Webinar: Conceptualizing a National Geospatial Software Institute

NSF SI2-S2I2 Conceptualization: Geospatial Software Institute (GSI)

Tuesday, December 17th: 3 – 4pm CT
Toward a sustainable social and technical ecosystem
to enable geospatial-inspired discovery and innovation

Steering Committee

Donna Cox
National Center for Supercomputing Applications/University of Illinois
Co-PI

Michael F. Goodchild
University of California - Santa Barbara
Chair of the advisory committee

Daniel S. Katz
National Center for Supercomputing Applications
Co-PI

Paul Morin
University of Minnesota
Co-PI

Margaret Palmer
University of Maryland
Co-PI

Anand Padmanabhan
University of Illinois at Urbana-Champaign
Project Manager

Shaowen Wang
University of Illinois at Urbana-Champaign
Chair, PI
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Advisory Committee

Luc Anselin
University of Chicago

Michael Batty
University College London

David Maidment
University of Texas-Austin

Victoria Stodden
University of Illinois at Urbana-Champaign

George Percivall
Open Geospatial Consortium

E. Lynn Usery
United States Geological Survey

Nancy Wilkins-Diehr
San Diego Supercomputer Center/University of California, San Diego
Background

Cyberinfrastructure for Sustained Scientific Innovation (CSSI) - Data and Software: Elements and Frameworks

PROGRAM SOLICITATION
NSF 18-531

REPLACES DOCUMENT(S):
NSF 17-500, NSF 17-526

National Science Foundation
Directorate for Computer & Information Science & Engineering
Office of Advanced Cyberinfrastructure

Directorate for Biological Sciences
Directorate for Education & Human Resources
Directorate for Engineering
Directorate for Geosciences
Directorate for Mathematical & Physical Sciences
Directorate for Social, Behavioral & Economic Sciences
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Existing NSF Software Institutes

- **MolSSI (Molecular Sciences Software Institute)**
- **IRIS-HEP (Institute for Research and Innovation in Software for High Energy Physics)**
- **SGCI (Software Gateway Community and Institute)**

**What is a gateway?**
Science gateways allow science & engineering communities to access shared data, software, computing services, instruments, educational materials, and other resources specific to their disciplines.

- **New to gateways?**
- **Building a gateway?**

Explore our services
Geospatial Data Complexity

- Dynamic
- Distributed sharing
- Heterogeneous
- Massive
- Multi-scale
- Privacy
- Quality
- Uncertainty
Geospatial Software

• Software for transforming geospatial (geo & spatial) data into information, knowledge, and intelligence

• Fusion of rapidly changing multidisciplinary sciences and technologies
Big Scientific and Societal Challenges

Increasing geospatial problems and questions!
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Matching Well with NSF’s Big Ideas

Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Conceptualization Process

• Three Workshops
• Community Survey
• AAG-UCGIS Summer School
Three Workshops
– 76 Position Papers Accepted

• Workshop 1: Mission and vision
  – https://gsi.cigi.illinois.edu/workshop/position-papers/

• Workshop 2: Use cases and core capabilities
  – https://gsi.cigi.illinois.edu/workshop2/position-papers/

• Workshop 3: Strategic plan and governance
  – https://gsi.cigi.illinois.edu/workshop3/position-papers/
Electronic Consent Form For Voluntary Participation in a Research Project  
(NSF SI2-S2I2 Conceptualization: Geospatial Software Institute):

Assessing the Needs and Practices of the Geospatial Software Community

This research project is being conducted by Shaowen Wang from the Department of Geography at the University of Illinois at Urbana-Champaign (UIUC). It has been funded by the National Science Foundation. Your participation is completely voluntary and you must be 18 years of age or older to participate. You are free to decline to participate. You may choose to withdraw from participation at any time without penalty or negative repercussion. The decision to participate, decline, or withdraw from participation will have no effect on your status at or future relations with the University of Illinois.

