CyberGISX for Reproducible Geospatial Research and Education

Anand Padmanabhan

CyberGIS Center for Advanced Digital and Spatial Studies
CyberInfrastructure and Geospatial Information Laboratory
Department of Geography and Geographic Information Science
University of Illinois at Urbana-Champaign

Geospatial Fellows
Urbana, Illinois
Oct 6, 2020

Outline

- Overview

- Behind the scenes

- User Interface

  - Registering for CyberGISX

  - Logging into CyberGISX

  - Editing your profile page

  - Submitting feedback and reporting problems

- Demo

WILEY

RESEARCH ARTICLE

CyberGIS-Jupyter for reproducible and scalable geospatial analytics

Dandong Yin[1,2] | Yan Liu[1,2,3] | Hao Hu[1,2,3] | Jeff Terstriep[3] | Xingchen Hong[3] | Anand Padmanabhan[1,2] | Shaowen Wang[1,2,3]

[1]Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Illinois
[2]CyberGIS Center for Advanced Digital and Spatial Studies, University of Illinois at Urbana-Champaign, Illinois
[3]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Illinois

Correspondence
Shaowen Wang, Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Illinois; or CyberGIS Center for Advanced Digital and Spatial Studies, University of Illinois at Urbana-Champaign, Illinois.
Email: shaowen@illinois.edu

Present Address
Department of Geography and Geographic Information Science Natural History Building 1301 W Green Street Champaign, IL 61801 USA

Summary

The interdisciplinary field of cyberGIS (geographic information science and systems (GIS) based on advanced cyberinfrastructure) has a major focus on data- and computation-intensive geospatial analytics. The rapidly growing needs across many application and science domains for such analytics based on disparate geospatial big data poses significant challenges to conventional GIS approaches. This paper describes CyberGIS-Jupyter, an innovative cyberGIS framework for achieving data-intensive, reproducible, and scalable geospatial analytics using Jupyter Notebook based on ROGER, the first cyberGIS supercomputer. The framework adapts the Notebook with built-in cyberGIS capabilities to accelerate gateway application development and sharing while associated data, analytics, and workflow runtime environments are encapsulated into application packages that can be elastically reproduced through cloud-computing approaches. As a desirable outcome, data-intensive and scalable geospatial analytics can be efficiently developed and improved and seamlessly reproduced among multidisciplinary users in a novel cyberGIS science gateway environment.

KEYWORDS

cloud computing, computational reproducibility, cyberGIS, geospatial big data

Yin, D., Liu, Y., Padmanabhan, A., Terstriep, J., Rush, J. and Wang, S. (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* (p. 18). ACM.
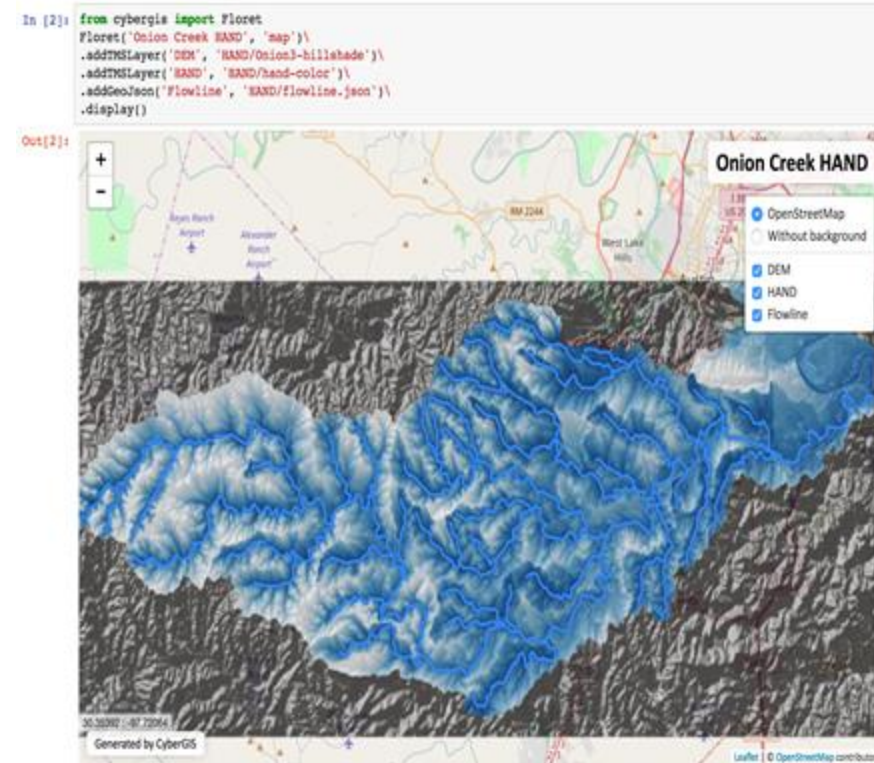
Purposes

