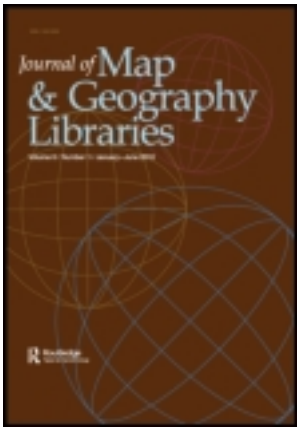

Journal of Map & Geography Libraries
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wmgl20

Applying GIS Methods to Public Health Research at Harvard University

Jeffrey C. Blossom [a], Julia L. Finkelstein [b a], Weihe Wendy Guan [a] & Bonnie Burns [a]

[a] Harvard University, Cambridge, MA, USA

[b] Cornell University, Ithaca, NY, USA

Available online: 20 Sep 2011

To cite this article: Jeffrey C. Blossom, Julia L. Finkelstein, Weihe Wendy Guan & Bonnie Burns (2011): Applying GIS Methods to Public Health Research at Harvard University, Journal of Map & Geography Libraries, 7:3, 349-376

To link to this article: http://dx.doi.org/10.1080/15420353.2011.599770

Journal of Map And Geography Libraries, 7:349–376, 2011

ISSN: 1542-0353 print / 1542-0361 online
DOI: 10.1080/15420353.2011.599770

Applying GIS Methods to Public Health Research at Harvard University

JEFFREY C. BLOSSOM
Harvard University, Cambridge, MA, USA

JULIA L. FINKELSTEIN
Cornell University, Ithaca, NY, USA; Harvard University, Cambridge, MA, USA

WEIHE WENDY GUAN and BONNIE BURNS
Harvard University, Cambridge, MA, USA

The Center for Geographic Analysis (CGA) at Harvard University supports research and teaching that relies on geographic information. This includes supporting geographic analysis for public health research at Harvard. This article reviews geographic concepts that apply to public health, pertinent data available in geographic format, and GIS analytical techniques. The work-flow methodology the CGA has developed for conducting research with geographic data will be presented, highlighting successful practices to follow and pitfalls to avoid. Applications of this work flow are illustrated through an in-depth discussion of specific case studies in public health research at the university.

KEYWORDS Health datasets, workflow, projects, geocoding, collaborative, webmap

INTRODUCTION

The Center for Geographic Analysis (CGA) at Harvard University supports research and teaching that relies on geographic information. In this role, CGA provides support for geographic analysis in public health research and other disciplines across the university. Harvard University researchers relying on geographic information are abundant and cover dozens of different disciplines. Since its inception in 2006, the CGA has served members of the public

Address correspondence to Jeffrey C. Blossom, Harvard University Center for Geographic Analysis, 1737 Cambridge St., Suite 350, Cambridge, MA 02138. E-mail: jblossom@cga.harvard.edu

health community at Harvard every month, with several months containing more than 50 active projects for public health or medical-related research or tasks. The majority of public health researchers aided by the CGA are from the Harvard School of Public Health and Harvard Medical School. On a less frequent basis, the CGA aids public health research for other Harvard entities including the Harvard Humanitarian Initiative and the Harvard Initiative for Global Health (Guan et al. 2011). The CGA's close collaboration with the Harvard Map Collection (HMC) and Harvard Geospatial Library (HGL) are essential in order to provide all researchers with the best possible service. The HMC holds 400,000 maps, more than 6,000 atlases, and thousands of reference books. HGL holds over 6,000 digital data layers ready for use in geographic information systems (GISs).

Discussed within the context of this existing infrastructure at Harvard, this article will draw on CGA's experience applying geographic analysis to public health research. Pertinent data available in geographic format will be reviewed, the full project workflow CGA employs to conduct projects will be articulated, and case studies in public health geography will be presented.

REVIEW OF GEOGRAPHIC DATA FOR USE IN SUPPORT OF PUBLIC HEALTH RESEARCH

Geographic data that support health research can be categorized as follows:

1. Data about health care capacities, such as health facilities, employment, and administration.
2. Data about the population and their health conditions and health care needs.
3. Data about the environment, both natural and social, that affect people's health.
4. Data about transportation and address location, which are essential in understanding accessibility for care, dissemination of medical supplies, transmission of infectious diseases, distribution of patients, and many other relationships among data of the above three categories.

We will examine each of these data categories, introduce the major data sources and popular data formats, and explain the most common usage of such data in health related research.

Data about Health Care Capacities

The US Department of Health and Human Services Health Resources and Services Administration (HRSA) maintains a Geospatial Data Warehouse and makes health care–related data available for download, including:

1. Health centers and look-alike sites.
2. Currently designated health professional shortage areas.
3. The National Sample Survey of Registered Nurses.
4. The Primary Care Service Areas project (a collaboration between HRSA and Dartmouth College).

Such data are in text or spreadsheet format, including postal addresses of the entities, which can be geocoded into spatial points data (see section below, "Data about Transportation and Address Location," for geocoding).

A summary of local health departments in the U.S. was published by Hua Lu and James Holt in 2009.[1] It was evident that local health services in the U.S. are fragmented, managed at different levels of local government agencies. It has been a challenge to develop a standard nationwide dataset to support effective collaboration.

Many states' GIS and health agencies maintain their own spatial health resource datasets for public use. Examples include Massachusetts[2] and New York.[3] In many states, such geospatial warehouses are often housed at state universities.

For data about health care providers, commercial data vendors are an alternative to public agencies. The ArcGIS Business Analyst is a licensed package of spatial data and analytical tools. It contains Infogroup Business Locations, which represents the locations of over 12 million private and public companies in the United States, among them health care facilities. It includes address information, estimated sales or assets, number of employees, franchise specialty information, and Standard Industrial Classification and North American Industry Classification System codes. These codes are four-digit numbers (standardized by the U.S. Government) that classify industries and are very useful in sifting out specific business types.

The availability of international health care data varies greatly from country to country, often depending on the national or provincial government agencies' data management practice. Health care provider information is harder to find in areas outside the United States The US National Geospatial-Intelligence Agency (NGA) maintains an international GeoNames service. It contains designations for medical centers, hospitals, and clinics.[4] The database is updated on a weekly basis, but locations are approximate and by no means complete for the globe.

Data about the Population and Health

Population data for health research can be further categorized into three categories. One is global population distribution and density data, which is aggregated and processed from nation-based census and other sources.[5] The second category is demographic health survey (DHS) data, which are

surveys in many countries describing disease outbreak rates and other health indicators of the population, often published with map coordinates of survey clusters.[6] The third category is the more widely known national census data from many countries such as the U.S.,[7] Great Britan,[8] and Israel,[9] just to mention a few.

Population data are usually derived from interpolation and projection models, which provide continuous coverage for the globe and an estimated value for the times when no census was conducted. Data are often organized into tables, graphs, or grid formats, which require further processing to match administrative polygon units in a GIS. Five of these global population datasets are listed below.

1. Gridded Population of the World, v3[10]
2. United Nations Population Information Network[11]
3. Home HYDE—the Netherlands Environmental Assessment Agency[12]
4. LandScan[6]
5. International Database—US Census Bureau[13]

DHS data are plentiful. The World Health Organization (WHO) maintains a searchable database online,[14] which includes data on mortality and health status, diseases, coverage of services, risk factors, health systems, and world health statistics. The WHO Global InfoBase is a data warehouse that collects, stores, and displays information on chronic diseases and their risk factors for all WHO member states. Data is presented to download in graph, map, and data-table formats.

The DHS program[15] collects, analyzes, and disseminates data on population, health, HIV, and nutrition through more than 200 surveys in over 75 countries. Data are organized in administrative units that can be mapped to the administrative polygons.

UNICEF maintains a Web site that contains statistical information on the well-being of women and children around the globe.[16] It also supports the Multiple Indicator Cluster Surveys (MICS), which is a major source of global development data downloadable for users having a log-in.[17] The data are organized by country and year in SPSS format. It requires further processing to match with the country boundaries for global statistical analysis.

In the United States, national health survey data are available from many sources. The National Center for Health Statistics,[18] part of the Centers for Disease Control (CDC), provides downloadable tabular data. Some data are mappable by address geocoding or by matching table records to survey geographic units, such as administrative units or metropolitan and micropolitan statistical areas (MMSAs).

