RENCI Geoanalytics

Data-driven maps to spark ideas and solve problems

A RENCI WHITE PAPER

renci

CONTACT INFORMATION

Jefferson Heard
Telephone: 919.445.9639
Email: jeff@renci.org

Summary

Data sets with a geographic component are essential to numerous applications. From the emergency manager responding to a natural disaster to the historian poring over ancient tax records, people rely on geographically-referenced data to meet a variety of research, business, and government needs.

As geographically-referenced data sets become ever larger and more complex, they hold ever greater potential for sparking insights and informing decisions. The tools to translate data into understanding, however, have struggled to keep pace. Some geographic analysis tools are so complex that only trained experts can use them. Others are limited to only certain data sources, or cannot easily be shared among collaborators.

RENCI's Geoanalytics platform provides intuitive, map-based visualizations to help transform data into decisions. Among other applications, this free, open-source software has been used to inform disaster response, study the spread of disease, and increase the accessibility of public records. An expanded user base and developer community would further enhance the tool's value and impact.

The Challenge

Mapping is one of the oldest examples of data visualization. From ancient times to today, people have created maps to understand and communicate our world. They answer questions: Where am I? They raise questions: Why is there so much cholera in this neighborhood? They spark insights: The hurricane will make landfall at noon.

Many of society's most pressing questions have a geographic component. We rely on mapping to discern patterns and make decisions related to weather and climate, public health, foreign policy, economics, energy, environmental conservation, and numerous other spheres.

The advent of Big Data has made it possible to tackle ever more complex questions. But while geographic data has become immensely easier to collect—anyone with a smartphone generates such data at every moment—using it to solve problems or gain insight poses greater challenges.

One challenge is sheer size. Today's data sets are often terabytes in size, or include rapidly evolving streams of information. Big Data requires new approaches to data storage, management, processing, interaction, and understanding. In addition, there is no standard format or cataloging system for geo-referenced data, making information from different sources often incompatible or difficult to find.

Another challenge inherent to working with geo-referenced data is that we often must represent it visually in order to make sense of it. For a researcher to assess the impacts of a drought on migratory birds, for example, she must first find the relevant data sets, then format them, then merge them with a mapping system before she can effectively discern patterns and communicate results. Sharing presents another hurdle. To allow other researchers to work with her map and its data, the researcher must somehow catalogue and archive the visualization and enable others to access it.

At a Glance

- Maps are meant to inform decisions. But as we amass ever larger and more complex data collections, it becomes increasingly difficult to represent location-based information in maps that are both powerful and intuitive.
- RENCI's Geoanalytics platform seamlessly combines large data sets to produce maps that take advantage of Big Data while remaining streamlined and comprehensible.
- These data-driven maps help illuminate patterns, communicate information, and solve problems for applications in research, business, and government.
- RENCI's free, open-source Geoanalytics software is available now and is being actively developed to enhance its value to users in diverse sectors.

The Team

Written by **Jefferson Heard**, RENCI Senior Research Software Developer, with support from **Anne Frances Johnson**. Design and editing support by **Karen Green.**

Vol. 2, No. 3 in the RENCI White Paper Series, July 2014
Renaissance Computing Institute
University of North Carolina at Chapel Hill
919-445-9640
www.renci.org
@renci

How to reference this paper

Heard, J. and Johnson, A. F. (2014): RENCI Geoanalytics: Data-driven maps to spark ideas and solve problems. RENCI, University of North Carolina at Chapel Hill. Text. http://dx.doi.org/10.7921/G06Q1V5N

Each of these steps—finding data, formatting it, visualizing it, sharing it—becomes increasingly difficult as the data collections become larger, more data sources are incorporated, more users become involved, or the intended tasks become more time-sensitive. For responders in a busy emergency command center, for example, any data point may be crucial and there is no time to waste.

