

# GISandbox: A Science Gateway for Geospatial Computing

Eric Shook

Department of Geography, Environment, and Society,  
University of Minnesota  
Minneapolis, MN  
eshook@umn.edu

Andrea Zonca

San Diego Supercomputer Center  
La Jolla, CA  
zonca@sdsc.edu

Davide Del Vento

Computational and Information Systems Laboratory,  
National Center for Atmospheric Research  
Boulder, CO  
ddvento@ucar.edu

Jun Wang

Pervasive Technology Institute, Indiana University  
Bloomington, IN  
wang208@iu.edu

## ABSTRACT

Science gateways provide easy access to domain-specific tools and data. The field of Geographic Information Science and Systems (GIS) uses myriad tools and datasets, which raises challenges in designing a science gateway to meet users' diverse research and teaching needs. We describe a new science gateway called the GISandbox that is designed to meet the needs of researchers and educators leveraging geospatial computing, which is situated at the nexus of GIS and computational science. The GISandbox is built on Jupyter Notebooks to create an easy, open, and flexible platform for geospatial computing. Jupyter Notebooks is a widely used interactive computing environment running in the browser that integrates live code, narrative, equations and images. We extend the Jupyter Notebook platform to enable users to run interactive notebooks on the cloud resource Jetstream or computationally-intensive notebooks on the Bridges supercomputer located at the Pittsburgh Supercomputing Center. A novel Job Management platform allows the user to easily submit a Jupyter Notebook for batch execution on Bridges (and eventually Comet), monitor the SLURM job, and retrieve output files. GISandbox Virtual Machines are created in Jetstream's Atmosphere interface and then deployed and configured using a series of Ansible scripts, which allow us to create an easily reproducible and scalable system. This paper outlines our vision for GISandbox, the current implementation, with a discussion looking toward the future and how the GISandbox could be used in other domains.

## KEYWORDS

Geographic Information Systems, GIS, Science Gateway

### ACM Reference Format:

Eric Shook, Davide Del Vento, Andrea Zonca, and Jun Wang. 2018. GISandbox: A Science Gateway for Geospatial Computing. In *PEARC '18: Practice*

---

Permission to make digital or hard copies of all or part 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 components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*PEARC '18, July 22–26, 2018, Pittsburgh, PA, USA*

© 2018 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.

ACM ISBN 978-1-4503-6446-1/18/07...\$15.00

<https://doi.org/10.1145/3219104.3219150>

*and Experience in Advanced Research Computing, July 22–26, 2018, Pittsburgh, PA, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3219104.3219150>*

## 1 INTRODUCTION

The world is faced with many complex challenges in food, energy, water, and mass migration. Global satellites, environmental sensors, smart cities, and mobile technologies provide rich geospatial data capturing global and societal dynamics to help address complex global challenges [19]. Geographic Information Science and Technologies (GIS&T) encompass the science domains, conceptualizations, tools, and technologies for working with and exploiting these geospatial data. This paper introduces a new science gateway for the GIS&T community called the GISandbox. This new science gateway is designed to lower the barriers to entry to cyberinfrastructure through creating an interactive, scalable, and reproducible system. We overview the vision, implementation details, and applications for GISandbox.

The GIS&T community is poised to become major users and contributors of cyberinfrastructure (CI) [2, 30]. The University Consortium for Geographic Information Science (UCGIS) is the leading organization for GIS&T. The 2016 UCGIS Symposium reported the need for advanced CI to support GIS&T research including: 1) improving processing for big, spatial-temporal data in real-time; 2) leveraging new sources of high-fidelity spatial data; 3) enhancing GIS&T accessibility; and 4) enabling interdisciplinary research [27]. The 2017 UCGIS Symposium theme became CyberGIS and Diverse Partnerships for Geospatial Workforce Development.

CyberGIS is an interdisciplinary field that combines cyberinfrastructure, geographic information science, and spatial analysis and modeling to tackle complex geospatial problems [28, 29]. It also represents a next-generation GIS based on high-performance computing, data-driven knowledge discovery, visualization and visual analytics, and collaborative problem-solving [29]. CyberGIS has had an impact in several domains ranging from hydrology and spatial econometrics to hazards and public health by enabling cyberinfrastructure-based geospatial problem solving [14, 21]. CyberGIS community members can help lead the broader GIS&T community in the adoption and advancement of cyberinfrastructure.