- Achieves data-intensive, reproducible, and scalable geospatial analytics using Jupyter Notebooks
  - Provides a holistic solution
  - Makes sharing code and workflows easy
- Reduces the barrier to accessing advanced cyberinfrastructure and cyberGIS capabilities
  - Exploits JupyterHub, cloud, and high performance computing resources

Capabilities

- Provides notebook servers with cyberGIS libraries and many geospatial software packages installed
  - Built-in cyberGIS capabilities to accelerate gateway application development
    - E.g., HAND Application
  - Data, analytics, and workflow runtime environments are encapsulated into application packages
- Deployment can be elastically scaled to accommodate the computational needs of cyberGIS users
  - Straightforward management and maintenance of computational infrastructure
  - Seamless scaling between Virtual ROGER and XSEDE JetStream

CyberGISX: Geovisualization
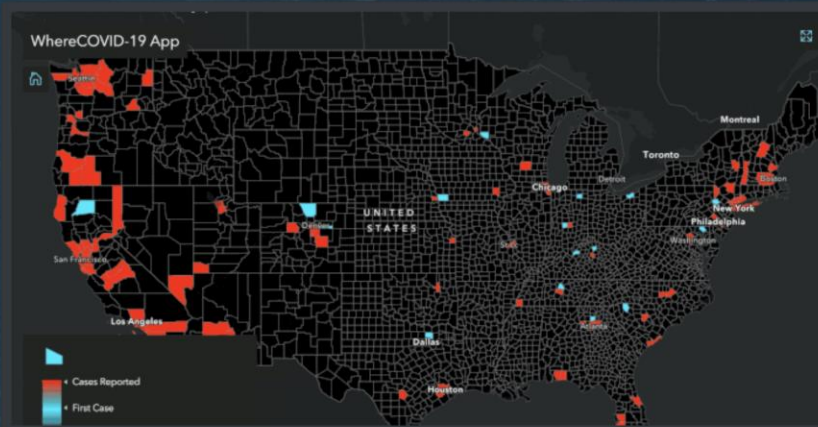
- Interactive map generation inside notebooks

- Support multiple formats of geospatial data

- Layer management, transparency, and styles

CyberGIS Center for Advanced Digital and Spatial Studies

http://cybergisx.org

https://cybergisx.cigi.illinois.edu/hub/login

Behind the Scenes

- Docker
- DockerSwarm
- JupyterHub

Technologies

Batch Job management

Cloud-based architecture

Data/storage synchronization

Bridging Local Cyberinfrastructure and XSEDE with CyberGIS-Jupyter

[Extended Abstract]

Anand Padmanabhan
Department of Geography and
Geographic Information Science
University of Illinois of Urbana-
Champaign
apadmana@illinois.edu

Dandong Yin
Department of Geography and
Geographic Information Science
University of Illinois of Urbana-
Champaign
dyin4@illinois.edu

Fangzheng Lyu
Department of Geography and
Geographic Information Science
University of Illinois of Urbana-
Champaign
flu8@illinois.edu

Shaowen Wang
Department of Geography and
Geographic Information Science
University of Illinois of Urbana-
Champaign
shaowen@illinois.edu

INTRODUCTION

The fabric of national and international cyberinfrastructure ecosystems for scientific discovery and innovation can be viewed as distributed computing environments composed of powerful supercomputers, various cloud computing resources, and numerous local cyberinfrastructure (including both cloud and HPC) resources. However, extensive computational work conducted by academic researchers are often siloed in one of these environments. Science Gateways [5, 6], by simplifying access to advanced cyberinfrastructure resources, have made significant progress on connecting these silos by enabling researchers in many fields to access advanced cyberinfrastructure through web browsers. In this context, this research bridges between national and local cyberinfrastructure resources through: (a) horizontal scaling of CyberGIS-Jupyter between the cloud resources provided by JetStream on the Extreme Science and Engineering Discovery Environment (XSEDE) and a VMWare-based cloud environment on Virtual ROGER, a local cyberinfrastructure resource hosted by the CyberGIS Center for Advanced Digital and Spatial Studies at the University of Illinois at Urbana-Champaign campus; and (b) enabling the submission of computationally intensive models to the batch systems of both Virtual ROGER and Comet (an XSEDE resource).