Behavioral Risk Factor Surveillance System (BRFSS), also sponsored by the CDC, provides downloadable GIS data.[19] These files contain data and documentation that are available in ZIP archive file format. The ZIP files

contain BRFSS data that are mapped for both the states and MMSAs. These data files are a subset of the BRFSS data intended for use with a GIS package. Complete datasets and documentation for these data years are available in the BRFSS and SMART (self-monitoring, analysis, and reporting technology) sections of the site.[20]

CDC's WONDER, Wide-ranging Online Data for Epidemiologic Research, is a menu-driven system that provides access to a wide array of public health information by the CDC.[21] It allows for access to statistical research data published by the CDC, as well as reference materials, reports, and guidelines on health-related topics. Public-use data sets about mortality (deaths), cancer incidence, HIV and AIDS, tuberculosis, vaccinations, natality (births), census, and many other topics are available for query. The requested data are summarized and analyzed, with dynamically calculated statistics, charts, and maps. The data are ready for use in desktop applications such as word processors, spreadsheet programs, or statistical and geographic analysis packages. File formats available include plain text (ASCII), Web pages (HTML), and spreadsheet files (tab-separated values).

The Health and Medical Care Archive (HMCA)[22] is the data archive of the Robert Wood Johnson Foundation, the largest philanthropy devoted exclusively to health and health care in the United States. Operated by the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan, HMCA preserves and disseminates data collected by selected research projects funded by the foundation and facilitates secondary analyses of the data. The HMCA includes health care provider locations, household survey data, community disease information, and many more datasets used in numerous studies.

Local scale health surveys in the United States can be found at city or county government Web sites, such as the New York City Community Health Survey[23] and the Los Angeles County Health Assessment.[24] Some other countries publish their health survey results online as well. Examples are the China Health and Nutrition Survey[25] and the India National Family Health Survey.[26]

Census and demographic data are usually published by the respective countries' census organizations, as well as by commercial data vendors, who process the raw census data from the government and add value by applying statistical, geographic, or other analyses to produce derived demographic data that is more convenient for end users. In the United States, census data can be obtained from the US Census Bureau FactFinder[27] and a number of publicly accessible or subscription-based data providers, such as Geolytics,[28] Social Explorer,[29] ICPSR,[30] and Esri Community Analyst.[31]

Many other countries offer their census data through a fee-based service online or upon request. Examples include the China census data[32] and the India census data.[33] Most census data have their own geographic base units, either the same as local administration units (such as counties, districts,

and townships in China), or a uniquely defined system (such as tracks, block groups, and blocks in the U.S.). The census data are usually organized in tabular format, which need to be matched to the geographic units for mapping or spatial analysis.

Data about the Environment

Many environmental conditions affect people's health. For example, the distribution of wetlands may affect the dispersion of malaria, while groundwater aquifer and the location of superfund sites may impact drinking water quality, which in turn affect residents' health. Because of the huge variety of data subjects that could belong to this category, we will not categorize data further by its content, but by the major source format instead, which includes remotely sensed image data, geo-referenced survey data, and statistical data.

The US Geological Survey (USGS) maintains a map-centric data discovery tool, Earth Explorer,[34] which provides public image data of different scales from global satellite images, national aerial photos, worldwide digital elevation models, land cover classifications, and weather monitoring images.

The commercial image data vendor, GeoEye, maintains an online image search tool GeoFUSE,[35] which allows users to find images they need and examine the cloud cover or other quality issues before submitting a purchase request. The Food and Agriculture Organization (FAO) of the United Nations maintains several statistical and spatial databases on agriculture, nutrition, fisheries, forestry, food aid, land use, and population.[36]

The US Environmental Protection Agency (EPA) maintains many public databases and tools for analyzing environmental data.[37] Its Geospatial Data Access Project provides downloadable files of environmental facilities or sites in various GIS formats, including extensible markup language (XML), keyhole markup language (KML), Esri Shapefile, and Esri Feature Class.[38] The EnviroMapper is an online mapping system for browsing environmental quality data interactively.[39]

The Esri Data and Maps and ArcGIS Business Analyst packages offer licensed users access to US parks and recreational land use data as well as food and beverage business locations and retail locations of tobacco and alcoholic beverages, and so on.

More specific local environmental data are often collected firsthand by researchers equipped with GPS receivers on the streets or in the field.

Data about Transportation and Address Location

Accurate and detailed transportation networks are critical for two purposes. One is for traffic routing, which provides information on the best route and

travel time between any two points. The other is for address geocoding, which is the process of converting postal addresses into longitude and latitude coordinates. Datasets used for traffic routing and for address geocoding may look identical when displayed on a map—both appear to be linear networks. However, they require different preparation and contain different attributes on the street or road segments.

The most critical property of a traffic routing network dataset is connectivity among street or road segments. Other attributes in support of traffic routines include speed limits, one-way directions, turn restrictions, and vehicle height or weight limits. On the other hand, street networks in support of address geocoding require each street segment to contain street names and house number ranges, as well as other postal address components such as city and state names or zip codes. The geocoding software looks for matches between input and reference data, and when matches are found, the corresponding location is interpolated from the street segment whose value range contains the matching address number. Many of the commercial street network datasets are prepared for both traffic routing and address geocoding functions. The two combined allow for instant routing between a pair of user input addresses.

It is worth noting that both traffic routing and address geocoding datasets are readily available for North America and Europe, but not for many other parts of the world. Without a street level geocoder, addresses may be matched to zip code zones, town or city centers, or other geographic features. The locations obtained from such matching process is less precise than that from address geocoding, but nevertheless useful for some research purposes.

Information about foreign geographic feature names can be obtained from the GEOnet Names Server (GNS), developed and maintained by the NGA. The GNS database is the official repository of foreign place-name decisions approved by the U.S. Board on Geographic Names.[40]

The Geographic Names Information System (GNIS) is the U.S. federal and national standard for geographic nomenclature,[41] containing information about physical and cultural geographic features of all types in the United States, associated areas, and Antarctica. It includes current and historical place names, but not roads and highways. GNIS was developed by the USGS in support of the U.S. Board on Geographic Names as the official repository of domestic geographic names. Data were collected through a broad program of partnerships with federal, state, and local government agencies and other authorized contributors. "Hospital" is a category among the GNIS data types.[42]

Both address geocoding (using a street network) and location identification (matching place names with entries) in the GNIS or GNS achieve the purpose of mapping tabular records or descriptive data to a geographic location. Given that the vast majority of traditional health data reside in tabular files

and other non-GIS file formats, many of which containing addresses—such as patient records, clinic records, and hospital records—geocoding is an essential feature. Once turned into locations, they can be studied through spatial analysis, revealing new patterns, relationships, trends, and meanings. The potential to identify address locations of patients raises a privacy issue that must be considered when using geocoded patient or other sensitive data. To preserve patient privacy, disassociating patient information such as names, Social Security numbers, health care system, or other individual identifiers from the geocoded address locations is necessary.

WORKFLOW FOR CONDUCTING PUBLIC HEALTH RESEARCH WITH GEOGRAPHIC DATA

Since its inception, the CGA has provided services in support of hundreds of Harvard research projects that have involved geographic information. To handle efficiently this large volume of services, a well-defined workflow has been developed, which is used by CGA personnel. This workflow involves several stages from initial consultation to final product creation, all of which are outlined below. When providing services for health researchers, there are specific considerations common to most clients that must be factored into the workflow cycle. This cycle involves the following general stages: initial consultation, project execution, and final delivery.

Initial Consultation

The initial consultation between a CGA staff member and a researcher interested in using geographic information usually originates from an e-mail or at the CGA help desk. The CGA general contact e-mail address (contact@help.cga.harvard.edu) is prominently displayed on the CGA Web site.[43] E-mails sent to this address are automatically forwarded to the in-boxes of several CGA staff members to ensure a response to all queries within one or two business days. In addition, e-mails sent to this address are automatically logged into a request-tracking ticketing system.[44] Each request ticket is assigned to the appropriate CGA member for resolution.