The Extreme Science and Engineering Discovery Environment (XSEDE) is the most advanced, powerful, and robust collection of integrated advanced digital resources and services in the world

**Table 1: XSEDE projects, users, and utilization from Jan. 1 2017-Dec. 31 2017** [<https://xdmod.ccr.buffalo.edu>].

XSEDE Usage by NSF Directorate	Projects	Active Users	Percent Utilization
Mathematical & Physical Sciences	910	2,162	20.07
Biological Sciences	632	1,064	8.85
Computer & Information Science, & Engineering	383	1,782	1.66
Engineering	365	777	7.88
<b>Geosciences</b>	<b>220</b>	<b>466</b>	<b>4.78</b>
<b>Social, Behavioral, &amp; Economic Sciences</b>	<b>70</b>	<b>62</b>	<b>0.18</b>
Humanities/Arts	11	10	0.03
Other	112	1,371	0.24

[25]. XSEDE allocates free computational resources for research and education. Yet, emerging data- and compute-intensive scientific disciplines, also known as the long-tail of science [16], are not as active on XSEDE compared to more traditional HPC disciplines. Research projects in the Geosciences and Social, Behavioral, and Economic Sciences, for example, have fewer projects and active users on XSEDE (Table 1). XSEDE has been making great outreach efforts by starting the Novel and Innovative Programs (NIP) to seek out projects and work with communities that do not typically use CI. XSEDE also offers Extended Collaborative Support Services, which provides expert staff to assist research teams using large-scale computing resources [17]. These efforts will help long tail science communities like those from Social, Behavioral, and Economic Sciences and Geosciences fully benefit from CI. A complimentary approach is to lower the barriers to entry to cyberinfrastructure for these long tail science communities through discipline-specific platforms, tools, and science gateways. GIS&T is one of few scientific disciplines that intersects many domains in the long tail of science including social sciences, health sciences, environmental sciences, and geosciences. Providing an easy-to-use platform that lowers the barrier to entry to CI for GIS&T has the potential to impact many of these disciplines.

## 2 GISANDBOX VISION

The GISandbox is a science gateway hosted on XSEDE resources for researchers and educators to learn about, experiment with, and advance GIS [7]. A key challenge for the GISandbox is to address the varied needs of researchers and educators that use myriad tools, methodologies, and theoretical frameworks. Interactive computing [18, 20] is important for certain applications of GIS aimed at decision making or exploring multiple scenarios, for example. Whereas data-intensive computing may be important for processing satellite imagery data, for example. The user-interface, computational demands, and data demands in these two common scenarios requires a flexible architecture that balances interactivity with scalability and both of these features may be incorporated into a computational workflow to solve a geospatial problem.

A key goal for the GISandbox is to lower the barrier to entry to cyberinfrastructure for the GIS&T community. By creating a science gateway that provides the computational infrastructure without “boxing in” researchers or instructors to the confines of a “toolbox” of tools, the GISandbox allows researchers and educators

to explore how to fully exploit cyberinfrastructure to tackle geospatial problems. By lowering the barriers to entry we expect that the GISandbox will lead to new innovative and discovery in solving problems, developing methods, and eventually new theoretical understanding.

The GISandbox is being built with the expert assistance of ECSS through an XSEDE Startup allocation (XSEDE TG-SES170014). The Science Gateway Community Institute (SGCI, [13]) Incubator team is in the process of conducting a Cybersecurity audit for the design and implementation of GISandbox (Spring 2018) to ensure the GISandbox is not only scalable, but also secure. The GISandbox uses JupyterHub, which permits hosting Jupyter Notebooks as a multi-user service, and it runs on the XSEDE environment. Jupyter was modified to use CILogon, a X.509 certification authority for cyberinfrastructure [3], to enable users to login using XSEDE credentials. By leveraging the scalability of JupyterHub running on XSEDE, GISandbox is designed to support large audiences and computationally intensive lessons. Further implementation details are provided in Section 3.

### 2.1 Related Work

We review several related science gateways and projects. The most related project is the CyberGIS-Jupyter framework [31]. This framework is built on cloud computing technologies running on the first CyberGIS supercomputer called ROGER. While both GISandbox and the CyberGIS-Jupyter framework benefit from the interactivity of Jupyter Notebooks, the key differences between these two approaches include the ability for GISandbox users to submit notebooks to multiple supercomputers to enhance scalability and a reproducible deployment system, both of which are detailed in Section 3. The CyberGIS Gateway is a science gateway for cyberGIS, which supports several applications that intersect disease analytics using social media, topography related analytics, and bioenergy supply chain optimization, for example [14]. Another related science gateway is Terra Populus, now called IPUMS-Terra, which is focused on integrating and harmonizing social data with environmental data [10]. Users can access massive demographic and environmental datasets including a suite of tools to integrate them into a number of different output types. A broader approach than a single science gateway is MyGeoHub that posits the benefits of hosting multiple research projects on a single platform, which allows each project to share resources, common capabilities, and enjoy new features developed by other projects [11]. The platform is