Specifically, we have developed a mechanism to provide access to a scalable JupyterHub platform together with cutting-edge cyberGIS software and hardware [4], called CyberGIS-Jupyter [7], which allows seamless access to high-performance computing (HPC) resources while shielding the complexity of managing cyberinfrastructure access from users. The user-friendly environment provided by CyberGIS-Jupyter along with computational scalability achieved through this research provides a powerful environment for conducting collaborative and reproducible research at scale with seamless access to advanced cyberinfrastructure at both local and national levels. The rest of this paper describes the architecture of our solution and articulates the corresponding implementation.

ARCHITECTURE AND IMPLEMENTATION

The architecture consists of four major layers: (a) user layer; (b) application layer; (c) cloud resources; and (d) HPC resources. The components of each of these layers and the interactions between them can be depicted in Figure 1. There are two levels at which users interact with cyberinfrastructure resources: (1) logging into JupyterHub and accessing their single user Jupyter Notebook server as an interactive session which runs transparently on local and national resources, and (2) submitting computationally intensive jobs that leverage HPC resources locally and on XSEDE.

Padmanabhan, A., Yin, D., Lyu, F., Wang, S. (2019) "Bridging Local Cyberinfrastructure and XSEDE with CyberGIS-Jupyter". In: *Proceedings of Practice and Experience in Advanced Research Computing (PEARC19)*, July 28-August 1, 2019, Chicago, IL, USA.

CyberGISX User Interface
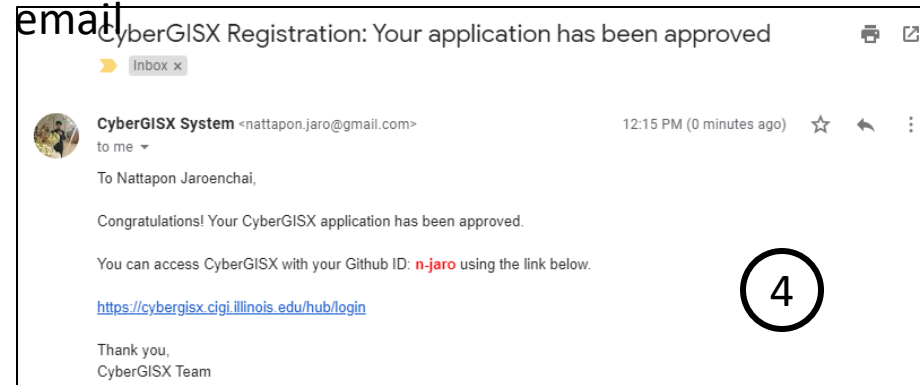
Registration part 1

