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

http://gsi.cigi.illinois.edu/
Twitter: #GSIfuture

Shaowen Wang

CyberGIS Center for Advanced Digital and Spatial Studies
CyberInfrastructure and Geospatial Information Laboratory
Department of Geography and Geographic Information Science (Primary)
Department of Computer Science
Department of Urban and Regional Planning
National Center for Supercomputing Applications
School of Information Sciences
University of Illinois at Urbana-Champaign

PEARC18/Pittsburgh, PA
July 24, 2018
Fostering a Sustainable Geospatial Software Ecosystem at Scale

Co-PIs

Donna Cox, NCSA/UIUC
Daniel Katz, NCSA/UIUC
Paul Morin
University of Minnesota
Margaret Palmer
University of Maryland
Steering Committee

• Shaowen Wang, UIUC (chair)
• Donna Cox, NCSA/UIUC
• Michael Goodchild, University of California – Santa Barbara (chair of the advisory committee)
• Daniel S. Katz, NCSA/UIUC
• Paul Morin, University of Minnesota
• Anand Padmanabhan, UIUC (project manager)
• Margaret Palmer, University of Maryland
External Advisory Committee

- Luc Anselin, University of Chicago
- Michael Batty, University College London
- Michael Goodchild, University of California – Santa Barbara (chair)
- David Maidment, University of Texas – Austin
- George Percivall, the Open Geospatial Consortium
- Victoria Stodden, UIUC
- E. Lynn Usery, USGS
- Nancy Wilkins-Diehr, San Diego Supercomputer Center/University of California, San Diego
Goal

• Conceptualize a Geospatial Software Institute (GSI) as a long-term hub of excellence to serve diverse research and education communities
Big Scientific and Societal Challenges

- Climate change
- Emergency management
- Food security
- Population growth
- Sustainability
  - Energy
  - Environment
  - Water
- Urbanization
- Etc.
Convergence

Geospatial Data & Software

Analysis

Modeling

Measurement
Fostering a Sustainable Geospatial Software Ecosystem at Scale

Matching Well with NSF’s Big Ideas

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
Context

- Computation- & Data-Intensive Applications and Sciences
- Geospatial Data Science
- CyberGIS
- Advanced Computing & Cyberinfrastructure

- Earth & Environment, Emergency Management, Food + Energy + Water Nexus, Sustainability, etc.
- Spatial Computational Theories & Methods
- Science & Technology
- Extreme-Scale Computing, NSF XSEDE, ROGER, etc.
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
Fostering a Sustainable Geospatial Software Ecosystem at Scale

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
This book elucidates how cyberGIS (that is, new-generation geographic information science and systems (GIS) based on advanced computing and cyberinfrastructure) transforms computation- and data-intensive geospatial discovery and innovation. It comprehensively addresses opportunities and challenges, roadmaps for research and development, and major progress, trends, and impacts of cyberGIS in the era of big data. The book serves as an authoritative source of information to fill the void of introducing this exciting and growing field. By providing a set of representative applications and science drivers of cyberGIS, this book demonstrates how cyberGIS has been advanced to enable cutting-edge scientific research and innovative geospatial application development. Such cyberGIS advances are contextualized as diverse but interrelated science and technology frontiers. The book also emphasizes several important social dimensions of cyberGIS such as for empowering deliberative civic engagement and enabling collaborative problem solving through structured participation. In sum, this book will be a great resource to students, academics, and geospatial professionals for learning cutting-edge cyberGIS, geospatial data science, high-performance computing, and related applications and sciences.
Fostering a Sustainable Geospatial Software Ecosystem at Scale

Ecosystem

Domain Communities and Sciences

- Emergency Management
- Food, Energy, Water Nexus
- National Security
- Biosciences
- Engineering
- Geosciences
- Social Sciences

Crosscutting

GSI Capabilities and Services

Geospatial Data Transformation
- Data Integration and Analysis
  - Data Access and Aggregation
  - Data Management
  - Data Processing
  - Quality Control
  - Visualization
  - Uncertainty Quantification

Software Integration and Interoperability
- Software Engineering
  - Continuous Build & Test Automation
  - Open Source Support
  - Scalability Test
  - Software Environments

Software Integration
- Cloud & HPC Deployment
- Interoperability
- Portability
- Reproducibility
- Scalability
- Workflow Management

Geospatial Middleware
- APIs
- Computational Intensity Libraries
- Data-Intensive Computing
- Geospatial Workflow
- Microservices

Collaborative Problem Solving

GSI Capabilities and Services

Online PSE
- Data & Model Integration
- Data & Software Repository
- Interactive Computing
- User Services
- Visual Analytics

Advanced CI Capabilities

Data Infrastructure
Software Infrastructure
Community Services
EOT

Hybrid Architecture

Cloud Computing
Data-Intensive Computing
HPC
Integrated Data Systems

* EOT – Education, Outreach, and Training; HPC – High Performance Computing; PSE – Problem-Solving Environment
**A Digital Divide**

Hydrologic Experimentation and Modeling

Data-Intensive & High-Performance Computing

From David Tarboton
Three Workshops

• Workshop 1: Mission and vision, January 28-30, 2018, Los Angeles, California

• Workshop 2: Use cases and core technical capabilities, July 15-17, 2018, Chicago, Illinois

• Workshop 3: Strategic plan and roadmap, January, 2019, Washington DC
Program of the First Workshop
http://gsi.cigi.illinois.edu/workshop/agenda/
Leadership

• Focus on fundamental scientific and societal challenges
• Prepare the future workforce
• Bridge the digital divide
• Enable open collaboration
  – Academia
  – Government
  – Industry
  – Etc.
• Foster innovation
Advanced Cyberinfrastructure (CI) Ecosystem

• 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
Education and Workforce Development

- 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
Research and Software Capabilities

• Enable transformative sciences
  – For example, autonomous data collection systems are producing incredible amounts of data and many communities are using geospatial data collected by such systems
  – Software development for systems processing such data is very diverse and uncoordinated.

• Allow any researchers anywhere to have easy access to geospatial big data and related software based on the integration of diverse data worlds

• Blend well with commercial and open source activities through leading the change of tackling challenging research problems rather than competing with industry
Program of the Second Workshop
http://gsi.cigi.illinois.edu/workshop2/agenda/
Fostering a Sustainable Geospatial Software Ecosystem at Scale
1st Community Survey


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
Survey Team

William Barley, UIUC

Rebecca Vandewalle, UIUC
Fostering a Sustainable Geospatial Software Ecosystem at Scale
Grand Opportunity!

Revolutionize discovery and innovation across many fields through synergistically advancing geospatial computing, data science, and software at scale
Fostering a Sustainable Geospatial Software Ecosystem at Scale

Acknowledgments

- National Science Foundation
  - BCS-0846655
  - OAC-1743184
  - OAC-1443080
  - OAC-1429699
  - OAC-1047916
  - XSEDE

- US Geological Survey
Thanks!

• Comments / Questions?

• Email: shaowen@illinois.edu