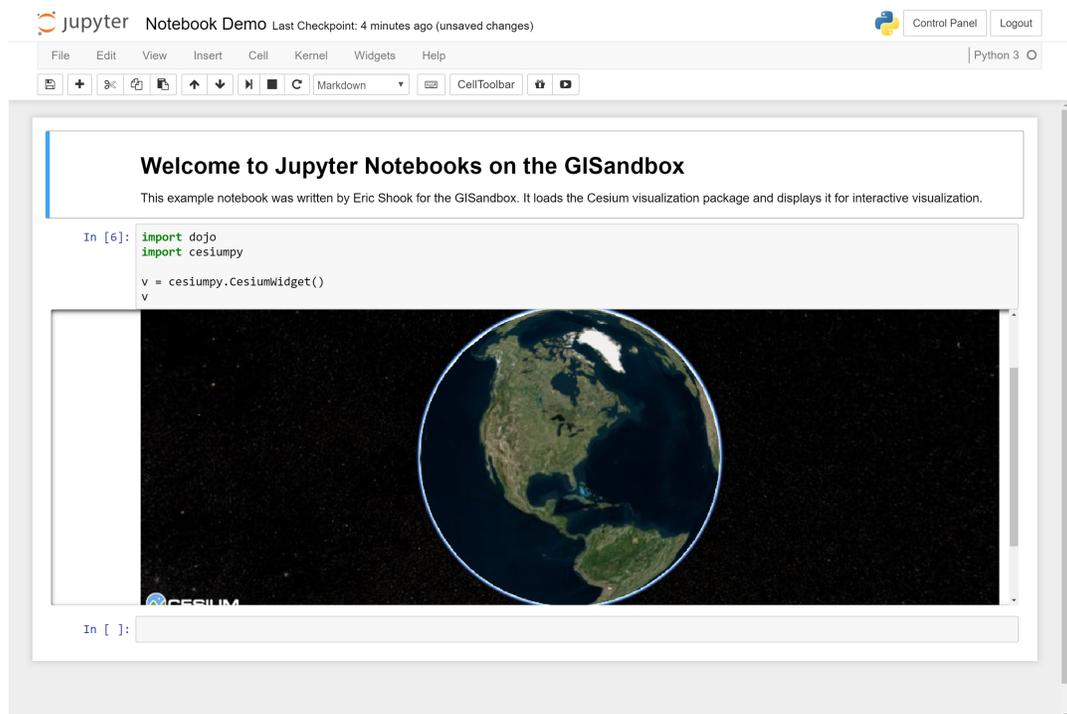


Figure 1: GISandbox Interface

based on the Hubzero cyberinfrastructure platform [15] and the sustainability model is worthy of consideration and could be explored in the future.

While there are excellent science gateways and platforms, currently no such gateway provides the *flexibility* of allowing users to write custom code to solve geospatial problems, the *interactivity* of combining code with visualizations through Jupyter Notebooks, the *scalability* for data-intensive applications, and a means of *reproducing* this computational environment as needs and technologies change in the future. The GISandbox fills this gap in capabilities for the GIS&T community to better leverage advanced cyberinfrastructure. The remainder of this manuscript details the implementation of the GISandbox followed by a brief discussion looking toward the future.

### 3 GISANDBOX IMPLEMENTATION

GISandbox integrates multiple components consisting of open-source and freely available software such as JupyterHub with custom-built software such as a notebook submission system in order to create a science gateway that will serve the diverse research and teaching needs of the GIS&T community. Software needs include analytical packages, modeling frameworks, and simulation toolkits to support data-intensive analytics and compute-intensive geospatial simulations. The GISandbox is in active development and the implementation is transitioning from prototype to production. We report the current status of the GISandbox and discuss future goals.

At the core of GISandbox is JupyterHub. From this core we add several modifications, extensions, and additional packages to improve the user experience and computational capabilities.

#### 3.1 Interactivity: JupyterHub and Jupyter Notebooks

The core user experience of the GISandbox implementation is a Jupyter Notebook (Figure 1). Jupyter Notebooks are web-based applications that support rich text, visualizations, and executable code in a single document [12]. The Jupyter platform has become synonymous with interactive computing and is used by thousands of data scientists around the world.