The CGA conducts a help desk every Tuesday afternoon both at the Harvard main campus in Cambridge and at the Harvard medical campus in the Longwood medical area in Boston. During this time, at least one CGA specialist is available at a computer lab at each location. These labs have high-powered desktop computers each with a variety of GIS software installed. This is an opportunity for any Harvard researcher to receive in-person consultation, hands-on troubleshooting assistance, or GIS data and software demonstrations from a CGA specialist. Appointments can be scheduled,

or people can just drop in to receive help. Each help desk consultation or help session administered is logged into the CGA request-tracking ticketing system.

Whether initiated through the contact e-mail or help desk, the nature of requests varies, from a quick two-minute answer to a multiyear project involving multiple stages, deliverables, and personnel requirements. The requesting client's knowledge of using geographic information and GIS may also range from just having heard about GIS to an experienced GIS practitioner. This wide range of project complexity and client understanding requires a flexible, customized approach to handling each request.

During the initial consultation, the scope of work must be determined as soon as possible. This can be defined in the initial request, such as "How to convert a table of longitude, latitude coordinates into a shape file for use on a map?" or "Is there a GIS dataset available that contains hospital locations for Florida?" These clear procedural or information requests can usually be handled in a single reply; for example, sending the client a tutorial to resolve the former request, and extracting the requested data in the latter request. Typically, however, the scope of work determination is less straightforward.

To define the scope of work properly, the project objective must first be determined. The researcher usually has a very clear objective, for example, "I want a map that displays one mile buffers around hospitals in Chicago, fast food restaurants within these buffers, and a table listing the number of fast food restaurants within one mile of each hospital." Once the objective is established, then a scope of work required to achieve the objective can be developed. In defining the scope of work, the following must be taken into consideration: (1) datasets required to make the map or perform the analysis, (2) procedures and methodologies necessary to perform the work, (3) tools required to execute these procedures, and (4) deliverables needed and type of each deliverable (i.e., a Web map, statistical table, or geodatabase). The scope of work is documented in written form such that each component of the scope is clearly communicated.

Project Execution

Once the scope of work is determined, there are several ways to execute the project. The CGA's first approach is to equip the researcher to perform the work him- or herself. Harvard University has several proprietary GIS software programs licensed on a site-wide level, and it supports the use of many free and open-source software programs as well. Every semester the CGA conducts six different free, instructor-led training sessions at both the Cambridge and Boston campuses. These are intended for students or researchers who know little about GIS but want to learn the basics to apply to their coursework or research. More intensive, two-week "GIS institutes"

Center for Geographic Analysis
Harvard University

Center for Geographic Analysis
Project Specification

Template last updated: February 17, 2011

Project Identification	
Project ID 2011-XXX	**Project Affiliation** School, Department
Project Title	
Contact Details	
Name	**Position**
Email	**Phone**
Project Specification	
Description	
Technical Overview	
Budget (*see CGA Financial Model Overview below*)	
Schedule	
Risk Risks and concerns	
Data Confidentiality Data for this project is (place an **x** next to all that apply): __ HRCI __ HCI __Non-confidential *(see Data Confidentiality overview below)*	
Phone	**Email** @cga.harvard.edu

Project Manager **Date**
Month/Day/Year

FIGURE 1 CGA Project Specification Document.

are held each summer and winter, geared at educating the graduate student-level researcher. The institute consists of lectures, lab exercises, discussions, and facility tours, culminating with each participant's presenting a GIS project featuring individual work. There are also many self-help tutorials and how-to documents available on the CGA Web site to help any Harvard personnel to use GIS.

Often it is not feasible for the researcher to perform the work. In these cases the CGA will schedule the work into their queue of service projects. For such projects, first-time clients receive 4 hours of CGA time for free, and then are charged $75 for each additional hour of work required. Using this consulting model has enabled the CGA to prioritize projects in order to meet project specific deadline schedules. A project specification document is filled out for these projects, wherein the scope of work, project methodology, budget, time line, and deliverables are documented and agreed upon by both parties (Figure 1).

The first stage of project execution involves gathering or creating the required datasets. As outlined above in the section Review of Geographic Data, etc., there are many datasets readily available for public health research,

and various methods to geocode tabular data that may be necessary for use. The appropriate datasets are thus acquired as needed for the specific area of interest.

Once all datasets are acquired or created, the necessary methodology is applied to perform the GIS analysis. The methodology required to complete a project could use any number of GIS technologies and procedures. An overview of the general project methodology used in completing the major groupings of requests CGA receives will be discussed below. Detailed methodology for several specific projects will be presented further on in the article. The types of service projects CGA provides can usually be grouped into one of the following major categories: (1) geocoding and census variable extraction, (2) map creation for print publications, (3) dynamic Web map creation, and (4) GIS analysis and visualization.

Geocoding and Census Variable Extraction

Many researchers desire specific census demographic characteristics regarding their data for use in statistical regression modeling or analysis. If a researcher knows the level of geography and geographic area needed for census information (for example, all census tracts for the state of California), the CGA may provide an assisting role in pointing the researcher toward one of several methods to obtain demographic data for his or her area of interest or performing the data extraction itself. Care is taken in this process to ensure that the proper unique identifiers exist in attribute fields common to both datasets so that a table join can be performed. If a researcher needs to determine which census geographical unit their data are in, then CGA will first geocode the dataset using one of the methods listed in the section Review of Geographic Data, etc., above. Then a spatial join using ArcGIS software will be performed between the geocoded data and the desired level of census geography. An ArcGIS spatial join evaluates the geographic location of every feature in an input dataset (the geocoded data) against the join dataset (census GIS data) and appends the attributes of both datasets together into a new dataset. As a final step in this process, the data are often exported to comma separated value (.csv) format, for import into a statistical analysis software program.

Map Creation for Print Publications

Map creation for book, journal, dissertation, thesis, and other publications is performed on a nearl-daily basis at CGA. Both general reference and thematic maps are made, depending on a researcher's request. The application of best practice cartographic principles regarding the map layout, scale, projection, color, symbology, and type are used on these maps. Often explaining

or demonstrating different variations of these principles with a researcher's data is necessary to produce the desired map. The audience for the map is taken into consideration, as is the final output media type. Printed maps are published in sizes ranging from 2 × 2 inches to 42 × 60 inches. The CGA has a 42-inch-wide plotter on which large format maps or posters can be plotted. In addition, maps are made in a wide variety of image formats and resolutions for use in presentations and publication on Web sites. Technology used to create the maps usually includes use of one or more GIS software programs (predominantly ArcGIS), and image-processing or graphic-design software.

Dynamic Web Map Creation

The ability to customize a variety of Web map application program interfaces (APIs) has created an increasing trend in dynamic Web map requests with various levels of functionality. The CGA publishes interactive Web maps using a variety of APIs including OpenLayers and Google Maps. Researchers readily recognize the utility of an interactive Web map embedded with their own data. This allows for entire research teams to view and interact with the same custom maps from anywhere in the world, fueling thought and collaboration. Technology used to create interactive Web maps is driven by user requests. The nature of the request (whether the researcher wants a handful of points with information window pop-up functionality or a Web map with complex symbolization capabilities and full database interaction) plays a large role in determining which software stacks are used to complete the request.

Enabling researchers to create their own Web maps is a major focus at the CGA. This has resulted in the CGA's building a series of Google Map mashup tutorials[45] and creating the WorldMap platform,[46] an open-source framework currently under development by the CGA designed for viewing and interpreting maps collaboratively. One can load data and maps from various sources in multiple formats and have flexibility in choosing how to create, share, and mix different datasets and maps together.

GIS Analysis and Visualization

The CGA aids public health researchers with many forms of GIS analysis. A critical first step in the workflow methodology of every GIS analysis project is recognizing what projected coordinate system is best to use for the type of analysis to be performed. Data layers are then projected onto the appropriate coordinate system. Some of the more common GIS analyses applied for public health researchers are the creation of straight-line and network buffers. Buffer creation enables proximity analysis, which, as described above, is a

heavily applied spatial analysis method in public health research. Performing overlay and intersect between two map layers (for example, intersecting land cover and village regions to produce land-cover-type percentage values per village) is another common type of analysis desired by public health researchers at Harvard. Interpolation techniques such as inverse distance weighting, kriging, point density, and others are applied to datasets if needed. The CGA provides geostatistical analysis consulting as needed, and calls on a network of professors and researchers at Harvard for referral of specific questions. For many GIS analysis projects, the methodology is such that batch processing, macros, or software programming is required to best complete the job. Different programming languages such as Python, Java, .NET, and PHP are used when necessary. If specific programming skills required to complete the work do not exist within CGA, subcontactors are hired to write code. Animated PowerPoint slides, .gif images, and video files in all sorts of formats are produced when the temporal nature of a geographic dataset needs to be visualized.