1. Go to the "Registration" page on the menu bar

2. Fill out the registration form

Registration part 2

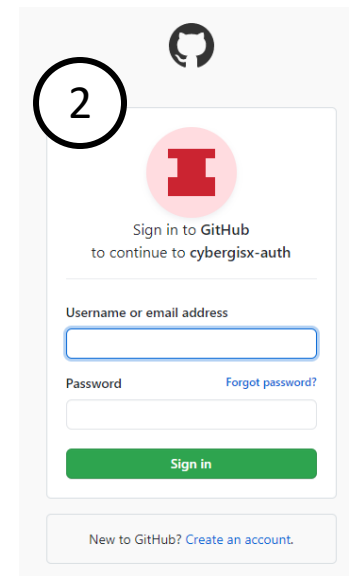Verify your email by clicking the link

3
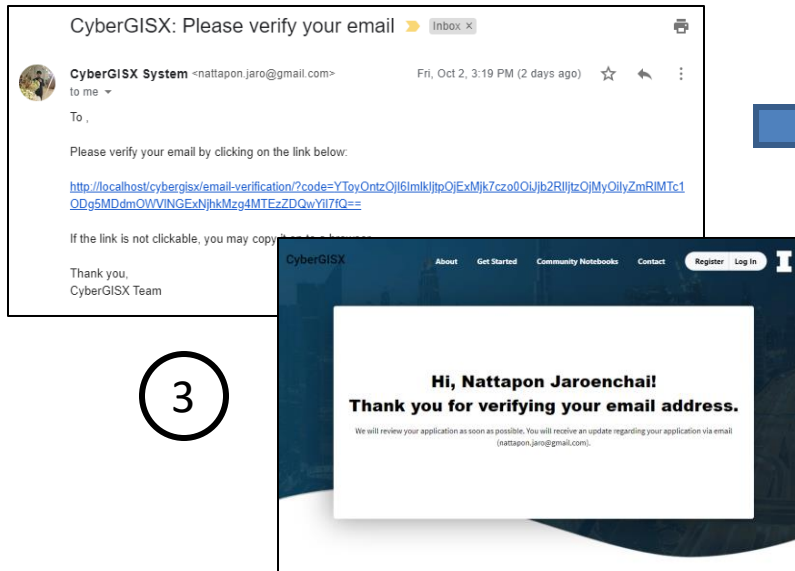
When you are approved, you will get an approval email

4

3. Verify your email address

4. Wait for approval email

Logging in part 1

Logging in part 2

There are 3 options:
1. Click the username to see your profile
2. Click "Edit Profile" to edit your profile page
3. Click "Log Out" to log out from this username
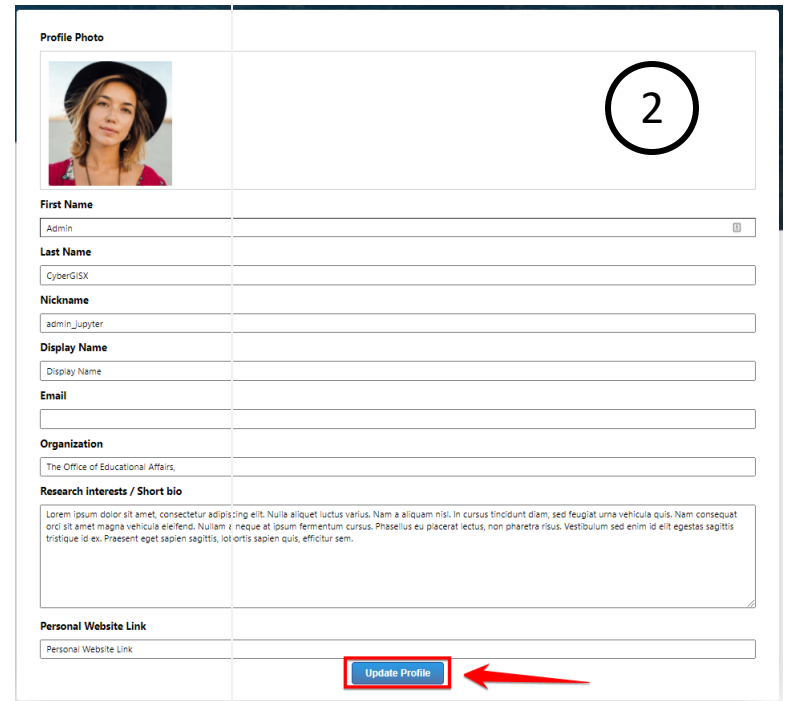
Hover over your username to see more menu items

3

Editing your profile page

1

After logging in, click "Edit Profile". You then will be redirected to the profile edit page.