Access to the GISandbox platform is provided via a web-exposed JupyterHub VM. JupyterHub extends Jupyter Notebooks to support multiple users in a centralized deployment. Authentication for JupyterHub is delegated via OAuth [9] and CILogon [3] to XSEDE. In this way, XSEDE users will be able to immediately access the GISandbox. To provide the needed functionalities, we needed the ability for CILogon to delegate to XSEDE, passing the proper username. Originally, CILogon used email addresses instead of XSEDE usernames. So we first implemented the possibility to configure custom claims for the username field in the OAuth implementation. Moreover, other configuration settings could not be made as default settings to match our needs, which made the login process for our users more complicated. To resolve this issue, we modified the implementation to allow the possibility of setting such defaults, and we configured our instances of JupyterHub to use these defaults. This process streamlines the login process for our

users. Last, but not least, we worked with CILogon developers to have our changes merged upstream. At the time of writing, to take advantage of CILogon, Jupyter developers will either need to use the unreleased trunk of the software or to manually patch it. However, the next release of JupyterHub's OAuthenticator will include our merged changes, which will be immediately available for everybody. See <https://github.com/jupyterhub/oauthenticator/issues/170> for details.

GISandbox runs JupyterHub instances as Debian-based Virtual Machines (VMs) on the Jetstream cloud resource [23]. Jetstream compliments traditional supercomputing resources such as Bridges at the Pittsburgh Supercomputing Center by providing providing a semi-permanent instance and web-accessible IP address to host JupyterHub over the web at <http://play.gisandbox.org>.

A variety of Python packages are deployed to facilitate GIS research and education, including but not limited to `numpy`, `scipy`, `folium`, `shapely`, `matplotlib`, `pandas`, `netCDF4` for both Python v2 and Python v3. In addition to the Python kernel, we support the R kernel for Jupyter Notebooks and deploy relevant scientific packages such as `rgdal` and `raster`. We also plan to support NetLogo for agent-based modeling (ABM) research [24], which is commonly employed by GIS&T researchers. For example, one project is developing a Paleoscape Model and is using NetLogo ABMs to simulate hunter-gatherer behavior in order to better understand human evolution and human origins [6, 22]. This project is planning to use the interactivity and scalability of GISandbox to run simulation models and manage output data.

### 3.2 Scalability: Batch Execution of Jupyter Notebooks

To support data-intensive problem solving, GISandbox allows users to run their Jupyter Notebook codes in parallel on Bridges, a supercomputer hosted at the Pittsburgh Supercomputing Center (PSC). We also plan to support running codes on Comet (at the San Diego Supercomputing Center) and Wrangler (at the Texas Advanced Computing Center). While interactive computing using Jupyter Notebooks is a critical feature to lower the barriers to entry, running Notebooks on a shared VM limits the amount of memory and number of cores available for data-intensive computations. We expand the capabilities of Jupyter Notebooks on Jetstream through a new package that allows batch execution of Notebooks without leaving the Jupyter Notebook interface.

As part of GISandbox, we have developed the `nsubmit` package to provide a simplified interface to leverage remote supercomputing resources for batch non-interactive execution of Jupyter Notebooks (<https://github.com/zonca/nsubmit>). Scientists will be able to utilize the GISandbox interactive Jupyter Notebooks environment to analyze data and prototype their algorithms. Once their interactive exploration is complete, they can execute their data analysis pipeline either (1) on larger datasets or (2) repeat it on hundreds of input files in parallel.

In the first scenario, the user can prototype their analysis on a smaller dataset on the limited memory resources on Jetstream. Once they are ready to analyze the full dataset they can leverage the `nsubmit` package to automatically copy the Jupyter Notebook to Bridges or Comet and execute it non-interactively on a Computing

node. The interface of `nsubmit` doesn't require any knowledge on how to submit jobs to a scheduler like SLURM. See for example how to submit a prototype Notebook (referred to as `job.ipynb` in this example) for remote execution on Comet (Code Example 1). Users will soon interact with `nsubmit` through a Jupyter Lab extension, which will enable them to monitor job status, kill a job, or hold a job without leaving the Notebook interface. This interface will replace the need to launch the notebook manually as illustrated in Code Example 1.