The goal of this project is to conduct research to understand the current needs of the geospatial software community. The main activities in the survey will involve answering a series of questions regarding the geospatial software that you currently use and what you use it for. We will also ask you a few questions about your education, occupation, and institutional affiliation. This activity should take you about 15 minutes to complete. While you may not directly benefit from your participation in this project, your response will allow the researchers to better understand the needs of the geospatial software community and to design innovative
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Survey Team at UIUC

William Barley  Dan Katz  Anand Padmanabhan  Rebecca Vandewalle  Shaowen Wang
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation
Toward a sustainable social and technical ecosystem
to enable geospatial-inspired discovery and innovation

Survey Findings – William Barley
In July 2019, a week-long summer school on Reproducible Problem Solving with CyberGIS and Geospatial Data Science will be co-led by the CyberGIS Center for Advanced Digital and Spatial Studies (CyberGIS Center) at the University of Illinois at Urbana-Champaign (UIUC), American Association of Geographers (AAG), and University Consortium for Geographic Information Science (UCGIS). Approximately 30 graduate students and early career scholars will learn to collaborate in developing novel solutions to complex problems and to take advantage of geospatial data science and cutting-edge scientific advances and technical capabilities of cyberGIS (e.g., CyberGIS-Jupyter and Virtual ROGER: cybergis.illinois.edu/infrastructures). Participants will experience the types of collaborative and professional interactions that are key to addressing reproducible geospatial problem solving in the context of computation- and/or data-intensive research involving confidential geospatial data. The program is built on the success of the inaugural UCGIS Summer School in 2017, and ideal for those working on interdisciplinary and transdisciplinary topics, including but not limited to: data-driven social and environmental sciences, digital humanities, geospatial artificial intelligence, and remote sensing big data.

This Summer School will be hosted on the UIUC campus at Urbana, Illinois, USA from Monday July 8 to Saturday July 13, 2019. If you are a graduate student or early-career scholar new to cyberGIS and geospatial data science and want to learn more about integrating them into your research, or are already working on cyberGIS and/or geospatial data science approaches, this Summer School will offer new and exciting opportunities for your professional development, and will help you build connections with others in related fields. UIUC will offer travel awards through generous support of the National Science Foundation (NSF) for participating in the Summer School.
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

https://www.nap.edu/catalog/25303/reproducibility-and-replicability-in-science
Computational Reproducibility
- Victoria Stodden

• Reproducible, transparent, and scalable geospatial software: Enable researchers to harness the geospatial data revolution for discovery and innovation by combining geospatial software and data at scale, in reproducible and transparent ways

• Structured guidance for computational reproducibility: Establish structured guidance for computational reproducibility in scientific research and education that are dependent on geospatial software
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation
Geospatial Digital Workforce:
Diana Sinton

• Increase the nation’s workforce capability and capacity to utilize geospatial big data and software for knowledge discovery supported by critical spatial thinking, and to further innovate geospatial software and advance related sciences
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

bit.ly/gvde_guidelines
Ethical and Open Geospatial Software: Coline Dony

• Promote a culture of ethical and open geospatial software driven by diverse communities
Toward a sustainable social and technical ecosystem
to enable geospatial-inspired discovery and innovation

Mission

• Transform geospatial software, cyberinfrastructure (CI), and data science across many fields to revolutionize diverse discovery and innovation by enhancing computational transparency and reproducibility
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Vision

• A sustainable social and technical ecosystem to enable geospatial-inspired innovation and discovery
Goals

• Reproducible, transparent, and scalable geospatial software: Enable researchers to harness the geospatial data revolution for discovery and innovation by combining geospatial software and data at scale, in reproducible and transparent ways

• Geospatial digital workforce: Increase the nation’s workforce capability and capacity to utilize geospatial big data and software for knowledge discovery supported by critical spatial thinking, and to further innovate geospatial software and advance related sciences

• Ethical and open geospatial software: Promote a culture of ethical and open geospatial software driven by diverse communities

• Structured guidance for computational reproducibility: Establish structured guidance for computational reproducibility in scientific research and education that are dependent on geospatial software

• High-performance and data-intensive geospatial software: Further the convergence of high-performance geospatial software with advancements in data-intensive and high-performance computing
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

The context of GSI’s mission, vision, and goals
Science Drivers – Geospatial Convergence

