

# *A Convergence Framework for Integrating CyberGIS Education and Research on I-GUIDE Platform*

Fangzheng Lyu  
*Dept. of Geography*  
*Virginia Polytechnic Institute and State University*  
Blacksburg, VA, United States  
[fangzheng@vt.edu](mailto:fangzheng@vt.edu)

Furqan Baig  
*Dept. of Geography & GIS*  
*University of Illinois at Urbana-Champaign*  
Urbana, IL, United States  
[fbaig@illinois.edu](mailto:fbaig@illinois.edu)

Yunfan Kang  
*Dept. of Geography & GIS*  
*University of Illinois at Urbana-Champaign*  
Urbana, IL, United States  
[yfkang@illinois.edu](mailto:yfkang@illinois.edu)

Erick Li  
*Dept. of Geography & GIS*  
*University of Illinois at Urbana-Champaign*  
Urbana, IL, United States  
[zhiyuan5@illinois.edu](mailto:zhiyuan5@illinois.edu)

Anand Padmanabhan  
*Dept. of Geography & GIS*  
*University of Illinois at Urbana-Champaign*  
Urbana, IL, United States  
[apadmana@illinois.edu](mailto:apadmana@illinois.edu)

Shaowen Wang  
*Dept. of Geography & GIS*  
*University of Illinois at Urbana-Champaign*  
Urbana, IL, United States  
[shaowen@illinois.edu](mailto:shaowen@illinois.edu)

**Abstract** — This paper described Data with Notebooks as a convergence framework for integrating cyberGIS education and research on the I-GUIDE Platform. The I-GUIDE Platform provides rich geospatial data and computing resources and adheres to FAIR (findable, accessible, interoperable, and reusable) data principles. The framework enhances these capabilities by harnessing scalable and reproducible geospatial analytics. Specifically, this framework supports the development of interactive cyberGIS education and research Jupyter Notebooks centered around geospatial datasets. A case study demonstrates the application of the framework, showing six cyberGIS research notebooks and four cyberGIS education notebooks based on a Twitter dataset. The case study and underlying framework highlight the potential of integrating cyberGIS research and education by making standardized datasets, interactive learning environments, and high-performance computing resources available on I-GUIDE Platform.

**Keywords**—*CyberGIS, GIS Education, I-GUIDE Platform*

## I. INTRODUCTION

The geospatial field is increasingly data- and computation-intensive. Handling geospatial big data, advancing geospatial artificial intelligence (GeoAI) models, and performing sophisticated analytics are now critical not only for conducting scientific research but also for effective learning in geospatial education programs [1]. Existing GIS training and education materials focus on key GIS concepts and teach students how to perform analysis using established platforms like ArcGIS and Google Earth Engine. However, with the growing popularity of geospatial big data and GeoAI, existing GIS educational resources are limited in their coverage of advanced analysis and modeling required for these emerging [2]. Consequently, there is a pressing need to advance GIS and GIS based on advanced computing and cyberinfrastructure (cyberGIS) education to include these data-intensive areas.

This paper introduces Data with Notebooks, a convergence framework for integrating cyberGIS education and research on the I-GUIDE Platform. We demonstrate interactive, data-intensive cyberGIS research and educational materials, leveraging the I-GUIDE Platform and its CyberGIS-Jupyter environment. CyberGIS-Jupyter integrates cyberGIS and Jupyter technologies for data-intensive, reproducible, and scalable geospatial analytics [3]. It encapsulates data, analytics, and workflow runtime environments into application packages reproducible through HPC and cloud computing [4]. And it has been widely used in different domains such as hydrology and health geography for geospatial knowledge discovery [5], [6]. The I-GUIDE Platform, developed by the Institute for Geospatial Understanding through an Integrative Discovery Environment, is designed to harness the vast, diverse, and distributed geospatial data at different spatial and temporal scales and make them broadly accessible and usable to convergence research and education enabled by cutting-edge cyberGIS and cyberinfrastructure.

The framework of Data with Notebooks designs a series of Jupyter Notebooks centered on geospatial datasets. From a research standpoint, this approach facilitates the publication of workflows and data, supporting open science and FAIR (Findability, Accessibility, Interoperability, and Reusability) data principles [7]. This framework enables data- and computation-intensive geospatial analysis utilizing High-Performance Computing (HPC) resources through the I-GUIDE platform. From the education perspective, the framework provides easily accessible cyberGIS learning resources. Additionally, it familiarizes learners with large-scale geospatial data analysis and modeling techniques.

