Crossing Borders

A column by Doug Richardson
Executive Director, American Association of Geographers

Dealing with Geoprivacy and Confidential Geospatial Data

Creating and analyzing geospatial data are now central to most scientific domains and ubiquitous in governments and businesses. However, the unique confidentiality characteristics of location data present special challenges for geospatial research and its societal applications.

The opportunities and benefits of using geospatial data for scientific and governmental collaboration are often constrained by the need to protect research subjects' locational privacy and confidentiality. When geospatial data is presented in maps and visualizations or combined with sensor data or other related datasets, it may be possible to identify individuals.

Challenges to Using and Sharing Confidential Geospatial Data

The ability to replicate and reproduce research is a cornerstone of the scientific method. To ensure that this is possible, researchers funded by the National Science Foundation (NSF) and the National Institutes of Health (NIH) who generate or use confidential geospatial data also need to be able to share that data so it can be safely accessed, analyzed, and built on by the larger scientific community.

But in the health sciences, data confidentiality is increasingly cited as the primary reason for resisting data sharing. NSF policy explicitly indicates, though, that researchers should share data "in a form that protects the privacy of individuals and subjects involved." Furthermore, the new Geospatial Data Act of 2018 also establishes new layers of data privacy oversight for most other federal agencies.

An Integrated Approach to a Persistent Dilemma

Fortunately, promising new options are emerging for addressing confidential geospatial data management concerns and restrictions. In the past five years, the American Association of Geographers (AAG), the University of Michigan's Inter-university Consortium for Political and Social Research (ICPSR), and the University of Illinois at Urbana-Champaign (UIUC) received NSF grants to conduct collaborative research on addressing key issues associated with using and sharing confidential geospatial data. As a team, we have already achieved proof of concept and developed an experimental and testable Geospatial Virtual Data Enclave (GVDE), a virtual machine environment designed specifically for geospatial confidentiality research needs. It lets researchers share, use, and analyze remotely hosted geospatial data on their desktop computers but doesn't allow them to download it. The GVDE includes advanced GIS, spatial statistical, and other analytical and modeling tools, as well as masking and encryption methods to enable anonymized maps or data visualization to also be removed from the GVDE after review.

While we are currently focused on addressing the unmet needs of NSF researchers who wish to share yet protect their confidential geospatial research data, we also see many ways in which governments and businesses can apply our core technology for their geoprivacy and geospatial confidentiality needs. The GVDE system we have developed is an integrated, robust, and reliable geospatial confidentiality management infrastructure that can be adapted to multiple other applications.

We are currently conducting research on four interrelated program components that are required to create and implement a robust and reliable GVDE system for widespread use by NSF and, later, adaptation by other entities that rely on confidential geospatial data. Our integrated approach to achieving this goal involves using NSF-funded research to do the following:

- **Develop the GVDE and its core functions.** We are undertaking a research and implementation program that's necessary to address the specific challenges of working with geospatial data in a secure, virtual environment. This component of the project will also evaluate and integrate various software tools and procedures (e.g., data management, GIS, analytics, modeling, and spatial statistics) to enable researchers who use confidential geospatial data to share it, access it, analyze it, replicate it, and build on research within the GVDE—all remotely and in a virtual environment.
- **Evaluate and implement masking and encryption capabilities for the GVDE.** To allow researchers to anonymize and then export the maps, analyses, and visualizations they derive from confidential geospatial data, we are also examining multiple masking and encryption methods to include in the GVDE. This will make it easier for researchers to use this kind of geospatial data in publications, presentations, and other public-facing bodies of work. This component of the project consists of testing anonymization methods and related disclosure risks for specific types of geospatial data, such as points, lines, polygons, rasters, and vectors. We are also evaluating this for newer sources of confidential geospatial data, such as GPS trajectories, crowdsourced information, and data from social media.
- **Develop a GVDE researcher credentialing system.** We want to ensure that only trained and trusted researchers have access to restricted data in the GVDE, as well as in multiple scientific and related data repositories. That's why we are developing and implementing an innovative, robust, and reliable system to provide authorized researchers with a durable digital identifier that operates as a sort of research passport. This will allow large numbers of credentialed researchers to safely access and use the GVDE, increasing opportunities for them to collaborate and build on the important and extensive body of geospatial and GIS research now being carried out.
- **Ensure the GVDE's sustainability.** To assure that researchers can make widespread use of the GVDE for years to come, the system will be maintained as part of ICPSR's portfolio of ongoing data management and stewardship services. This will ensure that the GVDE has the financial and technical support of ICPSR's 770 member universities and other institutions. Additionally, to expand use of the GVDE to the broader geospatial research community and NSF grantees, we are developing training and outreach activities that cover how to use the system, data confidentiality ethics, credentialing requirements, and policies and best practices. The GVDE operates in a secure cloud environment that meets both Federal Information Security Management Act (FISMA) and Federal Risk and Authorization Management Program (FedRAMP) compliance standards, which provides an Authority to Operate (ATO) for use by federal agencies, including the US Census Bureau, the Department of Defense, NIH, and the Internal Revenue Service (IRS), among many others.

Together, the AAG, ICPSR, and UIUC bring significant, complementary expertise and experience to the long-standing issues associated with sharing confidential geospatial data. Through its strong management, research, and outreach capacities, the AAG is uniquely positioned to lead and successfully implement this complex research program. With its extensive experience in dealing with privacy and confidentiality protection—and as the custodian of large, data-intensive social science archives—ICPSR is uniquely positioned to help develop the digital research passport being used with the GVDE. And because of its long history of doing cutting-edge research in GIScience, geospatial data analysis and management, and GPS data and methodologies, UIUC is uniquely positioned to oversee how to securely and efficiently protect analytical outputs, including maps, that anonymize data.

Practical Solutions for Working with Confidential Geospatial Data

Our scientific and technological research program is providing workable and sustainable solutions to key geospatial data confidentiality issues, both for research purposes and within broader society.

These solutions will enable more people to use GIS and other geospatial technologies for health and scientific research. The GVDE is creating important new research infrastructure that scientists can use to share confidential geospatial data so they can replicate and build on one another's work. This also has the potential to transform how businesses and government agencies apply geospatial data and research to their own ventures and can provide a new resource to help them comply with current and evolving data confidentiality policies and regulations.

For more information on the GVDE program, contact Doug Richardson at **drichardson@aag.org**.

This article was written with input from UIUC geography and GIScience professor Mei-Po Kwan and ICPSR director Margaret Levenstein.