in HPC. This allows adaptability to the varying use cases that present themselves both at USD as well as to users at other universities in the state university network.

### 2.2 Gateway User Management

Gateway user authentication and authorization is maintained by Keycloak [4], <https://www.keycloak.org/>. The USD gateway provides multiple methods for user access. USD users can use their university account to gain access to the gateway through federated access provided by CILogon, <http://www.cilogon.org/>, through Keycloak. Users from other universities and organizations can either use their organizational accounts or create a local account in the gateway. Once the gateway account is created, gateway administrators can decide on the level of access to grant for each user. Users can be at different levels ranging from gateway administrators to gateway basic users or group leaders.

### 2.3 Gateway Userspace

In the USD gateway users create ‘experiments’ in order to submit computational jobs to the USD High Performance Computing (HPC) systems (Figure 1). The gateway supports both SLURM and Grid Engine job submissions and monitors jobs until they are completed. Users can submit the experiments from the gateway and view real time status updates of the jobs as well as view the results through the gateway portal. Downloading results and sharing them with other gateway users are features available for all of the gateway users. Cancelling running jobs on hosts and cloning existing jobs are also features supported through this gateway.

Current work is also focusing on providing USD HPC cluster users the option of using a ‘community’ account in addition to their own individual accounts for job submission. Community accounts allow gateway users to execute tasks as a generic gateway user. Additionally, users will be able to select which account to use for job submission at the time they create experiments.

### 2.4 Gateway Administration

The USD gateway provides a dedicated set of web interfaces for gateway administrators called the ‘Admin Dashboard’, which provides interfaces required for managing gateway configurations and users. The Admin Dashboard features include interfaces to manage users, add software tools, maintain credential store keys for secure SSH communications [5], and add HPC configuration information. Job statistics are also available through the Admin Dashboard which includes gateway usage summaries, user specific software utilization, and status of HPC jobs and failures. These statistics are essential for a gateway administrator in order to best serve the gateway user community in resolving their issues and advising them on their HPC usage. Experiment statistics can be listed based on a date and time period provided by the administrator and further filtered by username, HPC resource, and software. Gateway administrators can then view the number of FAILED, RUNNING or CANCELLED experiments at a given time and understand the status of the gateway user activities.



**Figure 1: (Left) USD Legacy supports general computational tasks. (Right) USD Lawrence supports general computational tasks as well as high-memory and GPU partition requests.**

### 2.5 Gateway Theme

The USD gateway uses a default gateway client provided by the SciGaP platform with localizations and other customizations provided by the USD team. The gateway theme is one of the customizations done by the USD team. The theme consists of details on USD HPCs, how to get access to the gateway, user documentation, and contact details. The overall look and feel is consistent with USD’s overall web presence and guidelines.

### 2.6 Outreach

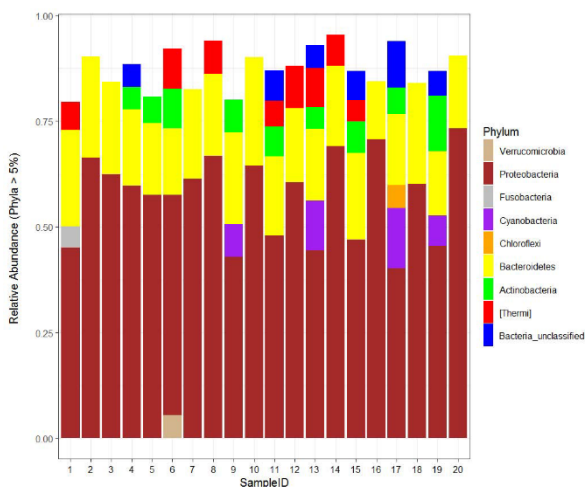
The USD gateway is currently focused on outreach events to establish the user base. Outreach is organized into two phases.

The first phase focuses on recruiting users from academic and research groups within USD, and the second phase is focused on intrastate university collaborations and research centers. Upcoming conferences will be used as the first outreach events for introducing the gateway to the state. Planned and ongoing outreach events are in the form of workshops, tutorials, webinars, and conferences. These events will include help from SciGaP gateway team members.

### 3 APPLICATIONS

The USD Science Gateway will be useful in two main respects, in line with the mission of the USDRCG. The first is to enhance the research capabilities of the institutional cyberinfrastructure; this includes both USD as well as other institutions in South Dakota. The second application that the Science Gateway will fill is in education and introducing cyberinfrastructure resources to new users.

#### 3.1 Research



**Figure 2: Bacterial community composition generated from operational taxonomic unit (OTU) tables (unpubl. data).**

Accessibility to the USD HPCs for research is an essential component of the overall cyberinfrastructure. The Science Gateway thus serves as an essential and primary source for user accessibility to USD’s HPC environment. The Science Gateway currently supports use of the following software in an HPC environment: Mothur, QIIME, NWChem, and Gaussian.

In order to test the efficacy of the USD Science Gateway with the applications of interest, we present an application use case with Mothur [6] for analyzing 16S rRNA gene sequencing data to elucidate bacterial community composition (Figure 2; OTU generation supported in the Science Gateway, figure created in R). The job specification page (Figure 3) and experiment summary (Figure 4) page for a Mothur experiment are given. Due to the large memory requirements of DNA sequence

alignment, we specify the job to use the ‘himem’ partition. We also have provided users the ability to choose the ‘preemptible’ (for general compute nodes) or ‘gpu’ (for GPU enabled nodes) partitions. Wall limits are also specified (120 minute limit).