Crosscutting

Geospatial Data & Software

Analysis

Modeling

Observation

Domain-centric
Principles of Engaging Science
Drivers: Mike Goodchild
Social Science Example:
Ned English
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Map Flood Inundation at Continental Scale

Catchments and Flowlines

Digital Elevation Model

Height Above Nearest Drainage (HAND)
(relative elevation of land surface cell above cell in stream to which it flows)

“Combination of detailed GIS representation of stream network and supercomputing to determine the flow is transformative – CyberGIS has delivered a major success for the nation!” – Dr. David Maidment @ 2015 CyberGIS All Hands Meeting
CyberGIS in a Nutshell

Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

**NSF CyberGIS Software Project**

~$4.8 million, Year: 2010-2017

**Principal Investigator**
- Shaowen Wang

**Co-Principal Investigators**
- Luc Anselin
- Budhendra Bhaduri
- Timothy Nyerges
- Nancy Wilkins-Diehr

**Senior Personnel**
- Michael Goodchild
- Sergio Rey
- Marc Snir
- David Tarboton
- E. Lynn Usery

**Chair of the Science Advisory Committee**
- Michael Goodchild

**Project Manager**
- Anand Padmanabhan

**Project Staff**
- ASU: Wenwen Li and Rob Pahle
- ORNL: Ranga Raju Vatsavai
- SDSC: Choonhan Youn
- UIUC: Yan Liu and Anand Padmanabhan
- Graduate and undergraduate students

**Industrial Partner: Esri**
- Steve Kopp and Dawn Wright
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Geospatial Discovery and Innovation

CyberGIS Toolkit

GI Solve Middleware

CyberGIS Gateway

Virtual R O G E R

http://cybergis.illinois.edu

www.xsede.org
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Geospatial Data Science

Geospatial Sciences & Technologies

CyberGIS + AI + Data Science

Cyberinfrastructure & Computational Sciences

Mathematical & Statistical Sciences
Geospatial Data Science @ Scale

- Geospatial
  - Distribution
  - Dependence
  - Integration
  - Heterogeneity
  - Representation
  - Uncertainty
  - Etc.

- Computational
  - Complexity vs. intensity
  - Reproducibility vs. validity
  - Performance vs. reliability
  - Etc.

NSF DIBBs: Scalable Capabilities for Spatial Data Synthesis
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

A Digital Divide - David Tarboton

Hydrologic Experimentation and Modeling

Data-Intensive & High-Performance Computing

```bash
#!/bin/bash
vi
chmod
PBS -l nodes=4:ppn=8
grep
awk
mpiexec
```
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

It’s Time to Shift Emphasis Away from Code Sharing

Building well-documented, citable frameworks for Earth data analysis will encourage scientific replicability by addressing the underlying issues that inhibit code sharing.

As a community of Earth scientists, we need to develop holistically designed numerical toolboxes to ensure accuracy, transparency, and replicability. Credit: Adam S. Befus

By Chad A. Greene and Koustubh Thirumalai  © 20 February 2019

Have you ever watched a student struggle to perform a seemingly straightforward analytical procedure? It may be a routine preprocessing step, like detrending a time series or removing a seasonal cycle, but somehow the simple operations can stymie a student for weeks. It’s tempting to assume that young people with their short attention spans are unwilling or unable to think through the task at hand, Students may have little choice but to blindly tinker with code until things seem to work.

https://eos.org/opinions/its-time-to-shift-emphasis-away-from-code-sharing
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

HAND and Flood Emergency Response

This Jupyter notebook illustrates the HAND workflow and its use in example flood emergency scenarios. The study area is Onion Creek (HUC10 code 1203020554)

It is also a demonstration of conducting geospatial analysis with open source toolkits (gdal) + online Jupyter interface

- We use CyberGis’ accelerated TauDEM version for d8 and d10 flow direction calculation
- More about TauDEM
CyberGIS-Jupyter

Authors: Dandong Yin
Owners: Fangzheng Lu · Dandong Yin
Resource type: Web App Resource
Created: Sep 26, 2018 at 5:53 p.m.
Last updated: Sep 26, 2018 at 6:56 p.m. by Dandong Yin