---

#### Code Example 1 Simple notebook launch example

---

```
from nsubmit import cluster
comet = cluster.get("comet")
comet.launch_job("job_title", "job.ipynb", hours=10)
```

---

In the background, the package uses a multiplexed SSH connection (so authentication needs to happen only once) to copy the Notebook to Bridges or Comet, create a SLURM job file and submit it for execution on the scheduler (Figure 2 outlines the overall architecture using Bridges as an example). GISandbox requires users to authenticate to the remote supercomputer once and uses multiple channels of a single multiplexed connection for job and data management and coordination. Currently, authentication happens in a Jupyter terminal at the start of a work session, but we are evaluating incorporating authentication into the Jupyter Lab interface. The `nsubmit` package assumes that each user has already access to a XSEDE allocation (either as part of the GISandbox or their own allocation) where they can charge their computational hours. The user can then check job status and access the result on the filesystem mounted locally via SSHFS. The `nsubmit` package also supports mounting remote storage volume from Comet or Bridges locally in the home folder of each user's home folder on Jetstream. We are working on simplifying this further by providing a Jupyter Lab that allows the user to just click on a button, choose parameters to create a remote job, and monitor the job status in a separate panel.

The second scenario targets advanced users wanting to repeat the same analysis on tens or hundreds of different input files (e.g., parameter sweeping or tiles of remote sensing data). Here the user can parametrize the Jupyter Notebooks using environmental variables and then submit remote jobs by customizing the environment for every job submission (Code Example 2). All of the jobs will be submitted to the target Supercomputer where they would often run concurrently to quickly process all of the input files. In this case, users will need to manually launch jobs using the `nsubmit` package to provide finer-grained control compared to the single job launch through a Jupyter Lab interface in the first scenario. Examples for single and multiple job submission are available at the `nsubmit` github page.

The Jupyter Notebook is executed inside a Singularity container on the Supercomputer so that we can ensure that the code in the Notebook has access to the same software environment installed on the GISandbox interactive Jupyter Notebooks on Jetstream.

We rely on `sshfs` to mount the remote file systems in user-space. XSEDE resources are interconnected with large bandwidth and low latency networks. Our testing showed that metadata operations like browsing the file system and listing files are so quick they are

---

**Code Example 2** Example parameter sweeping job submission using the nbsubmit package.

---

```
for i in range(10):
    comet.launch_job("job_run_{}".format(i), "job.ipynb",
                    environment={"INPUTFILE":"input_data_{}.h5".format(i)})
```

---

not distinguishable from local disks. Although we anticipate that data-intensive operations will occur on Bridges or Comet supercomputers where the data reside on local parallel file systems to maintain top-level performance.

The purpose of this feature is to make it easy and seamless for the scientists to switch between batch processing on the Supercomputer and interactive data analysis on the Jupyter Notebook. The users will be able to execute (light) interactive data processing on the Jupyter Notebooks on the GISandbox, and once the analysis prototype is finalized, they can submit it on a Supercomputer where they can exploit larger memory, more CPU cores, and faster disk access with more storage capacity.

### 3.3 Reproducibility: Automated Deployment using Ansible

Creating a science gateway is a labor intensive process that often involves customized configurations to create a usable and scalable system for a particular science domain, which can be difficult to reproduce. To alleviate this challenge, the GISandbox is automatically deployed using Ansible, an effective provisioning system [1]. The server side deployment is almost completely automatized, to guarantee reproducibility and consistency of the deployments and re-deployments. The only steps that are not automatic are: The generation of the github deploy keys for the automatic log backup described below. The install and renewal of the CA SSL/TLS certificates, because it has some possible failure modes better addressed manually. Moreover, at time of writing we are working with XSEDE experts to automate the deployment of security patches using the OS updates repository.

A variety of small and standard, but important, details are also taken care of, for example we configure our servers to default the communication to https with SSL/TLS Certificates provided by an official Certificate Authority, and we manage the DNS via DynDNS (Dynamic Domain Name System) to make sure our service is visible to our users with a consistent URL, despite the possible changes to the underlying IP addresses provided by the Jetstream cloud provider.

All the server logs are automatically pushed on private repositories on github, easily allowing detailed post-mortem in case of catastrophic server crashes or security incidents where the validity of the logs on the servers may not be trusted anymore. Obviously one always hope that such incidents will never happen (and in fact they have not happened so far), but we are convinced that it is important to have the infrastructure in place to be ready. The github configuration step is not automatically handled by Ansible, because it requires manual intervention in the github GUI, which requires manual intervention. Pushing up the logs to github has the advantage of being simple (once properly configured with deploy keys, it is just a git add + git push command in a cron job),

fast and efficient (git automatically packs only the differences, and compresses them).

## 4 CONCLUSION

The GISandbox is a new science gateway for the GIS&T community. The design of GISandbox has been centered on three principles: interactivity, scalability, and reproducibility. We use Jupyter Notebooks as an easy interface that supports interactive computing. We developed nbsubmit to enable GISandbox users to leverage supercomputers to execute their Notebooks on large and many datasets, which enables scalable computing when problems exceed the limited computational capabilities of interactive Jupyter Notebooks running on a VM. GISandbox is built and deployed using Ansible scripts to create a reproducible environment that can be built and rebuilt on-demand.