The tools available to people who work with geographic data have struggled to keep pace with the rise of Big Data. A traditional geographic information system (GIS) allows users to build and analyze custom databases. Services such as Google Maps allow data to be published to the Web. But while open source tools enhance the utility of both of these tools, they can be rapidly outstripped by large data sets, and it can be difficult to combine multiple, often incompatible, data sources or facilitate collaboration among users. GIS systems often take a "kitchen sink" approach that overloads users with options and information, making it possible to solve complex problems but difficult to solve simple or common ones. In addition, these systems require extensive training to use, and their development cycle encourages the creation of data and logic silos that ultimately isolate researchers from one another.

Until recently, geo-referenced data has largely been integrated, cataloged, visualized, and shared on a case-by-case basis. No standard infrastructure existed to allow users to examine data and share their insights in an intuitive, map-based visual format. Data captured by one group of scientists through sensors and field studies were often unavailable to other researchers, or were stored in formats that were incompatible with data collections from other researchers. Even open source systems that could integrate instead created silos and failed to effectively facilitate data sharing, analysis, management, and visualization.

Ideas into Action

With these problems in mind, RENCI's Geoanalytics platform was developed in 2011 to aid in several scientific initiatives. It was released as an open source software in 2012 and has been actively developed and deployed in numerous applications since then.

RENCI's unique cyberinfrastructure allows users to work with multiple sources of geo-referenced data in an intuitive, map-based interface. The platform links data points with zip codes, latitude and longitude, or street addresses to create a Web-based visualization that can be easily accessed and shared to conduct research or solve problems. This flexible infrastructure can be extended and customized to suit the needs of any project that would benefit from integrating geo-referenced data with a Web-based mapping system.

RENCI Geoanalytics was developed by unifying multiple open source python geography projects into a content management and collaboration framework. The framework aims to implement, and in some instances extend, standards set by the Open Geospatial Consortium for geographic data and visualization interchange (http://www.opengeospatial.org). These standards allow researchers to share data from disparate sources in an interchangeable way and make it easier to work with data across platforms and disciplines.

Key features of this cyberinfrastructure include:

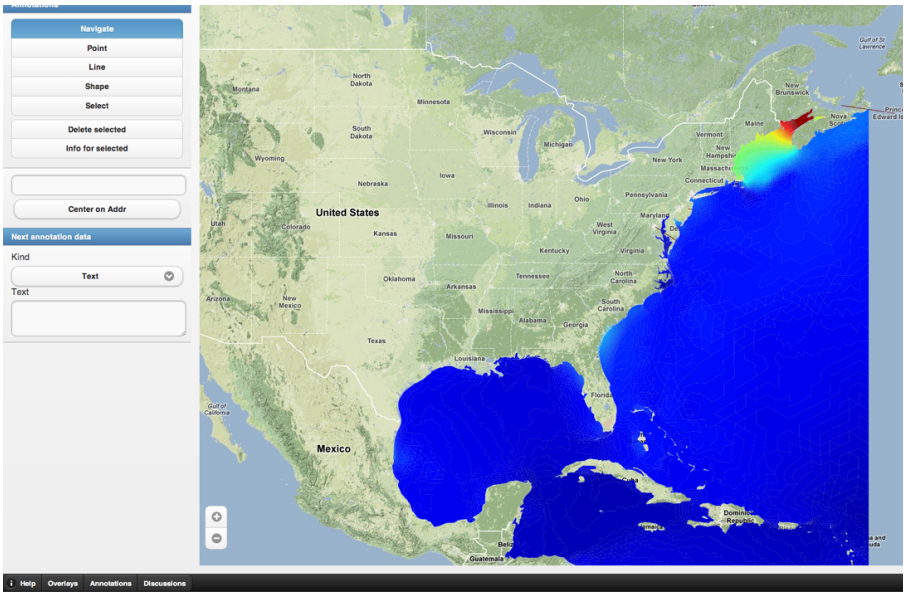
- Cataloguing: RENCI's Geoanalytics platform is designed to vet, recommend, and federate open source geography tools to make it easier to find and use available geographic data sets.
- **Scaling:** A data management and analytics layer incorporating iRODS, open source GIS software, RENCI's supercomputing resources, and a distributed task queue allow the Geoanalytics platform to scale to meet Big Data needs, support frequent updates, and fulfill appropriate access patterns and management requirements.
- **User-Centered Design:** The system provides pathways to easily accomplish common tasks, reducing the complexity of getting things done. In addition, a federated set of client-side software can be used to rapidly develop browser-based or mobile web applications.
- **Adaptability:** To remain flexible, the Geoanalytics team rapidly develops and deploys prototypes and solutions. The system's standards-based web service layer integrates open-source GIS and provides for data interoperability and rapid application development.