Abstract

Jupyter environment set in CyberGIS Center for interaction with HPC

Subject Keywords

HPC, CyberGIS

How to cite


This resource is shared under the Creative Commons Attribution CC BY.

http://creativecommons.org/licenses/by/4.0/
CyberGIS-Jupyter for reproducible and scalable geospatial analytics

Dandong Yin\textsuperscript{1,2} | Yan Liu\textsuperscript{1,2,3} | Hao Hu\textsuperscript{1,2,3} | Jeff Terstriep\textsuperscript{3} | Xingchen Hong\textsuperscript{3} | Anand Padmanabhan\textsuperscript{1,2} | Shaowen Wang\textsuperscript{1,2,3}

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
CyberGIS-Jupyter Architecture
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Organization Structure and Governance

Advisory Board

Biannual meeting

Executive Committee

Executive Director (PI), Deputy Director, Co-PIs, Focus Area Leads, Managing Director, Chair of the Advisory Board

Focus Areas

Science Drivers
- Activities & Efforts

Core Capabilities and Services
- Activities & Efforts

Education and Workforce Development
- Activities & Efforts

Evaluation and Sustainability
- Activities & Efforts

Outreach and Partnerships
- Activities & Efforts

Geospatial Software Institute
Education and Workforce Development
- Eric Shook

- Equip geospatial communities with rigorous computational and data sciences and software engineering skills
- Meet users where they are and have capabilities for users who are not savvy computationally
- Combine formal and informal education for nurturing and serving diverse learning communities
Advanced CI Ecosystem – Anand Padmanabhan

• Engage and support communities (e.g., business, humanities, and social sciences) that are currently not well represented in the national and international CI ecosystem
• Serve as a conduit for bringing capabilities, processes and people together to tackle complex scientific problems while cross-fertilizing innovations of geospatial sciences and software
• Integrate with and leverage advanced CI (e.g., NSF Big Data Hubs, CyberGIS, TRIPODS, and XSEDE) to achieve high-quality, interoperable, and scalable software for broad impacts
Core Capabilities and Services – Carol Song

• **Deep**: To help users of geospatial data and software to tackle the challenges of scale and complexity of utilizing geospatial data in modeling, analytics, visualization and decision making to solve high-impact research problems and societal challenges

• **Wide**: To enable broader use of geospatial software and data especially by non-traditional and diverse geospatial software users, and train the next generation of researchers and workforce in creating and using geospatial software following the FAIR science principles

• **Transparent**: To promote and enable transparency and reproducibility of data-driven research and innovation by engaging both geospatial data producers and consumers in developing structured guidance, and employing an inclusive governance model
Partnerships – George Percivall

• Academic & International
  – AAG
  – AGILE
  – AGU
  – CyberGIS
  – ESIP
  – GIScience
  – UCGIS
  – XSEDE
  – Etc.

• Government
  – CDC
  – DOE
  – EPA
  – NASA
  – NGA
  – NIH
  – USGS
  – Etc.

• Industry
  – DigitalGlobe
  – Esri
  – Google
  – HDF
  – Kitware
  – LimnoTech
  – OGC
  – Etc.
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Outreach – Donna Cox
Sustainability – Dan Katz
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation

Evaluation – Margaret Palmer
Leadership

• Focus on fundamental scientific and societal challenges
• Prepare the future workforce
• Bridge the digital divide
• Foster open collaboration
• Enable discovery and innovation
Grand Opportunity!

Revolutionize discovery and innovation across many fields through synergistically advancing geospatial cyberinfrastructure, sciences, and software!
Toward a sustainable social and technical ecosystem to enable geospatial-inspired discovery and innovation
Community Inputs
– So Important!

• For this webinar
  – https://gsi.cigi.illinois.edu
  – Recording will be available in early January 2020

• Strategic plan

• Twitter hashtag
  – #GSIfuture
Acknowledgments

- National Science Foundation
  - BCS-0846655
  - IIS-1354329
  - OAC-1743184
  - OAC-1443080
  - OAC-1429699
  - OAC-1664119
  - OAC-1551492
  - OAC-1047916
  - XSEDE
Thanks!

• Comments / Questions?

• Email: shaowen@illinois.edu