Final Project Delivery

After the project execution stage concludes, final product delivery occurs. The end products to be delivered are always documented in the project specification document, which is updated and communicated to the client if the scope of work changes during the project execution stage. Project results might be tabular or GIS datasets, printed hard copy maps of various sizes, static map images for inclusion into print or Web publications, or interactive Web sites. For tabular or GIS dataset delivery, these data are sent to the researcher through e-mail or secure file transfer. A field key explaining what the data attributes or column names are is always included. Often a document listing the input data used and procedures employed to produce the data is also included. This usually completes the project workflow for tabular or GIS dataset delivery. For map, image, or Web site delivery, there is nearly always a revision process involved, and CGA factors this into the project specification. A draft map will be sent to the client for review and comment, then CGA will revise the map based on the comments and republish. This process is repeated until the researcher is satisfied with the map.

Once a dataset or map is deemed acceptable, the CGA specialist provides necessary metadata for the product and documents any processing procedures. This procedures document is saved into the project folder on CGA's file system and passed on to the researcher if requested. All project work is saved into individual project folders using a standard naming convention that includes the year the project was performed, project number, and client name. Subfolders within each project folder also have standardized names, so any CGA specialist is able to access and understand any project

data if the need arises in the future. Project data are saved on a Netapp[47] file storage system. After two years of inactivity, project data are archived on flash memory storage media to free up space on the Netapp. In delivering Web maps hosted on the CGA Web map servers, Web sites are hosted for clients for one year. Subsequently, the researcher may opt to ensure permanence of a custom Web site for a yearly fee. Once the product, metadata, and documentation are delivered to the client's satisfaction, an invoice for the work is sent. Datasets produced that might be of use to others can be published into HGL with the client's permission.

CURRENT GIS RESEARCH IN PUBLIC HEALTH AT HARVARD UNIVERSITY

Many research projects are currently underway at Harvard that involve geographic analysis in the field of public health. Five of these projects will be highlighted in this section.

Nurses' Health Studies

The Nurses' Health Studies are among the most significant and longest-running epidemiological studies of women's health, established at the Department of Nutrition of the Harvard School of Public Health. Since 1976, investigators have followed over 238,000 registered nurses to examine risk factors for major noncommunicable diseases. This study brings together medical and public health researchers from the Harvard School of Public Health, Harvard Medical School, Brigham and Women's Hospital, Dana Farber Cancer Institute, Boston Children's Hospital, and Beth Israel Deaconess Medical Center. The Nurses' Health Study, established in 1976 by Dr. Frank Speizer, and the Nurses' Health Study II, established in 1989 by Dr. Walter Willett, are landmark epidemiological studies on women's health and noncommunicable diseases.

GIS regression approaches and geocoded residential addresses are being used to estimate different exposures and investigate associations between health outcomes and the environment. The following sections will highlight some of the major analyses being conducted in this study.

Spatio-Temporal Estimation of Particulate Matter Exposure

Public health researchers have been investigating the health effects of particulate matter (PM) air pollution as one of the components of the Nurses' Health Study. Yanosky et al. (2008) used GIS techniques to analyze data from

the Nurses' Health Study; investigators used semiempirical models to predict spatially and temporally resolved long-term average outdoor concentrations of PM to successfully predict chronic fine and coarse particulate exposures for the northeastern and midwestern United States (Yanosky et al. 2009).

Researchers used a GIS-based spatial smoothing model to predict monthly outdoor PM10 concentrations (Yanosky, Paciorek, and Suh 2008), which included monthly smooth spatial terms and smooth regression terms of GIS-derived and meteorological predictors. Final model performance was strong (cross-validation $R^2 = 0.62$), with little bias (-0.4 μg m^{-3}) and high precision (6.4 μg m^{-3}). The final model performed better than a model with seasonal spatial terms (cross-validation $R^2 = 0.54$), and performed well in both urban and rural areas and across seasons and years. The addition of GIS-derived and meteorological predictors improved predictive performance over spatial smoothing (cross-validation $R^2 = 0.51$) or inverse distance weighted interpolation (cross-validation $R^2 = 0.29$) methods alone and increased the spatial resolution of predictions. The strong model performance demonstrated the suitability of these GIS-based spatial smoothing methods to estimate individual-specific chronic PM10 exposures for large populations.

Spatio-Temporal Estimation of PM Exposure and its Effects on Cardiovascular Disease

Researchers have applied GIS methods to investigate associations between environmental risk factors and health outcomes, using data from the Nurses' Health Study. In an analysis conducted by Puett et al. (2009), investigators examined the association of chronic particulate exposures with incident nonfatal myocardial infarction (MI), fatal coronary heart disease (CHD), and all-cause mortality in a prospective cohort of 66,250 women from the Nurses' Health Study in the northeastern United States (Puett et al. 2009). Researchers developed a spatio-temporal model to estimate monthly PM10 and $PM_{2.5}$ exposure between 1988 and 2002, using government monitoring data, geocoded residential addresses, and covariates calculated using GIS overlay analysis between address locations and census tract boundaries. Multivariate models included hypertension, family history of MI, hypercholesterolemia, body mass index (BMI, continuous), physical activity (<3, 3 to <9, 9 to <18, 18 to <27, or ≥27 metabolic equivalent hrs/week), smoking status (never, former, or current), diabetes, median house value, and household income for census tract of residence, season, and state of residence; these were stratified by age in months. GIS techniques were used to account for uncertainty and truncation in available data sources. In addition to the primary exposure, defined as the average exposure to $PM_{2.5}$ and $PM_{10-2.5}$ in the 12 months before the outcome of interest, other windows of exposure were considered, namely, 1, 3, 24, 36, and 48 months prior to exposure; $PM_{2.5}$ and

$PM_{10-2.5}$ were also assessed in separate single- and two-pollutant models. All multivariate models were also stratified by age in months and adjusted for state of residence (indicator variables), year (linear term), and season (indicator variables), to adjust for large-scale spatial mortality patterns that might be related to factors apart from pollution. Authors found increased risk of all-cause mortality (hazard ratio [HR], 1.26; 95% confidence interval [CI], 1.02−1.54) and fatal CHD (HR = 2.02; 95% CI, 1.07−3.78) associated with each 10-μg/m(3) increase in annual PM(2.5) exposure (Puett et al. 2008; Puett et al. 2009). Findings demonstrated that chronic PM(2.5) exposure was associated with risk of all-cause and cardiovascular mortality.

Association Between Residences in U.S. Northern Latitudes and Rheumatoid Arthritis

In order to examine the geographic variation in the occurrence of rheumatoid arthritis (RA), investigators analyzed geocoded addresses and incident case diagnosis of RA (or censoring of controls) in the Nurses' Health Study between 1988 and 2002. Generalized additive models were used to predict a continuous surface adjusted for known risk factors. Permutation tests were conducted to test for the importance of location and identify areas with statistically significant increased risk, compared with the entire study area. Spatial analyses demonstrated that women residing in high northern latitudes may be at greater risk for RA ($P < 0.05$). Findings also demonstrated the utility of applying GIS methods to traditional large-scale epidemiological studies to pose new research questions and generate hypotheses for future investigation.

The AfricaMap Project

AfricaMap is an open-source software project developed to support academic research and teaching and bring together resources from a variety of disciplines in a single geographic environment. This open-source system allows for the investigation, analysis, visualization, and communication of multidisciplinary, multisource and multiformat data, organized spatially and temporally.

This project was the first application of WorldMap, developed at the Center for Geographic Analysis (Lewis and Guan 2010). Since its Beta release in November of 2008, the AfricaMap framework has been implemented in several different locations including metro Boston, Chicago, East Asia, Vermont geological sites, Harvard Forest, and Paris. These Web mapping applications are used by individual researchers, and they have been incorporated into courses at Harvard University.