2

Edit your information and then click "Update Profile" to save your progress.
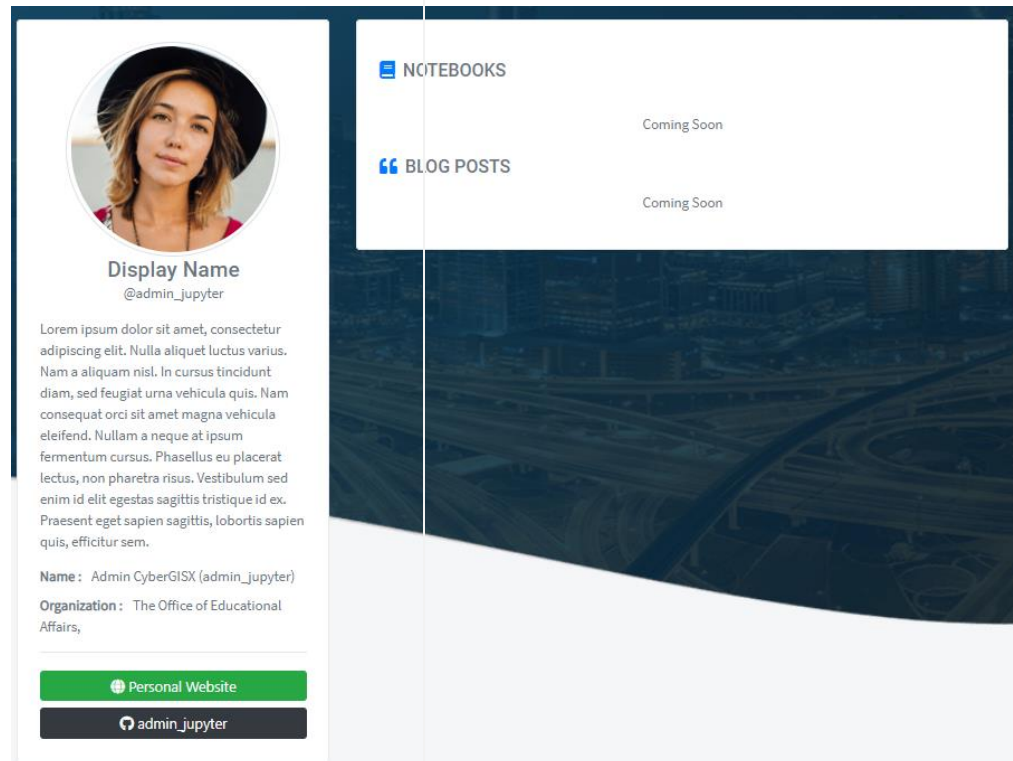
Recommended profile image size: 300x300.

About your profile page

Your profile page is public. You can share it anywhere and anyone can access your profile page.

The link to your profile page is:
http://cyberigsx.org/user-profile/{username}

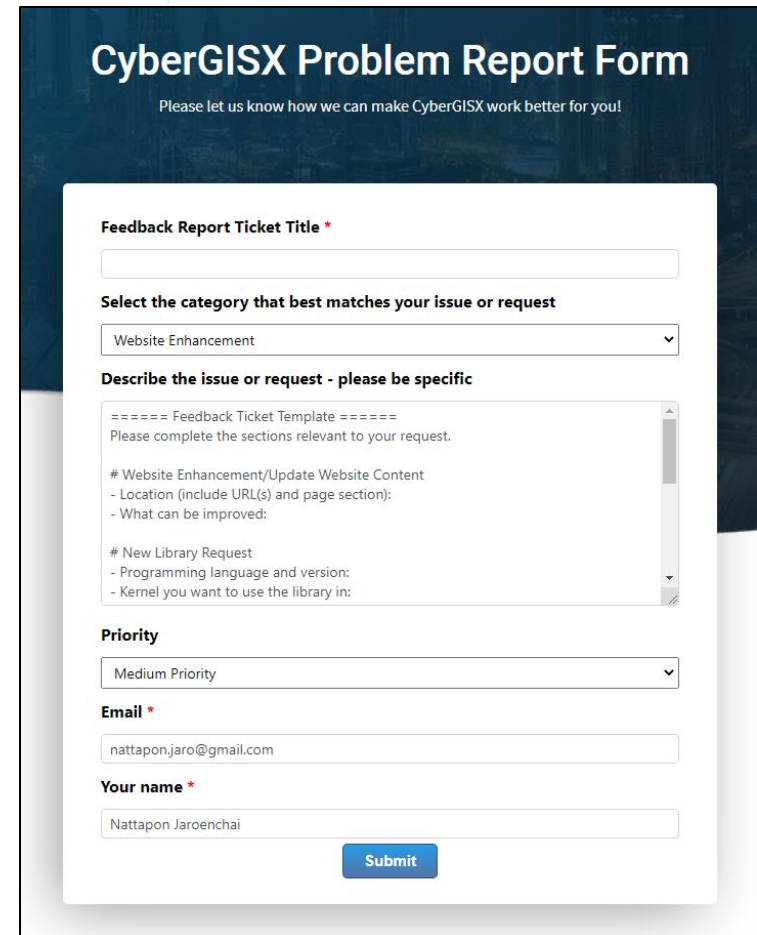**Note:** Notebooks and Blog post profile page sections are still under development.

Submitting feedback and reporting problems

If you have feedback, want to submit a request, or find a problem, you can report it by going to "Contact > Problem Report" on the CyberGISX main menu.

We have provided a template in the description field. Please fill as much information as you can in the template section that corresponds to your request or issue.

Then click "Submit".

Demo – Using CyberGISX

Acknowledgments

Questions?

Contact

apadmana@Illinois.edu