Because maps are meant to be used, the Geoanalytics platform has been closely tied to its users and applications throughout its development. It has been used to meet diverse needs in the areas of emergency management and mitigation, solar power, decision support, epidemiology, hydrology, and weather modeling, among others.

Example 1: Enabling Science-Based Decisions

When anticipating or responding to disasters, emergency managers must rapidly make decisions based on the information they have available. This information comes from a variety of sources, including environmental sensors, workers on the ground, and past experience.

With funding from the National Weather Service, RENCI developed a prototype Geoanalytics-based system to help North Carolina emergency managers more easily access, analyze, and share data during weather emergencies such as hurricanes. The resulting tool, dubbed "The Big Board," combines teleconferencing with a Web-based shared workspace that emergency managers can use to collaboratively analyze and respond to conditions as they develop.



A screenshot of the Big Board, which overlays data onto maps and allows participants in a session to query and annotate the map using the controls at right.

The sophisticated system merges geo-referenced weather and climate data with Web-based mapping and video conferencing capabilities. Users create a map-based virtual conference room and annotate it in real time, working independently or with other emergency managers to analyze information at various levels of detail. With the users in mind, the system was designed to allow annotations to be grouped and customized to present a streamlined view and avoid miring emergency managers in "information overload" during a rapidly developing situation.

Example 2: A Window to the Past

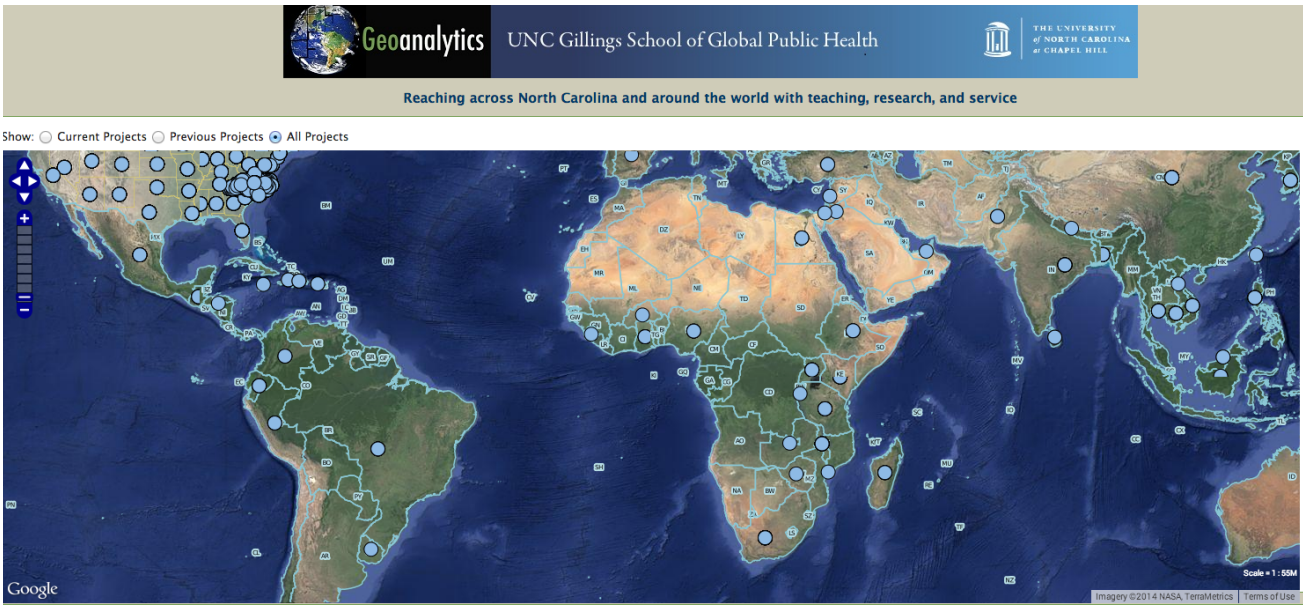
Children stand in front of a home in Asheville's Southside neighborhood in the mid-1960s. (photo courtesy of Housing Authority of the City of Asheville Records, Special Collections, D. H. Ramsey Library, UNC at Asheville).