There are numerous applications and uses for the GISandbox. We highlight several example uses for the GISandbox. Some of these examples are being actively piloted and others are planned pilots for summer 2018. First, is executing NetLogo ABMs for geospatial research. The Paleoscape Model project is planning on running potentially hundreds of thousands of scenarios and simulations to better understand hunter-gatherer behaviors given different climatic and environmental conditions [22]. The ABMs are part of a larger computational workflow that can be constructed and executed on the GISandbox. Second, is remote sensing and satellite imagery data analytics. This area of research tends to apply computational processes across enormous numbers of two-dimensional raster tiles. Sentinel-2 is a two satellite Earth observation mission led by the European Space Agency that produces 13 band multi-spectral data at spatial resolutions of 10m, 20m, and 60m [5]. Sentinel-2 satellites revisit the same location every 5 days (sometimes even more frequently) producing terabytes of spatiotemporal data that can be used for Land Use and Land Cover Change (LULCC) research among other areas [26]. This type of research matches the design of GISandbox quite well, because the analytical methods and workflows can be designed using a single tile or a small subset of tiles. Developers can use the interactivity of notebooks to rapid prototype and quickly visualize the output to improve development productivity. Once the analytical procedures are developed and tested they can scaled to all the tiles using supercomputers through nbsubmit. A project is underway to use satellite imagery data to better understand and map agriculture around the world, which will require advanced cyberinfrastructure to handle the complex analytics and big spatial data demands. Third, is analyzing user-contributed and user-generated content in the form of volunteered geographic information (VGI) such as OpenStreetMap data [8] or social media data from platforms such as Twitter [21]. Fourth, an emerging area for GIS&T is the application of machine learning and deep learning to geospatial data. A popular application in this area is image classification using satellite imagery data, but research