**Figure 3: An example submission page for a Mothur experiment (alignment).**

#### 3.2 Education

Science Gateways provide a user-friendly interface for utilizing computational resources which traditionally require specialized training for access and use. For this reason, the USD Science Gateway is an important tool for introducing high-performance computing clusters in a classroom or workshop setting. In addition to current initiatives for use in laboratory level mentorship, this Gateway is planned to be used in coursework and workshops. Due to the multi institutional intent of this Science Gateway, use in educational settings across institutions, including regional workshops and classes taught at multiple institutions are planned.

#### 3.3 Intrastate Institutional Access

The USDRCG Science Gateway has been setup to allow multi-institutional access. USD currently operates many cyberinfrastructure resources not available at many smaller state universities. Due to this limitation and USD’s unique status as

one of the main state providers of research cyberinfrastructure, access through the Science Gateway has been made available to all public and private universities, and their collaborators, in South Dakota.

The screenshot shows the 'Experiment Summary' page for a Mothur experiment. The page is titled 'Experiment Summary' and includes a navigation bar with 'Project', 'Experiment', and 'Storage' tabs. The main content area displays the following information:

- Experiment ID:** Mothur\_microbiomealignment\_turtle\_1\_9d11ed8f-966e-4667-b026-00d246893f
- Name:** Mothur\_microbiomealignment\_turtle\_1
- Description:** Testing with Mothur
- Project:** Turtle-Microbiome
- Owner:** usderoma2017
- Application:** Mothur
- Compute Resource:** lawrence.usd.edu
- Experiment Status:** EXECUTING

A table below the main details shows job information:

Name	ID	Status	Creation Time
A1831281457	4590	QUEUED	05/06/2018, 1:08 AM - GMT-0400 (EDT)

Additional sections include 'Notifications To' with 'Creation Time' and 'Last Modified Time' both set to 05/06/2018, 1:08 AM - GMT-0400 (EDT). 'Wall Time' is 120, 'CPU Count' is 24, and 'Node Count' is 1. The 'Queue' is 'gpu'. 'Inputs' include 'Data File: turtle.htm.config.good.unique.good.filter.unique.precluster.pick.fasta' and 'Batch File: batch\_align.txt'. 'Outputs' state 'Experiment hasn't completed. Experiment Status is: EXECUTING'. 'Storage Directory' is 'Open'. 'Sharing Details' shows 'Order By: Username' and a message 'This has not been shared'.

Figure 4: Experiment Summary page for a Mothur experiment.

#### 4 FUTURE WORK AND CONCLUSIONS

While the USDRCG gateway has thus far been developed to provide an exceptional resource for new users, there remains work for its future enhancement. Providing access to licensed software through the gateway is an important future capability. This is due to some applications requiring additional permissions beyond the current access provided by basic gateway accounts. Currently the SciGaP team is working on extending its sharing mechanisms [7] to support user groups to the gateway. Access to a software or selected number of softwares can be provided to a set of users through a group. Through a group, the gateway will also be able to provide different levels of access to the HPC resources. In addition to the aforementioned future work, USDRCG is continuing to make more software applications available on the USD Science Gateway. Additionally, incorporation of the Science Gateway is planned for future workshops and training events, especially with an introductory focus to USD cyberinfrastructure. This gateway is therefore already proving useful in fulfilling a vital role for new and existing users at South Dakota's universities. Through continuing efforts, the USD Science Gateway will continue as a critical component of the overall cyberinfrastructure at USD.

#### ACKNOWLEDGEMENTS

The USD Research Computing Group Science Gateway was implemented with support from the Science Gateways Community Institute (NSF award #1547611). Development of the Apache Airavata used to develop the science gateway is supported by NSF award #1339774. The USD Lawrence cluster was obtained through a combination of state and federal funding: a FY16 SD Board of Regents Research and Development Innovation award, and National Science Foundation Major Research Instrumentation award #1626516. The USD Legacy cluster was obtained through an NIH Institutional Development award (#P20GM103443). Additional Legacy upgrades have been funded by both NIH and USD College of Arts & Sciences.

#### REFERENCES

- [1] 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), 4252-4268.
- [2] Pierce, M., Marru, S., Miller, M. A., Majumdar, A., and Demeler, B. (2013). *Science Gateway Operational Sustainability: Adopting a Platform-as-a Service Approach*.
- [3] Marru, S., Gunathilake, L., Herath, C., Tangchaisin, P., Pierce, M., Mattmann, C., Singh, R., Gunarathne, T., Chinthaka, E., Gardler, R., and Slominski, A. (2011). Apache airavata: a framework for distributed applications and computational workflows. In *Proceedings of the 2011 ACM Workshop on Gateway Computing Environments* (pp. 21-28). ACM.
- [4] Christie, M. A., Bhandar, A., Nakandala, S., Marru, S., Abeyasinghe, E., Pamidighantam, S., and Pierce, M. E. (2017). Using Keycloak for Gateway Authentication and Authorization.
- [5] Kanewala, T. A., Marru, S., Basney, J., and Pierce, M. (2014). A credential store for multi-tenant science gateways. In *Cluster, Cloud and Grid Computing (CCGrid)*, pp. 445-454. IEEE.
- [6] Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., and Sahl, J. W. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23), 7537-7541.
- [7] Nakandala, S., Marru, S., Piece, M., Pamidighantam, S., Yoshimoto, K., Schwartz, T., Sivagnanam, S., Majumdar, A., and Miller, M. A. (2017). Apache Airavata Sharing Service: A Tool for Enabling User Collaboration in Science Gateways. In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact* (p. 20). ACM.