Over the past century, Americans in hundreds of thousands of neighborhoods have been uprooted and relocated, their homes torn down and replaced with highways, parks, high-rise buildings, and other structures intended to reform cities and drive economic growth. This process, called "urban renewal," has complex economic, political, cultural, and ethical dimensions.

A collaborative research project by the City of Asheville, N.C., Duke University, and the University of North Carolina uses RENCI's Geoanalytics platform to illuminate the impacts of urban renewal in Southside, a historically African-American community in Asheville. Researchers and city residents use a Web-based map of Southside powered by Geoanalytics to collect, analyze, and share historical records such as photographs, documents, audio recordings, and videos that tell the stories of residents suddenly faced with the loss of their homes and neighborhoods.

The goal of the project is to shed light on the big-picture impacts of urban renewal by giving context to its nuanced ethical and cultural aspects. Researchers hope that this window to the past can help inform better policies for the future.

Example 3: Impact You Can See



In 2013 alone, researchers at the UNC Gillings School of Global Public Health netted 414 research grants and published more than 1,700 peer-reviewed research articles in areas such as cancer, obesity, and health disparities. The impacts of these research projects and public health interventions reverberate across the globe, but it is difficult to grasp the reach and diversity of the school's work from numbers alone.

A website built around RENCI's Geoanalytics tool helps the Gillings School assess and communicate the worldwide impacts of its work. Drawing up-to-date information from a comprehensive archive of all funded research projects maintained by UNC, the Gillings School "Where We Work" website lets users quickly find information about current and previous public health school projects using a world map to drill down on specific locations of interest (http://ga.renci.org/static/sphmaps/uncsph.html).

The school uses the tool to illustrate its impact to prospective students, faculty members, state legislators, and community leaders.

The Upshot

Using maps to analyze data used to require a high level of specialized expertise and complex, custom-built databases. Despite the increasing availability of large, valuable stores of publicly-accessible data, researchers, businesses, and governments have continued to struggle to transform data into usable, map-based visual interfaces that can be used to generate insights and inform decisions.

The goal of RENCI's Geoanalytics platform is to allow people to use location-based data in a way that is fundamentally easier and more fluid than has ever before been possible. By responding to the demands of Big Data and focusing on user needs, this technology offers a flexible solution to put powerful—yet intuitive—maps in the hands of researchers and decision makers.

The Big Picture

As maps have always done, RENCI's Geoanalytics framework enhances our understanding of the world. Maps yield clues about health and disease, our changing environment, our history, and the impacts of our policies. What is new about this technology is that it offers a way to seamlessly bring together vast, complex data sources and multiple collaborators where previous technologies would have created incompatible silos or buckled under the weight of too much information.

The Geoanalytics team is actively seeking collaborators and contributors to grow this endeavor. At the current stage, documentation—for both developers and users—is the most pressing need. In addition, we seek to expand our diverse user base to extend this technology into new realms, incorporate new features, and continue to increase the real-world impact of RENCI Geoanalytics.

About RENCI

RENCI is an institute of the University of North Carolina at Chapel Hill that develops and deploys advanced technologies to enable research discoveries and practical innovations. RENCI partners with researchers, policy makers, and technology leaders to engage and solve the challenging problems that affect North Carolina, our nation and the world. The institute was launched in 2004 as a collaborative effort involving UNC Chapel Hill, Duke University and North Carolina State University. For more information, see www.renci.org.

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL