Web-based Geovisualization and Geocollaboration: Applications to Public Health

Robert E. Roth[1], Anthony Robinson[1], Michael Stryker[1], Alan M. MacEachren[1], Eugene J. Lengerich[2], Etien Koua[3]
reroth@psu.edu | arobinson@psu.edu | stryker@psu.edu | maceachren@psu.edu | elengerich@psu.edu | elkoua@gmail.com

[1]GeoVISTA Center, Department of Geography, The Pennsylvania State University, University Park, Pennsylvania, USA
[2]Department of Public Health Sciences, The Pennsylvania State College of Medicine, Hershey, Pennsylvania, USA
[3]United Nations Development Programme, Addis Abba, Ethiopia

ABSTRACT: This paper describes the design and implementation of three web-based geovisualization and geocollaboration applications developed for the domain of public health. Each was implemented using Web 2.0 architecture. First, the Pennsylvania Cancer Atlas is a web-based geovisualization tool for the exploration of county-level cancer incidence rates using multiple interactive and live-linked statistical representations. Second, the Health GeoJunction is a web-based geovisualization and geocollaboration tool for geographic, temporal, and attribute filtering of articles from the PubMed medical research database. Finally, the Geo-Explication Web Portal is a web-based geocollaboration tool for the annotation and dissemination of geovisualization tools, training materials, and analysis artifacts collected from the application of geovisualization tools and techniques.

KEYWORDS: geovisualization, geocollaboration, statistical graphics, Web 2.0, public health, epidemiology

INTRODUCTION

Geovisualization involves the interactive exploration of geographically-referenced information graphics in order to prompt visual thinking about complex geographic patterns and processes (Slocum et al. 2005). In epidemiology, visual thinking informs decision-making regarding the etiologic investigation of disease incidence, deployment of limited public health resources, and adoption of health policies and regulations. Geovisualization bridges the gap between traditional cartography, concerned with creating maps that communicate information on geographic phenomena that are already understood, and scientific visualization, concerned with allowing users to 'see' the phenomenon in new ways to uncover or clarify things that are unknown or only partially known about the phenomenon (e.g., MacEachren 1994). Geovisualization software commonly provides users with multiple interactive, live-linked views of geographic information to encourage visual exploration in support of hypothesis generation and knowledge construction. While geovisualization was originally conceived as a private activity conducted by an individual user (e.g., DiBiase 1990), an emerging sphere of research has acknowledged the possibility and utility of spatial technologies that allow for the visual exploration of geographic information by multiple users in a group setting (Headley et al. 2002; e.g., MacEachren 2000; MacEachren and Brewer 2004; Slocum et al. 2001); this new area of research is referred to as geocollaboration. Although considered separate disciplines, the theoretical underpinnings of geovisualization and geocollaboration overlap significantly and the distinction between the two fields is often blurred. This paper introduces and describes three applications of geovisualization and geocollaboration to public health currently under development at the Penn State Geographic Visualization Science, Technology, and Applications (GeoVISTA) Center: (1) the Pennsylvania Cancer Atlas (PA-CA), (2) the Health GeoJunction, and (3) the GeoExplication Web Portal (G-EX Portal).

The three applications presented here are similar in their online accessibility and Web 2.0 architecture. Most early geovisualization tools were developed as desktop applications. When available online, access of geovisualization tools often follows a Web 1.0 model of file sharing, where the application and accompanying datasets are downloaded in their entirety first and ran locally. The applications we present here are designed specifically for online access using Web 2.0 architecture. The Web 2.0 concept describes a transition from the Internet as a simple computer-to-computer transfer mechanism (i.e., things moving through it) to the Internet as a platform atop which disparate data and services are interwoven for personalized use (i.e., things occurring on it) (O'Reilly 2007). Web 2.0 technologies are characterized by a separation of form (i.e., the data being analyzed) from function (i.e., the interface for exploring this data) through the compartmentalization of the application into at least three components: (1) one or more front-end clients for interacting with the data, (2) one or more back-end repositories for storing and indexing the data, and (3) at least one intermediate script or web service for relaying requests and data between the front and back-ends. The three applications described here use Web 2.0 architecture, each using a different variant

as shown in Figure 1. The breadth and depth of Web 2.0 technologies are expanding exponentially, and, like so many other domains, it can be envisioned that the web-based geovisualization and geocollaboration applications will significantly augment (or possibly even fully replace) their desktop counterparts as the performance and reliability of Web 2.0 architecture continues to improve.

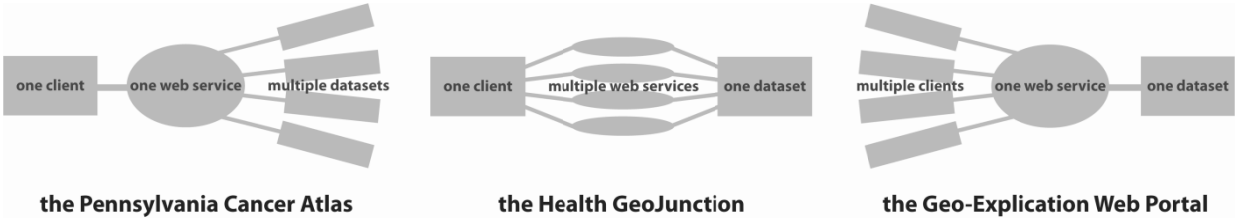

Figure 1: Three variants of Web 2.0 architecture. Compartmentalization among multiple clients, web services, and datasets is fundamental to Web 2.0 architecture.

THE PENNSYLVANIA CANCER ATLAS

The Pennsylvania Cancer Atlas (PA-CA) is a web-based geovisualization tool for the exploration of county-level cancer data for the state of Pennsylvania (Figure 2). The target audience for the PA-CA includes professionals in state and local public health departments who have a particular focus on cancer surveillance and control, policy makers in other government agencies or NGOs focused on public health issues, and individuals who work for voluntary organizations. The PA-CA uses a dataset compiled by the Pennsylvania Cancer Registry containing incidence rates for colorectal and prostate cancer from 1994-2002 stratified across age, gender, race, and stage at diagnosis. The design of the PA-CA was inspired by the State Cancer Profiles site (http://statecancerprofiles.cancer.gov), a web-based epidemiological tool that provides statistical graphics for analysis of cancer rates in each state. The architecture of the PA-CA includes an Adobe Flash client, a GeoServer web service intermediary that enables geographic querying, and geographically referenced datasets indexed in the PostGreSQL database. Although the current implementation features four datasets (ascending colon, descending colon, rectal, and prostate cancer), the tool is designed to use other county-level datasets with no modification. The PA-CA can be viewed online at http://www.geovista.psu.edu/grants/cdc/.

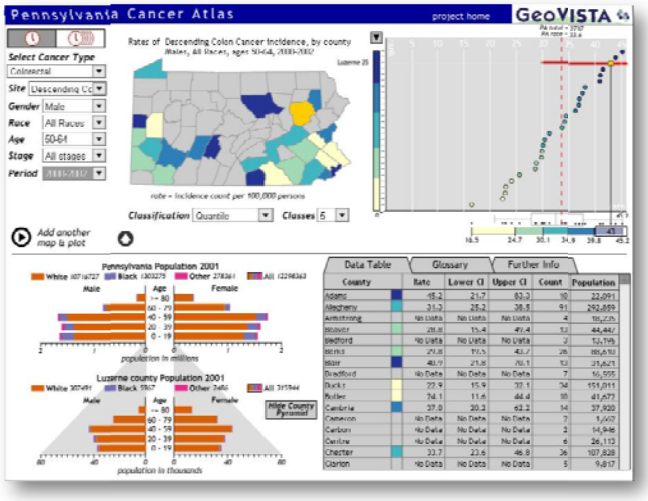

Figure 2: The Pennsylvania Cancer Atlas, a web-based geovisualization tool for the exploration of county-level cancer data for the state of Pennsylvania.

Four interactive, live-linked statistical graphics comprise the application: (1) a choropleth map displaying cancer rates given user-defined parameters, (2) a cumulative frequency plot illustrating the distribution of county values, (3) a population pyramid showing an age, gender and race profile for the entire state of Pennsylvania (and a comparative population pyramid for a single county when one is selected in the choropleth map), and (4) a table providing specific rates and summary statistics for each county. Brushing a data item in any of the views highlights the same data item in the other three views, allowing for visual exploration of unique places across multiple statistical graphics. It is also possible to animate a single parameter configuration through all available years to visually assess temporal trends in the choropleth map, cumulative frequency plot, and table views. Finally, users may explore choropleth maps and cumulative frequency plots using two different parameter configurations, allowing for users to visually compare spatial and statistical patterns across age, gender, race, stage of diagnosis, and time. More information concerning the implementation of the PA-CA and the user-centered design process applied during its design and evaluation is available in Bhowmick et al. (2008).

THE HEALTH GEOJUNCTION

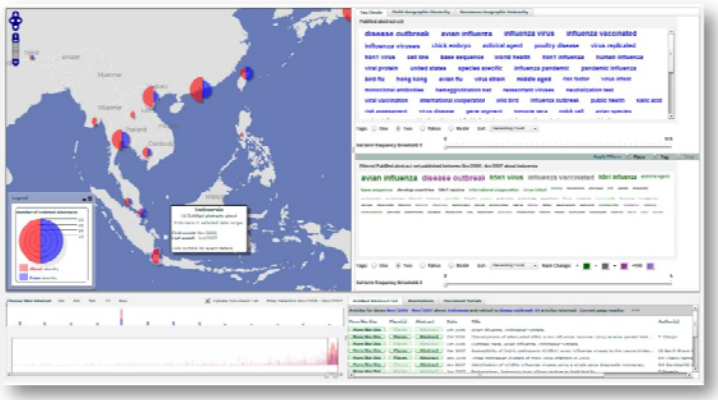

Figure 3: The Health GeoJunction, a web-based geovisualization tool for the exploration of infectious diseases across the globe.

The Health GeoJunction is a web-based geovisual analytics tool that maps information extracted from scientific literature and public health reports to assist identification of cross-connections among these extracted information fragments (Figure 3). The target audience for the Health GeoJunction includes researchers and analysts involved with monitoring and responding to disease events. The current prototype uses scientific articles about avian influenza from the PubMed medical database. The application allows analysts to quickly build a geographically-grounded understanding of the current science and public health policy related to distributed public health threats and events. The architecture of the Health GeoJunction features an Adobe Flex client, a wide array of internally-developed open-source API web service intermediaries, and a PostGreSQL database that stores and indexes the PubMed documents. The Health GeoJunction is a true Web 2.0 mash-up, currently implementing eight web services: (1) the GeoVISTA FactXtractor web service that extracts keywords and place names from relevant PubMed documents, (2) a custom java servlet that provides geographic, temporal, and keyword querying of the document collection, (3) a GeoServer web map service (WMS) providing a geographic reference map, (4) the Yahoo! web service for automated keyword extraction, (5-6) custom one term and two term keyword extraction services, (7) a custom web service for producing multiple types of tag clouds, and (8) the Apache Lucene indexing system for determining similarity between two text documents.

Four interactive, live-linked views of the data are provided in the application: (1) the FlexLayers map viewer (porting OpenLayers JavaScript™ for rendering a tiled WMS into Flex™) showing the frequency of documents written *about* each country on the left side of a split proportional symbol and the frequency of documents written *in* the country on the right half of the proportional symbol, (2) an interactive timeline for temporal filtering of the documents, (3) a set of two tag clouds, the first showing the frequency of keywords for the full PubMed document space and a second that can be filtered by place, time, and keyword, and (4) a list of PubMed documents that conform to the filtering parameters. Navigation of the application begins by showing all PubMed documents relating to avian influenza, is followed by filtering the documents by geography, time, or attribute using the map, timeline, and tag cloud views respectively, and concludes with selection of individual PubMed documents that may be of interest in the list view (or a return to the filter interface if no documents of interest are returned). This approach follows Shneiderman's (1996, 337) information visualization design mantra of "overview first, zoom and filter, then details-on-demand." More information on the design and implementation of the Health GeoJunction is available in Stryker et al. (2008).

THE GEO-EXPLICATION PORTAL

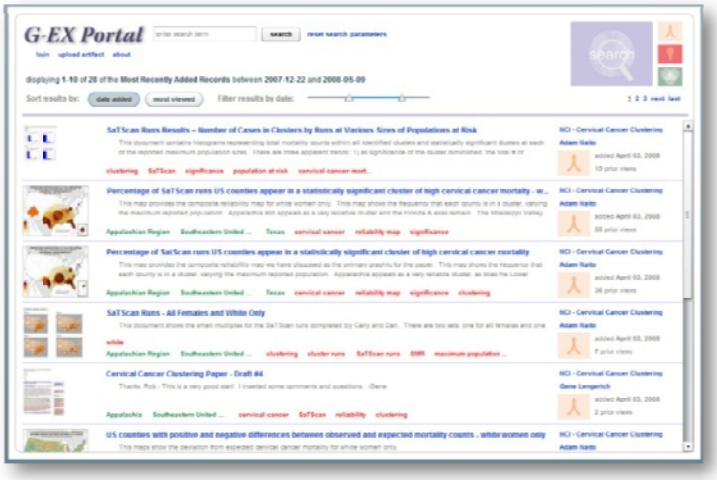

Figure 4: The Geo-Explication (G-EX) Portal, a web-based geo-collaboration tool for group geovisualization.