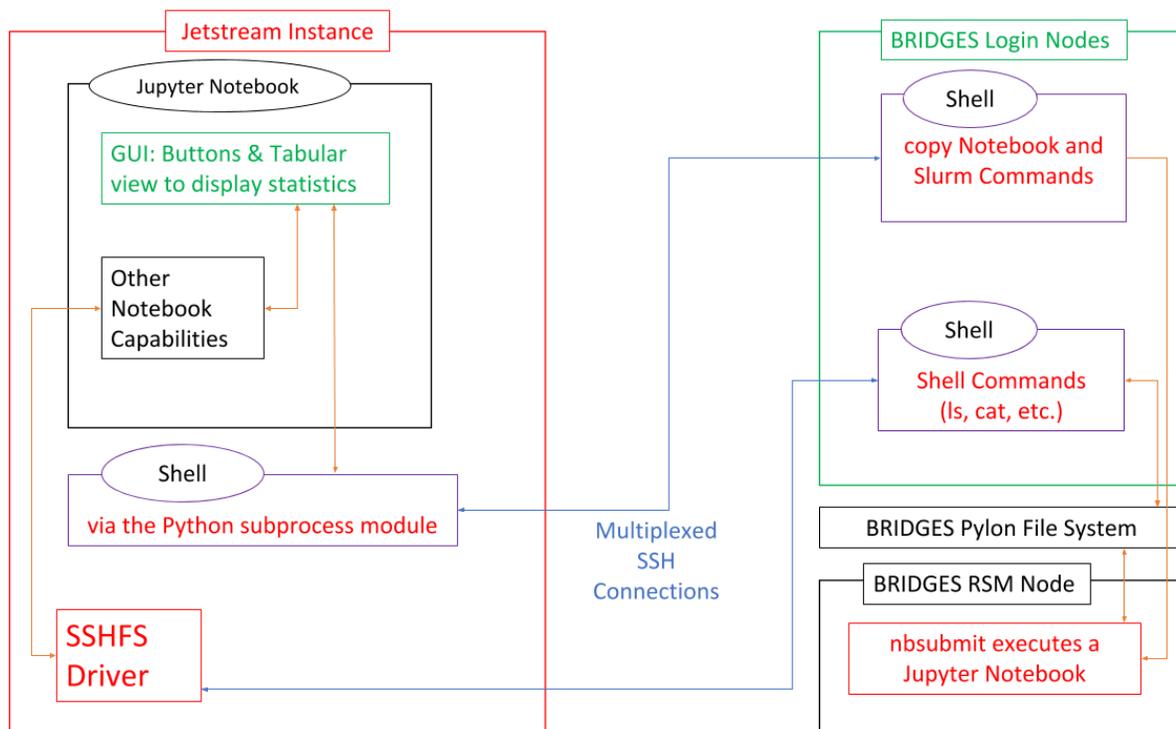


Figure 2: GISandbox Architecture

is still in the early stages and we are bound to see more exciting discoveries in the rapidly evolving space. The GISandbox will be installing Google TensorFlow in the future to support these latest research needs. We are in the early exploration phases in this area with a pilot project using the Functional Map of the World data [4]. Finally, a collaborative group in the GIS&T community is working to build curriculum materials that could use the GISandbox as a platform for classroom activities and lab exercises.

The development for GISandbox continues. It is built on free and open source software and improvements from this project have been pushed upstream to their respective code repositories. We aim to continue to improve the usability and utility of GISandbox. We plan to explore the Interactive Computing Protocol available as part of JupyterHub to add a third option situated between interactive Notebooks and the nbsubmit package. As we transition from prototype to production we continue to invite project partners to use and test the system. In the future we plan to write introductory documentation to help users transition from desktop to GISandbox and the proper use of high-performance computing, which will likely include pointing to existing documentation such as CI-Tutor (citur.org) or Cornell Virtual Workshops (<https://cvw.cac.cornell.edu/>). Ultimately, we aim for the GISandbox to become a common platform for GIScience research and education and we continue to work toward our ambitious goal.

## ACKNOWLEDGMENTS

The authors would like to thank Sergiu Sanielevici and Nancy Wilkins-Diehr for their guidance during the initial formulation of the GISandbox. The authors would like to thank Mark Krenz and the Science Gateway Community Institute for his continued assistance in a cybersecurity audit as part of a Science Gateway Community Institute Incubator project. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.

## REFERENCES

- [1] Ansible. 2018. Ansible is Simple IT Automation. <https://www.ansible.com/>. (2018).
- [2] Marc P Armstrong. 2000. Geography and computational science. *Annals of the Association of American Geographers* 90, 1 (2000), 146–156.
- [3] Jim Basney, Terry Fleury, and Jeff Gaynor. 2014. CILogon: A federated X.509 certification authority for cyberinfrastructure logon. *Concurrency and Computation: Practice and Experience* 26, 13 (2014), 2225–2239.
- [4] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. 2017. Functional Map of the World. *arXiv preprint arXiv:1711.07846* (2017).
- [5] M Drusch, U Del Bello, S Carlier, O Colin, V Fernandez, F Gascon, B Hoersch, C Isola, P Laberinti, P Martimort, et al. 2012. Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. *Remote sensing of Environment* 120 (2012), 25–36.
- [6] Erich C Fisher, Miryam Bar-Matthews, Antonietta Jerardino, and Curtis W Marean. 2010. Middle and Late Pleistocene paleoscape modeling along the southern coast of South Africa. *Quaternary Science Reviews* 29, 11-12 (2010), 1382–1398.
- [7] GISandbox. 2018. GISandbox Homepage. <http://gisandbox.org>. (2018).
- [8] Mordechai Haklay and Patrick Weber. 2008. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing* 7, 4 (2008), 12–18.
- [9] Dick Hardt. 2012. The OAuth 2.0 authorization framework. <https://tools.ietf.org/html/rfc6749>. (2012).

- [10] David Haynes, Steve Manson, and Eric Shook. 2017. Terra Populus: architecture for integrated big geospatial services. *Transactions in GIS* 21, 3 (2017), 546–559.
- [11] Rajesh Kalyanam, Lan Zhao, Carol Song, Larry Biehl, Derrick Kearney, I Luk Kim, Jaewoo Shin, Nelson Villoria, and Venkatesh Merwade. 2018. MyGeoHub: A sustainable and evolving geospatial science gateway. *Future Generation Computer Systems* (2018).
- [12] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, Sylvain Corlay, et al. 2016. Jupyter Notebooks—a publishing format for reproducible computational workflows.. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. 87–90. <https://doi.org/10.3233/978-1-61499-649-1-87>
- [13] Katherine A Lawrence, Michael Zentner, Nancy Wilkins-Diehr, Julie A Wernert, Marlon Pierce, Suresh Marru, and Scott Michael. 2015. Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community. *Concurrency and Computation: Practice and Experience* 27, 16 (2015), 4252–4268.
- [14] Yan Liu, Anand Padmanabhan, and Shaowen Wang. 2015. CyberGIS Gateway for enabling data-rich geospatial research and education. *Concurrency and Computation: Practice and Experience* 27, 2 (2015), 395–407.
- [15] Michael McLennan and Rick Kennell. 2010. HUBzero: a platform for dissemination and collaboration in computational science and engineering. *Computing in Science & Engineering* 12, 2 (2010).
- [16] Richard L Moore, Chaitan Baru, Diane Baxter, Geoffrey C Fox, Amit Majumdar, Phillip Papadopoulos, Wayne Pfeiffer, Robert S Sinkovits, Shawn Strande, Mahidhar Tatineni, et al. 2014. Gateways to discovery: Cyberinfrastructure for the long tail of science. In *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*. ACM, 39.
- [17] Nancy Wilkins-Diehr, Sergiu Sanielevici, Jay Alameda, John Cazes, Lonnie Crosby, Marlon Pierce, and Ralph Roskies. 2016. An overview of the XSEDE extended collaborative support program. In *High Performance Computer Applications - 6th International Conference, ISUM 2015, Revised Selected Papers (Communications in Computer and Information Science)*, Vol. 595. Springer Verlag, Germany, 3–13. [https://doi.org/10.1007/978-3-319-32243-8\\_1](https://doi.org/10.1007/978-3-319-32243-8_1)
- [18] Fernando Pérez and Brian E Granger. 2007. IPython: a system for interactive scientific computing. *Computing in Science & Engineering* 9, 3 (2007).
- [19] Shashi Shekhar, Steven K Feiner, and Walid G Aref. 2015. Spatial computing. *Commun. ACM* 59, 1 (2015), 72–81.
- [20] Helen Shen. 2014. Interactive notebooks: Sharing the code. *Nature News* 515, 7525 (2014), 151.
- [21] Eric Shook and Victoria K Turner. 2016. The socio-environmental data explorer (SEDE): a social media-enhanced decision support system to explore risk perception to hazard events. *Cartography and Geographic Information Science* 43, 5 (2016), 427–441.
- [22] Eric Shook, Colin Wren, Curtis W Marean, Alastair J Potts, Janet Franklin, Francois Engelbrecht, David O'Neal, Marco Janssen, Erich Fisher, Kim Hill, et al. 2015. Paleoscape model of coastal South Africa during modern human origins: progress in scaling and coupling climate, vegetation, and agent-based models on XSEDE. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*. ACM, 2.
- [23] Craig A Stewart, David Y Hancock, Matthew Vaughn, Jeremy Fischer, Tim Cockerill, Lee Liming, Nirav Merchant, Therese Miller, John Michael Lowe, Daniel C Stanzione, et al. 2016. Jetstream: performance, early experiences, and early results. In *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale*. ACM, 22.
- [24] Seth Tisue and Uri Wilensky. 2004. Netlogo: A simple environment for modeling complexity. In *International Conference on Complex Systems*, Vol. 21. Boston, MA, 16–21.
- [25] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D Peterson, et al. 2014. XSEDE: accelerating scientific discovery. *Computing in Science & Engineering* 16, 5 (2014), 62–74.
- [26] BL Turner, William B Meyer, David L Skole, et al. 1994. Global land-use/land-cover change: towards an integrated study. *Ambio. Stockholm* 23, 1 (1994), 91–95.
- [27] UCGIS. 2016. Report from the 2016 UCGIS Symposium. [http://ucgis.org/sites/default/files/document-sharing-form-files/209/UCGIS\\_Symposium\\_2016\\_Final\\_Report.pdf](http://ucgis.org/sites/default/files/document-sharing-form-files/209/UCGIS_Symposium_2016_Final_Report.pdf). (2016).
- [28] Shaowen Wang. 2010. A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers* 100, 3 (2010), 535–557.
- [29] Shaowen Wang, Luc Anselin, Budhendra Bhaduri, Christopher Crosby, Michael F Goodchild, Yan Liu, and Timothy L Nyerges. 2013. CyberGIS software: a synthetic review and integration roadmap. *International Journal of Geographical Information Science* 27, 11 (2013), 2122–2145.
- [30] Chaowei Yang, Robert Raskin, Michael Goodchild, and Mark Gahegan. 2010. Geospatial cyberinfrastructure: past, present and future. *Computers, Environment and Urban Systems* 34, 4 (2010), 264–277.
- [31] Dandong Yin, Yan Liu, Anand Padmanabhan, Jeff Terstriep, Johnathan Rush, and Shaowen Wang. 2017. A CyberGIS-Jupyter Framework for Geospatial Analytics at Scale. In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*. ACM, 18.