## II. DATA WITH NOTEBOOKS

Data with Notebooks consists of four major components: the I-GUIDE platform, CyberGIS-Jupyter, cyberGIS Education Notebooks, and cyberGIS Research Notebooks, all centered around a core dataset (see Fig. 1). The geospatial middleware, consisting of the I-GUIDE Platform and CyberGIS-Jupyter, provides backend computing capabilities. Meanwhile, a series of GIS Education and Research Jupyter Notebooks are developed using the core dataset to teach fundamental GIS concepts and conduct GIS research.



Fig. 1. The architecture of Data with Notebook.

Data with Notebooks implements the FAIR principles of data management. Findability is ensured by assigning globally unique identifiers to each dataset in the I-GUIDE platform, providing metadata (author, description, etc.), and supporting search functionality for data, notebooks, publications and education resources. Accessibility is achieved through a backend NoSQL database for metadata storage and ensuring the data easily accessible on the I-GUIDE platform. Interoperability is facilitated by using a broadly applicable language for knowledge representation and including references to other data, notebooks, and external links for each dataset. Lastly, reusability is supported by comprehensive dataset descriptions on the I-GUIDE platform, with data and notebooks reproducible in a consistent Python environment using CyberGIS-Jupyter.

### A. I-GUIDE Platform

The I-GUIDE platform (<https://go.illinois.edu/IGP>) is an online knowledge discovery environment integrating data, domain knowledge, communities, and learning resources. It comprises four major components including datasets, notebooks, publications and educational resources (see Fig. 2). The platform hosts data using Amazon S3 and Google Cloud Service (GCV), ensuring user accessibility. Centered on a core dataset, it hosts GIS research and education Jupyter Notebooks executable in the CyberGIS-Jupyter environment. Publications provide data and notebook-related information, while educational resources offer teaching materials for GIS concepts. Additionally, an OpenSearch-enabled search function facilitates information retrieval. The I-GUIDE platform provides a scalable knowledge sharing and discovery system, linking datasets, notebooks, publications, and open educational resources to support I-GUIDE's convergence work and related communities.

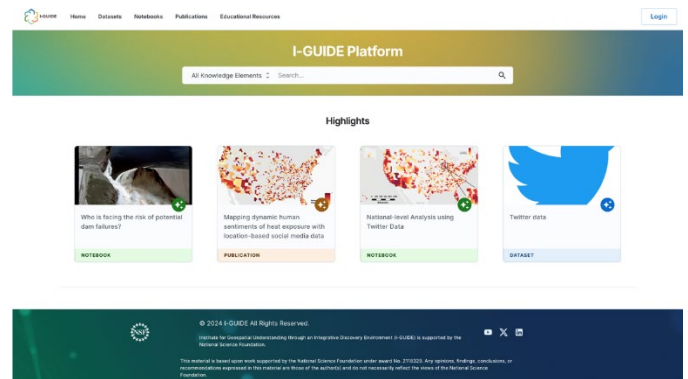


Fig. 2. I-GUIDE platform.

## III. A CASE STUDY ON TWITTER DATA

Location-Based Social Media (LBSM) data, such as Twitter data, has become valuable for GIS research due to its accessibility, low cost, and large scale [8]. Despite its importance in GIS research, LBSM data analysis presents challenges due to volume and analytical complexity. This case study uses a 20 GB Twitter dataset collected on September 26<sup>th</sup> and September 27<sup>th</sup>, 2021, using the *tweepy* library. Centered around the Twitter data, this case study provides an example of designing Jupyter Notebooks using CyberGIS-Jupyter on I-GUIDE platform for GIS education and research.

Figure 3 shows notebooks developed using Twitter data for the Data with Notebooks. Six GIS research notebooks include an introductory notebook on data collection and processing, and five advanced notebooks on data visualization and sentiment analysis of heat exposure at various scales using the Twitter data. Four GIS education notebooks cover data types for teaching basic GIS concepts, and map projection, choropleth maps, and the Modifiable Areal Unit Problem for more advanced GIS concepts, all utilizing the Twitter dataset.

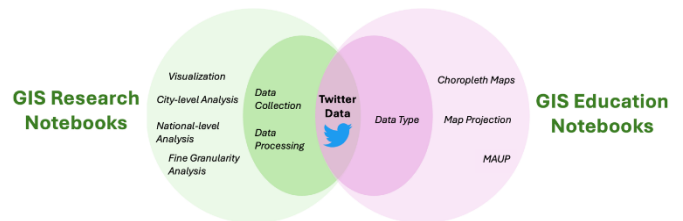


Fig. 3. GIS research and teaching notebooks.

### A. GIS Research Notebooks

Integration of advanced natural language processing (NLP) techniques with Twitter data enables sentiment analysis of heat exposure expressions [9]. This analysis can be crucial for the wellbeing and health of urban population. This series of research question-driven notebooks centers on a key Twitter dataset, addressing four primary research questions: 1) methods for Twitter data and related data collection and processing; 2) techniques for Twitter data visualization using Python; 3) evaluation and mapping of human sentiments of heat exposure at various scales; and 4) analysis on correlation between human sentiments of heat exposure derived from Twitter data with socioeconomic indicators.

To answer these research questions, we design a series of notebooks covers diverse aspects of analyzing heat exposure sentiments using Twitter data (see Fig. 4). It encompasses data collection from multiple sources; city-level analysis of Chicago at different spatial resolutions; hourly temporal analysis; data visualization techniques; integration of socioeconomic indicators with Twitter data for correlation analysis; and a notebook for national-scale sentiment analysis. The notebooks demonstrate methods for acquiring, processing, analyzing, and visualizing location-based social media data to understand human sentiments of heat exposure across various spatial and temporal scales.

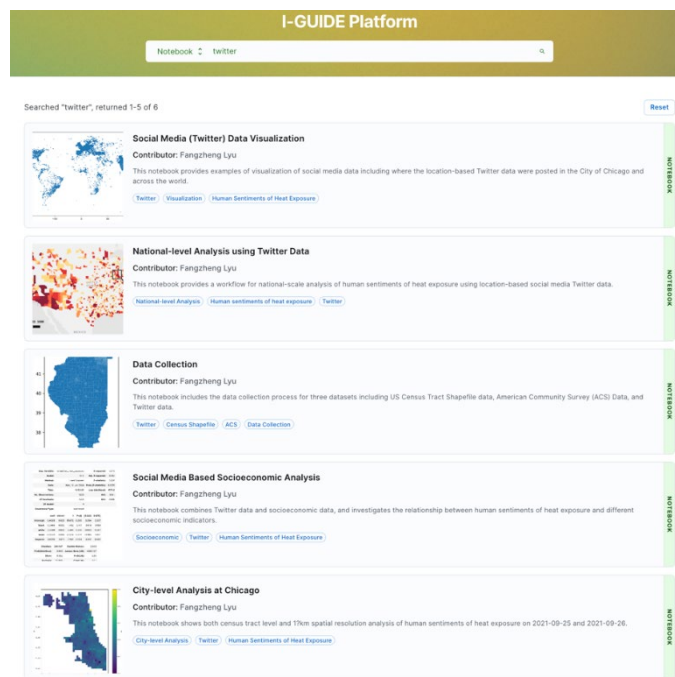


Fig. 4. GIS notebooks on I-GUIDE platform.

Data with Notebooks approach offers an efficient approach for data and code deposition and retrieval, while enhancing the visibility of published research. This approach facilitates improved data management and accessibility, potentially increasing the impact and reproducibility of scientific work.

### B. GIS Education Notebooks

Similar to the GIS research notebooks, this case study uses the Data with Notebooks approach to develop a series of GIS education notebooks centered on Twitter data. These educational notebooks not only help the students get familiar with geospatial big data analysis, but also, by reusing the same core dataset across notebooks, reduce the learning curve, enabling students to focus on new concepts while building upon existing knowledge. The CyberGIS-Jupyter platform provides a consistent workflow runtime environment with kernels containing necessary GIS-related Python libraries, enhancing the learning experience. Furthermore, The I-GUIDE platform provides students and users with efficient access to the learning materials.

We developed four notebooks centered on Twitter data, addressing key GIS concepts. These cover GIS data types

(vector and raster), choropleth maps for quantitative data visualization, map projections for transforming Earth's surface onto two-dimensional planes, and the Modifiable Areal Unit Problem (MAUP) in spatial analysis. The notebooks provide foundational knowledge in spatial data representation, visualization, cartographic techniques, and spatial analysis considerations, using Twitter data as a practical example.

### C. Connecting Data, Notebooks, Runtime Environment and Geospatial Problems

The I-GUIDE platform facilitates key functionalities for Data with Notebooks, notably the interconnectivity between data, Jupyter notebooks, runtime environments, and geospatial problems. To illustrate this integration, we examine the National-level Analysis using Twitter Data Jupyter notebook (see Fig. 5).

#### National-level Analysis using Twitter Data

Author: Fangzheng Lyu

This notebook provides a workflow for national-scale analysis of human sentiments of heat exposure using location-based social media Twitter data.

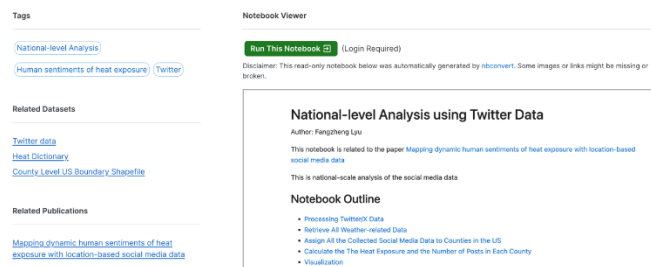


Fig. 5. National-level analysis notebook.

This Jupyter notebook is problem-driven and connects to a cyberGIS geospatial problem. It addresses the specific research question: "Can we analyze and visualize human sentiments of heat exposure using Twitter data at a national scale for the United States?" The notebook is also linked to datasets in the I-GUIDE platform including Twitter data, heat dictionary dataset, and county-level US boundary shapefile data, as indicated in the "Related dataset" section. And on the other hand, each dataset is associated with a set of relevant notebooks. The I-GUIDE platform employs keyword-based connections between notebooks, data resources, and publications, facilitating comprehensive searching and exploration. Furthermore, the Jupyter notebook is connected to the CyberGIS-Jupyter runtime environment. Users can access a pre-configured online Jupyter Lab environment by selecting "Run This Notebook," enabling interactive learning and execution of cyberGIS research and education notebooks.

This integrated framework of Data with Notebooks on the I-GUIDE Platform creates a cohesive ecosystem that seamlessly connects cyberGIS research and education components, enhancing both the learning experience and preparation for geospatial research challenges.

## IV. CONCLUSION

This paper introduces Data with Notebooks, a convergence framework for integrating cyberGIS education and research on I-GUIDE Platform. Data with Notebooks comprises four major

components: the I-GUIDE platform, cyberGIS-Jupyter, cyberGIS education Notebooks, and cyberGIS research Notebooks, all centered around a core geospatial dataset. The geospatial middleware, including the I-GUIDE Platform and CyberGIS-Jupyter, provides computational capabilities on the backend. Meanwhile, a series of cyberGIS education and research Jupyter Notebooks is developed using the core dataset to demonstrate foundational GIS concepts and facilitate cyberGIS research activities.

In the case study with Twitter data, six cyberGIS research notebooks and four cyberGIS education notebooks are developed. Data with Notebooks offers a comprehensive, accessible, and data-centric approach to GIS education and research by integrating consistent datasets, interactive Jupyter notebooks, high-performance computing resources, and standardized environments on the I-GUIDE platform, thereby reducing learning curves, enhancing reproducibility, and preparing students for real-world geospatial challenges.

#### REFERENCES

- [1] F. Bowlick, E. Shook, K. Kemp, and A. Padmanabhan, "Building CyberLiteracy through interactive lessons: Design, development, and deployment of jupyter notebooks for geospatial computing education," *Earth and Space Science Open Archive*, Earth and Space Science Open Archive, 13-Jul-2021.
- [2] R. C. Vandewalle, W. C. Barley, A. Padmanabhan, D. S. Katz, and S. Wang, "Understanding the multifaceted geospatial software ecosystem: a survey approach," *Geogr. Inf. Syst.*, vol. 35, no. 11, pp. 2168–2186, Nov. 2021.
- [3] S. Wang, "A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis," *Ann. Assoc. Am. Geogr.*, vol. 100, no. 3, pp. 535–557, Jun. 2010.
- [4] D. Yin *et al.*, "CyberGIS-Jupyter for reproducible and scalable geospatial analytics," *Concurr. Comput.*, vol. 31, no. 11, p. e5040, Jun. 2019.
- [5] F. Lyu *et al.*, "Reproducible Hydrological Modeling with CyberGIS-Jupyter," in *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning)*, Chicago IL USA, 2019.
- [6] J.-Y. Kang *et al.*, "Rapidly measuring spatial accessibility of COVID-19 healthcare resources: a case study of Illinois, USA," *Int. J. Health Geogr.*, vol. 19, no. 1, p. 36, Sep. 2020.
- [7] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, p. 160018, Mar. 2016.
- [8] D. Sui and M. Goodchild, "The convergence of GIS and social media: challenges for GIScience," *Geogr. Inf. Syst.*, vol. 25, no. 11, pp. 1737–1748, Nov. 2011.
- [9] F. Lyu, L. Zhou, J. Park, F. Baig, and S. Wang, "Mapping Dynamic Human Sentiments of Heat Exposure with Location-Based Social Media Data," *International Journal of Geographical Information Science*, vol. 38, no. 7, pp. 1291-1314, Apr. 2024.