The Geo-Explication (G-EX) Portal is a web-based geocollaboration tool for the dissemination of analysis artifacts, datasets, and training materials derived from geovisual analysis (Figure 4). The G-EX Portal is targeted toward researchers who use ESDA and geovisualization software similar to the tools developed at the GeoVISTA center. The G-EX Portal allows for users to upload artifacts derived from analysis sessions; these range from simple screenshots or textual notes to geospatial datasets and multimedia tutorials. Design of the G-EX Portal was in part inspired by

del.icio.us (http://del.icio.us), Facebook (http://www.facebook.com), and YouTube (http://www.youtube.com), sites that allow for the annotation and dissemination of user-generated content based on the Web 2.0 model of social networking. The basic architecture of the G-EX Portal includes a Flex client, a Java web service intermediary, and a PostGreSQL database indexing the analysis artifacts. However, the web service was written so that other GeoVISTA web-based and desktop applications can communicate with it, allowing analysis artifacts to be sent to the repository during analysis and bypassing the Flex client altogether. These analysis artifacts are then available for interpretation, annotation, and commentary when users visit the Portal.

G-EX users are presented with links to four modules upon entrance into the application: (1) the 'Search' module, allowing the user to browse, filter, and synthesize analysis artifacts in the repository to which the user has access, (2) the 'Collaborate' module, allowing the user to view, discuss, and annotate analysis artifacts associated with the user's projects, (3) the 'Learn' module, allowing the user to browse and view multimedia tutorials and examples on geovisualization tools or analysis techniques (tutorials on the PA-CA and the Health GeoJunction, for example), and (4) the 'Review' module, allowing the user to review the analysis reports of others. Users may also log into their personal G-EX profile space and upload analysis artifacts to the repository. More details concerning the conceptualization and design of the G-EX Portal are available in Robinson et al. (2007).

REFERENCES

Bhowmick, T., A. C. Robinson, A. Gruver, A. M. MacEachren, and E. Lengerich. 2008. Distributed usability evaluation of the Pennsylvania Cancer Atlas. *International Journal of Health Geographics* 7 (36).

DiBiase, D. 1990. Visualization in the earth sciences. *Earth and Mineral Sciences* 59 (2):13-18.

Headley, N. R., M. Billinghurst, L. Postner, R. May, and H. Kato. 2002. Explorations in the use of augmented reality for geographic visualization. *Presence* 11 (2):119-133.

MacEachren, A. M. 1994. Visualization in modern cartography: Setting the agenda. In *Visualization in Modern Cartography*, eds. A. M. MacEachren and D. R. F. Taylor. New York, NY: Guilford.

———. 2000. Cartography and GIS: Facilitating collaboration. *Progress in Human Geography* 24 (3):445-456.

MacEachren, A. M., and I. Brewer. 2004. Developing a conceptual framework for visually-enabled geocollaboration. *International Journal of Geographic Information Science* 18 (1):1-34.

O'Reilly, T. 2007. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, January 1 2007, 17.

Robinson, A. C., E. Koua, F. Hardisty, and A. M. MacEachren. 2007. The G-EX Portal: Web-based Dissemination of Geovisual Analytic Results. Paper read at ICA Commission on Visualization and Virtual Environments Workshop 'From Geovisualization Toward Geovisual Analytics', August 2-3, at Helsinki, Finland.

Shneiderman, B. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization. Paper read at IEEE Conference on Visual Languages, at Boulder, CO.

Slocum, T. A., C. Blok, B. Jiang, A. Koussoulakou, D. R. Montello, S. Fuhrmann, and N. R. Hedley. 2001. Cognitive and usability issues in geovisualization. *Cartography and Geographic Information Science* 28 (1):61-75.

Slocum, T. A., R. B. McMaster, F. C. Kessler, and H. H. Howard. 2005. *Thematic Cartography and Geographic Visualization*. Second ed. Upper Saddle River, NJ, USA: Pearson Prentice Hall.

Stryker, M., I. Turton, and A. M. MacEachren. 2008. Health GeoJunction: Geovisualization of news and scientific publications to support situation awareness. Paper read at Geospatial Visual Analytics Workshop, GIScience 2008, September 23, at Park City, UT.