AfricaMap consists of a set of public digital base maps of the continent, viewable dynamically at a range of scales and composed of the best

cartographic mapping publicly available. A gazetteer provides rapid navigation to specific locations across a vast landscape. As more detailed mapping becomes available, it is added to the system. Because of its decentralized architecture, there is (in theory) no hardware or software limitation to the amount of data that could be incorporated. Although currently focused on Africa, the Web-based mapping framework behind the project could be used to organize information for any region of the world. This model aggregates data using maps, rather than disciplines, authors, subjects, or indices.

AfricaMap serves the needs of researchers in multiple disciplines interested in Africa such as public health and medicine, including a common Web-accessible set of current and historical maps of Africa, a comprehensive gazetteer of African place names, and a repository for spatial and nonspatial datasets for research projects on Africa.

The AfricaMap project represents a framework for organizing African data from a variety of disciplines in a single environment. This allows researchers to explore public health research questions with the breadth and depth of data from other disciplines in a single environment. For example, researchers examining malarial transmission patterns in sub-Saharan Africa can explore environmental factors, vector sources, habitat reservoirs, soil type, land usage, urban planning, agriculture, socio-demographic data, and population health in a single environment and develop hypotheses for further analysis and investigation (Figure 2). Varied spatial information is also available to view and download for further study, including historical maps, topographic maps, historical trade routes, anthropological and ethnographic data, and epidemiological data.

AfricaMap supports collaborative public health research and teaching on Africa. This project brings together public health researchers from across Harvard University—including the Harvard School of Public Health, Harvard Medical School and affiliated hospitals, Harvard Humanitarian Initiative, Harvard Initiative for Global Health, and Faculty of Arts and Sciences—from a variety of disciplines including public health, medicine, infectious diseases, nutrition, epidemiology, biostatistics, demography, geography, sociology, and anthropology. Public health researchers can utilize the Projects layer in AfricaMap to search for ongoing or historic research projects, examine relevant publications and funding sources, and identify investigators and potential collaborators in a single environment. AfricaMap facilitates discovery and collaboration among researchers from across disciplines to examine health research questions and geographic analysis in Africa.

Child Physical Activity in the Built Environment

Childhood obesity is a public health issue that has risen to epidemic proportions in the past few decades (Koplan, Liverman, and Kraak 2005).

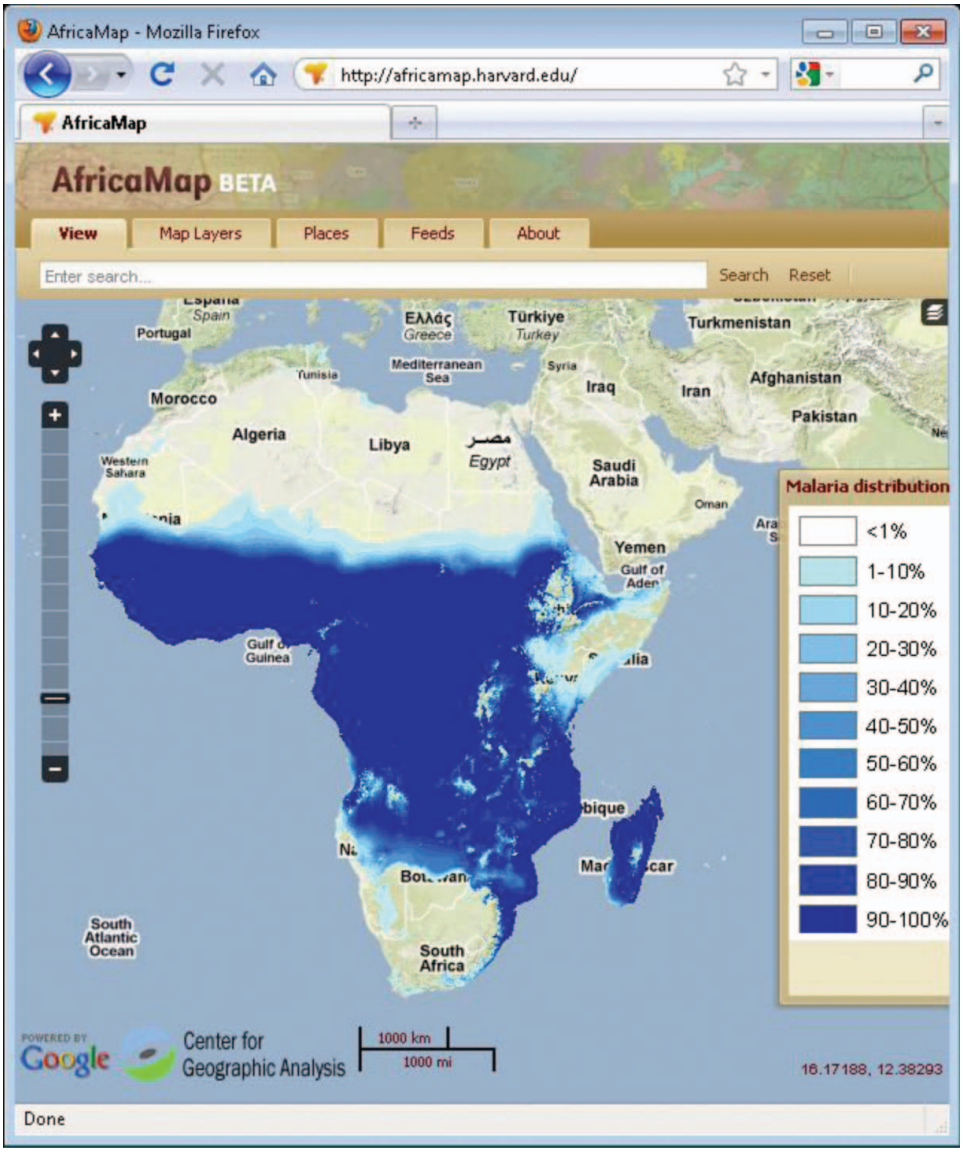

FIGURE 2 AfricaMap displaying malaria distribution.

Between 1970 and 2004 the prevalence of obesity almost tripled among U.S. preschoolers and adolescents and quadrupled among children aged 6 to 11 years (Ogden et al. 2006). Poor daily diet and lack of physical activity frequently lead to childhood obesity, and numerous factors figure into shaping these two behaviors, including personal and cultural beliefs, environmental conditions, societal influences, health care access, and individual physiology (National Institute of Health 2006).[48]

Recommendations by the American Academy of Pediatrics to prevent childhood obesity include getting at least 1 hour of physical activity per day, limiting sugary beverages and high-fat, fast food consumption, and switching dietary habits to include low-fat dairy products and high-fiber and calcium-rich foods.[49] It is common knowledge that physical activity can be increased through activities such as organized sports, but no intervention has yet sought to increase physical activity by increasing unstructured physical activities that include the use of one's built environment (Ogden et al. 2006). This includes sidewalks, open space, playgrounds, parks, and other areas freely accessible to children in which physical activity can occur. Assessing an individual's comprehensive use of the built environment throughout a given day in an attempt to find ways to enhance physical activity has had little previous study, and it is the focus of a current study[48] conducted by Harvard-affiliated medical professionals from Massachusetts General Hospital.

The study aims to collect objective information on adolescents' use of the built environment using GPS receivers and accelerometers (devices that measure one's movement or physical activity). By using GIS overlay analysis with the GPS and accelerometer data, the built environment elements most associated with physical activity in adolescents can be identified. Variations in use patterns of the built environment by age, gender, and socio-demographics can then be assessed. This objective analysis of where children are active or inactive throughout the course of an entire day may lead to discovery of methods for children to more effectively use the built environment for exercise.

The CGA is involved in this study by providing consultation regarding appropriate GPS devices with which to equip the children, writing a script that will automatically join GPS readings and accelerometer readings by date and time (using 30-second intervals), and performing overlay analysis using GIS. During December 2009, a preliminary study was conducted involving middle school children wearing a GPS and accelerometer for 7 consecutive days (Figure 3). The GPS and accelerometer data were joined by date and time. Locations were classified into categories in GIS using base-map data provided by MassGIS[50] (Figure 4).

Once classified, activity levels and locations were graphed for visual analysis and loaded into SAS for statistical analysis. This study is in the early stages, and it will collect activity and location information for the same group of children over several years.

The Surgical Safety Web Map

The Safe Surgery Saves Lives (SSSL) program has a mission of improving surgical care worldwide by ensuring adherence to proven standards of care in all countries.[51] Sponsored by WHO, the SSSL team is headquartered at the

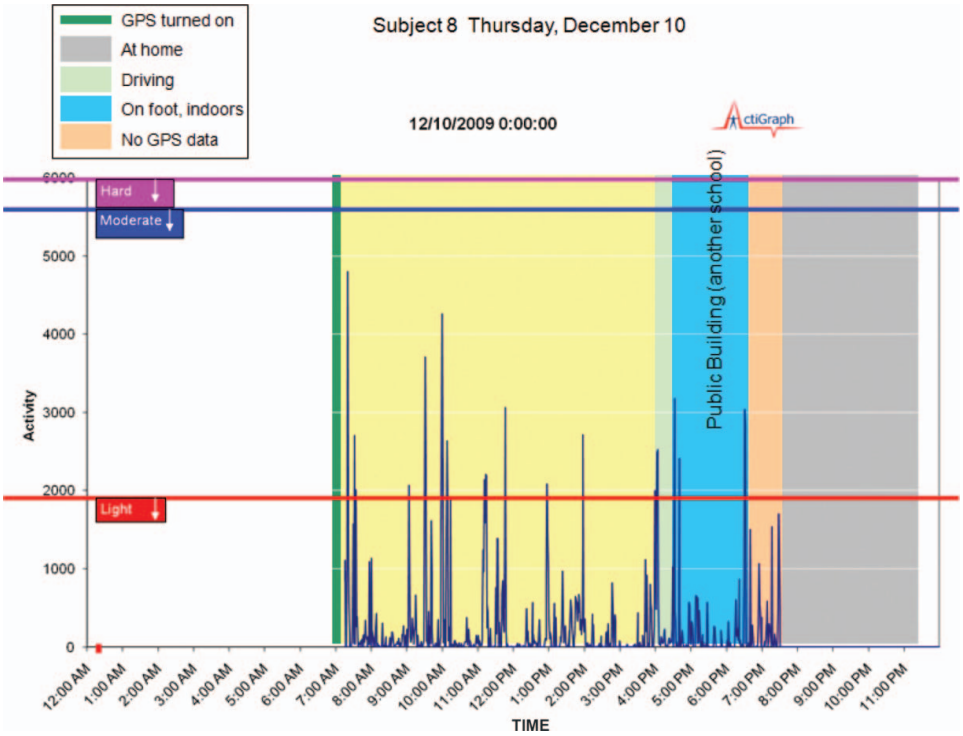

FIGURE 3 Graph of physical activity and locations.

FIGURE 4 GPS location classifications on a MassGIS orthophoto.

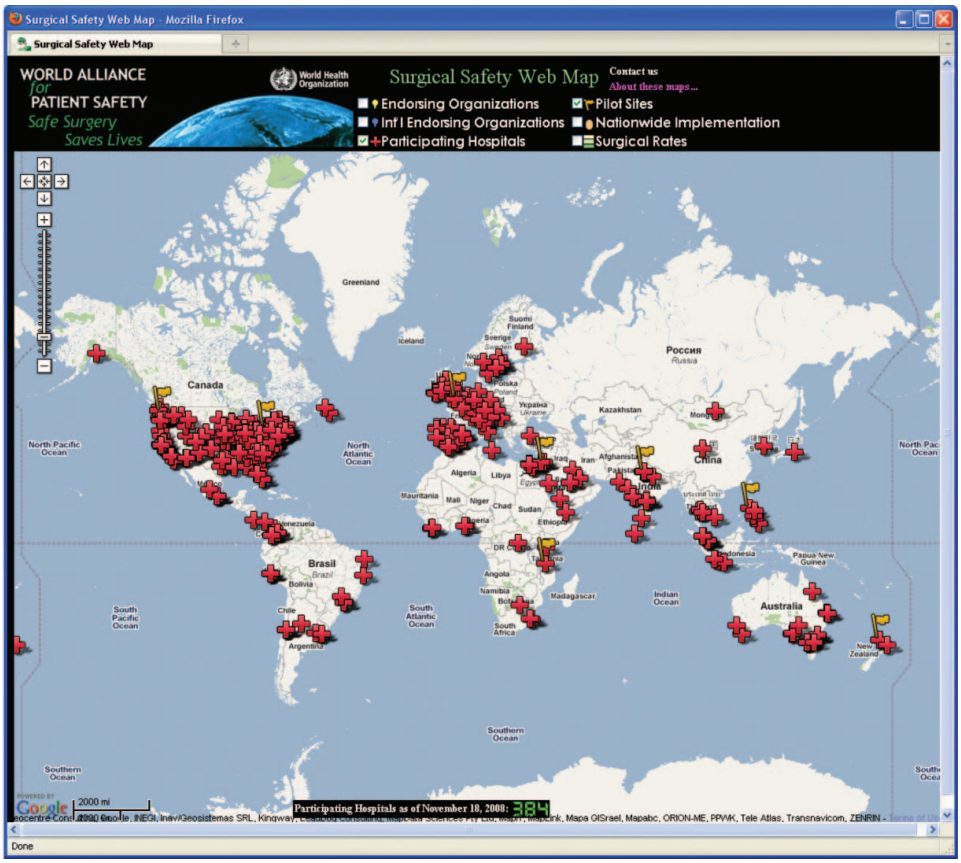

FIGURE 5 The Surgical Safety Web map from November, 2008, with 384 participating hospitals.

Harvard School of Public Health and led by Dr. Atul Gawande, an endocrine surgeon from Boston's Brigham and Women's Hospital. To improve surgical care worldwide, the team created a checklist (19 steps to be completed during any operation), tested the checklist in clinical settings in eight hospitals worldwide, and found that using the checklist reduced major complications from surgery by 36% and the rate of death resulting from surgery by 43% (Gawande 2009). These study results were sufficient justification to begin a worldwide dissemination effort to get as many hospitals as possible to use the checklist. In the summer of 2007, members from the SSSL team came to the CGA help desk to inquire about adding a mapping component to their project.

The team wanted a Web map displaying (1) distinct markers at the hospital locations that had registered to use the checklist and hospitals that were actively using the checklist and (2) a thematic map showing surgical rates per country, normalized by population (Figures 5 and 6). They wanted to

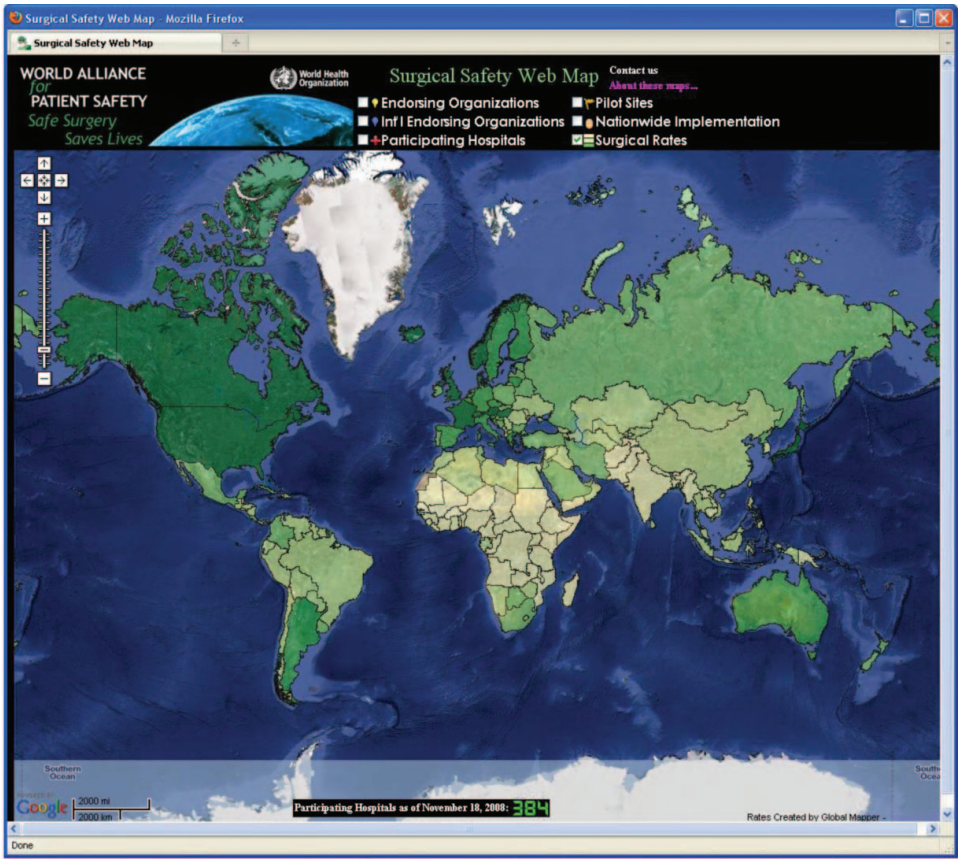

FIGURE 6 The Surgical Rates Thematic map overlay.

be able to update the map as new hospitals signed on, and wanted the map to have popup window functionality displaying the hospital name, city, and country upon a mouse click. The team wanted to use the map both internally to track progress of the checklist dissemination and for hospital information access and externally as a marketing tool to show hospital administrators, chief surgeons, and other decision makers the wide acceptance and proximity of hospitals using the checklist. The SSSL team liked the look, feel, and functionality of Google Maps, and wanted to use this as their basemap platform. In December 2007 the CGA published the Surgical Safety Web Map with the then 25 participating hospitals displayed as a KML file and the surgical rates map as a screen overlay.

The team quickly put the map to use, featuring it in presentations and referring potential checklist adaptors to it as their dissemination effort continued. New participating hospitals were sent to the CGA for inclusion on the map each month, and a new KML file was generated and published.

Team members also started using the map as a way to view the current status of the dissemination effort. Members could use the map to quickly get an idea of the participating hospitals in their area of interest, which are available worldwide to anyone with an Internet connection. This use of the map prompted a second phase in the Web map development. The team wanted to be able to update the map themselves as soon as hospitals signed up for the checklist. Also desired was an ability to access more information about the hospitals through the map such as the number of beds, number of physicians, contact person information, and whether devices such as pulse oximiters were available. The SSSL team wanted to be able to view and update this hospital information, yet restrict information access to the general public to the hospital name, city, and country. At this point the total hospital count was nearing 400, which exceeded the number of points that could be rendered on a Google Map mashup at once. Also, it was difficult to discern individual hospitals on the map in cities with many participating hospitals in close proximity. Because of this, the team wanted a list of hospitals in the viewable map area to be displayed interactively.

To accommodate the new Web map functionality requests, CGA transferred the hospital data layer into a PostgresSQL database, geospatially enabled with the PostGIS module. Mapserver was used to render the hospitals, and hospitals were symbolized according to whether they were registered or actively using the checklist. To enable the SSSL team to upload new hospitals, a data import program was written in Java, and the SSSL team was trained on how to geocode new hospitals using Batchgeocode.com to obtain longitude and latitude coordinates. The coordinates found for each new hospital were saved in columns in an MS Excel template that matched the PostGIS database schema. Once the template was populated with the hospital information and coordinates, it was saved into .csv format and uploaded to the database through the data import program. This upload program requires a login to access, ascertaining that the .csv file is in the right format. Once uploaded, the new hospitals immediately appear on the map, and each time the map is rendered it draws all the points in the database according to their coordinate locations. In addition, when a hospital is clicked on to identify it, a "For more hospital information" message appears in the information window that, when clicked on, prompts the user to enter a username and password. Upon authentication, a form appears listing all the hospital information in editable text forms. The longitude-latitude forms are editable as well, enabling position refinement of hospital locations. A new sidebar was added to the map, and programmed to display hospitals in the current view in a tabular fashion, enabling one to click on a hospital name in the list and see it identified on the map with its information displayed (Figure 7).

The final task was writing another Java program that allows for an export of the full database back to .csv format after logging in. This enabled

FIGURE 7 Updated Surgical Safety Web Map.

any member of the team to save out the latest version of the database, and interact with it in Excel, a database, or statistical program.

Once the new functionality was tested and rolled into the production Web map, the CGA's role of providing monthly update services was no longer needed. The SSSL team was now fully enabled to update the database and the map on their own. The database and map continue to be updated, and, as of this writing, they contain 3,924 hospitals that have signed up to use the checklist.

Enabling Health Researchers to Use GIS: The *Lancet* Publication

CGA's role of enabling public health researchers to use GIS technology is showcased in an article in the November 29, 2010, issue of *The Lancet*: "Health Professionals for a New Century: Transforming Education to Strengthen Health Systems in an Interdependent World" (Figure 8). The cover of this issue features global cartogram maps, as do several figures within the issue. The maps were produced by Dr. Ananda S. Bandyopadhyay, a

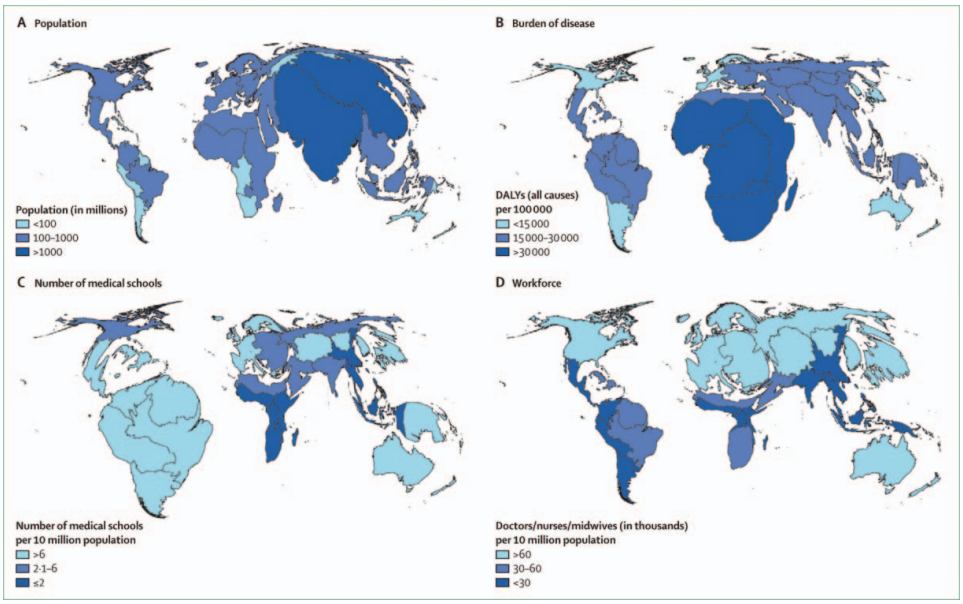

FIGURE 8 Map illustrations as published in *The Lancet*, 376 (9756), Julio Frenk, Lincoln Chen, Zulfiqar A. Bhutta, et al. Health professionals for a new century: Transforming education to strengthen health systems in an interdependent world. Pp. 1923–1957, © 2010, with permission from Elsevier.

graduate of the Harvard School of Public Health (HSPH) Master of Public Health Program. In performing his research at HSPH, Dr. Bandyopadhyay came to the CGA help desk to learn how to make maps and perform spatial analysis. Through CGA workshop training courses, the CGA GIS Institute, and several help desk sessions he became proficient in using many GIS analytical and cartographic tools. Dr. Bandyopadhyay was asked to join the *Lancet* research team, and his GIS skills enabled the production of these descriptive, informative maps, creating an unanticipated benefit to the publication that became one of its focal points.

CONCLUSION

The GIS infrastructure at Harvard, in addition to the services that the HGL, HMC, and CGA offer, provides a robust environment in which GIS can be applied to the realm of public health research. This article highlighted pertinent geographic data, the project workflow of the CGA, and case studies in public health research. These topics are prevalent in the Harvard research community today, and they are poised to grow and develop in the future.

NOTES

1. ESRI 2009 International Users Conference Proceedings Web site: http://proceedings.esri.com/library/userconf/health09/docs/monday/constructing_a_geospatial_database_of_united_states_local_health_departments.pdf
2. The Commonwealth of Massachusetts GIS Web site: http://www.mass.gov/mgis/
3. New York GIS Clearinghouse Web site: http://www.nysgis.state.ny.us/gisdata/
4. GeoNames Search Web site: http://geonames.nga.mil/ggmagaz
5. Oak Ridge National Laboratory Web site: http://www.ornl.gov/sci/landscan/
6. Demographic and Health Surveys Web site: http://www.measuredhs.com/
7. U.S. Census Bureau Web site: http://www.census.gov
8. United Kingdom Office for National Statistics Web site: http://www.ons.gov.uk/census/index.html
9. Israel Census of Population and Housing Web site: http://www.cbs.gov.il/mifkad/e-mifk.htm
10. Gridded Population of the World Web site: http://sedac.ciesin.columbia.edu/gpw/
11. United Nations Population Information Network Web site: http://www.un.org/popin/
12. PBL Netherlands Environmental Assessment Agency History Database of the Global Environment Web site: http://themasites.pbl.nl/en/themasites/hyde/index.html
13. U.S. Census Bureau International Data Base Web site: http://www.census.gov/ipc/www/idb/index.php
14. World Health Organization Data and Statistics Web site: http://www.who.int/research/en
15. Demographic and Health Surveys Web site: http://www.measuredhs.com/
16. UNICEF Childinfo Monitoring the Situation of Children and Women Web site: http://www.childinfo.org
17. UNICEF Childinfo MICS Web site: http://www.childinfo.org/mics.html
18. Centers for Disease Control and Prevention Web site: http://www.cdc.gov/nchs/
19. Centers for Disease Control Behavioral Risk Factor Surveillance System Web site: http://www.cdc.gov/brfss/maps/gis_data.htm
20. Centers for Disease Control Behavioral Risk Factor Surveillance System City and County Data Web site: http://apps.nccd.cdc.gov/BRFSS-SMART/index.asp
21. Centers for Disease Control WONDER Online Database Web site: http://wonder.cdc.gov
22. Health and Medical Care Archive of the Robert Wood Johnson Foundation Web site: http://www.icpsr.umich.edu/icpsrWeb/HMCA/index.jsp
23. New York City Department of Health and Mental Hygiene Web site: http://www.nyc.gov/html/doh/html/survey/survey.shtml
24. Los Angeles County Public Health Assessment Web site: http://www.publichealth.lacounty.gov/ha/HA_DATA.htm
25. The University of North Carolina Population Center China Health and Nutrition Web site: http://www.cpc.unc.edu/projects/china
26. National Family Health Survey of India Web site: http://www.nfhsindia.org/odata.html
27. U.S. Census Bureau American FactFinder Web site: http://factfinder.census.gov/home/saff/main.html?_lang=en
28. GeoLytics Web site: http://www.geolytics.com
29. Social Explorer Web Site: http://www.socialexplorer.com/pub/home/home.aspx
30. Inter-University Consortium for Political and Social Research Web site: http://www.icpsr.umich.edu/icpsrWeb/ICPSR/
31. ESRI Community Analyst Web site: http://communityanalyst.esri.com
32. All China Data Center: China Data Online Web site: http://chinadataonline.org/
33. Government of India Ministry of Home Affairs Census of India Web site: http://censusindia.gov.in
34. U.S.G.S. EarthExplorer Web site: http://edcsns17.cr.usgs.gov/NewEarthExplorer/
35. GeoEye GeoFUSE Online Maps Web site: http://geofuse.geoeye.com/maps/Map.aspx
36. Food and Agriculture Organization of the United Nations Web site: http://www.fao.org/corp/statistics/en/
37. United States Environmental Protection Agency Envirofacts Web site: http://www.epa.gov/enviro/index.html

38. U.S. Environmental Protection Agency Geospatial Data Download Service Web site: http://www.epa.gov/enviro/geo_data.html

39. U.S. EPA EnviroMapper Web site: http://www.epa.gov/emefdata/em4ef.home

40. U.S. Geological Survey Board on Geographic Names, Foreign Names Web site: http://geonames.usgs.gov/foreign/index.html

41. U.S. Geological Survey Board on Geographic Names, Domestic Names Web site: http://geonames.usgs.gov/domestic/index.html

42. GeoNames Forum Web site: http://forum.geonames.org/gforum/posts/list/5.page

43. Center for Geographic Analysis at Harvard University Web site: http://gis.harvard.edu

44. Best Practical RT: Request Tracker Web site: http://bestpractical.com/rt

45. Center for Geographic Analysis at Harvard University Google Maps Tutorial Web site: http://maps.cga.harvard.edu/gmaps_instruction/

46. WorldMap Alpha Web site: http://worldmap.harvard.edu

47. NetApp Web site: http://www.netapp.com/us

48. American Academy of Pediatrics Prevention and Treatment of Childhood Overweight and Obesity Web site: http://www.aap.org/obesity/families.html?technology=1

49. National Institute of Health Research Portfolio Online Reporting Tools Project Information Web site: http://projectreporter.nih.gov/project_info_description.cfm?aid=7962059&icde=7202284

50. The Commonwealth of Massachusetts GIS Web site: http://www.mass.gov/mgis/

51. World Health Organization Safe Surgery Saves Lives Web site: http://www.who.int/patientsafety/safesurgery/en

REFERENCES

Frenk, J., et al. 2010. Health professionals for a new century: Transforming education to strengthen health systems in an interdependent world. *Lancet* 376 (1): 1–3.

Gawande, A. 2009. *The Checklist Manifesto: How to Get Things Right*. New York: Henry Holt.

Guan, W. W., B. Burns, J. L. Finkelstein, and J. C. Blossom. 2011. Enabling geographic research across disciplines: Building an institutional infrastructure for geographic analysis at Harvard University. *Journal of Map & Geography Libraries* 7 (1): 36–60.

Higgs, G. 2004. A literature review of the use of GIS-based measures of access to health care services. *Health Services & Outcomes Research Methodology* 5 (1): 119–139.

Introduction to the "Working Group Report on Future Research Directions in Childhood Obesity Prevention and Treatment." 2007. *National Heart Lung and Blood Institute, National Institutes of Health*. http://www.nhlbi.nih.gov/meetings/workshops/child–obesity/index.htm

Jerrett, M., S. Gale, and C. Kontgis. 2010. Spatial modeling in environmental and public health research. *Int J Environ Res Public Health April* 7 (4): 1302–1329.

Koplan, J. P., C. T. Liverman, and V. I. Kraak (Eds.). 2005. Institute of Medicine of the National Academies. *Preventing Childhood Obesity*. Washington, DC: National Academy Press.

Lewis, B., and W. Guan. 2010. Jump-starting the next level of online geospatial collaboration: Lessons from AfricaMap. *Advances in Web-based GIS, Mapping Services and Applications*. Li, Dragicevic, and Veenendaal (Eds). Boca Raton, FL: CRC Press.

Martens, P., and L. Hall. 2000. Malaria on the move: Human population movement and malaria Transmission. *Emerg. Infect Dis. March/April* 6 (2): 103–109.

Ogden, C. L., M. D. Carroll, L. R. Curtin, M. A. McDowell, C. J. Tabak, and K. M. Flegal. 2006. Prevalence of overweight and obesity in the United States, 1999–2004. *Journal of the American Medical Association* 295 (13): 1549–1555.

Puett, R. C., J. E. Hart, J. D. Yanosky, C. Paciorek, J. Schwartz, H. Suh, F. E. Speizer, and F. Laden. 2009. Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. *Environmental Health Perspectives November* 117 (11): 1697–1701.

Puett, R. C., J. Schwartz, J. E. Hart, J. D. Yanosky, F. E. Speizer, H. Suh, C. J. Paciorek, L. M. Neas, and F. Laden. 2008. Chronic particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. *American Journal of Epidemiology* 168 (10): 1161–1168.

Yanosky, J., C. Paciorek, and H. Suh. 2009. Predicting chronic fine and coarse particulate exposures using spatio-temporal models for the Northeastern and Midwestern US. *Environ Health Perspect* 117 (1): 522–529.

Yanosky, J. D., C. J. Paciorek, J. Schwartz, F. Laden, R. C. Puett, and H. Suh. 2008. Spatio-temporal modeling of chronic PM10 exposure for the Nurses' Health Study. *Atmos. Environ.* 42 (18): 4047–4062.