Integrating a Geographic Information System with Electronic Medical Records

Kwatei Jones-Quartey, Scott Petricig, Rita Weinstein, and Kevin Gravesande

Abstract
The integration of Geographic Information Systems (GIS) and Electronic Medical Record (EMR) systems has been predominantly implemented in an epidemiological context. This work aims to integrate a GIS with the EMR system of a major urban medical center, as an efficient administrative tool for daily use rather than a means of tracking disease. With this GIS, medical personnel may create maps to locate patient addresses. Furthermore, they may view socio-economic and environmental variables relating specifically to the patient’s residence area. The paper details methodology and materials utilized, as well as the path to expansion of the system.

1. Introduction
Geographic Information Systems (GIS) have an untested potential when integrated into Electronic Medical Record (EMR) systems: to assess and assist in providing real-time healthcare service to clients. An integrated system can allow clinicians to explore, identify, and implement preventive measures to inhibit the spread of diseases, thus producing a powerful tool.

The integration of a GIS into an EMR system presents the following principal problem—the two systems are completely unrelated. The former uses geographical data such as coordinates, layers, shapes, and other attributes to produce maps and location data. The latter presents a person’s medical profile maintained at a medical institution or at a health practitioner’s office. Yet since a person’s health is very much a function of his/her environment, a GIS does have a role to play from the standpoint of geography. Indeed this role may be very significant, as a review of literature pertaining to this subject shows; furthermore, “geography” takes on an enhanced frame of reference. Certainly, the justification for such a merging of the two data systems relates wholly to how a GIS might enhance the typical EMR.

In New York City, medical practitioners at all levels of the health profession encounter various environmental difficulties relating to the plying of their profession. We may enumerate the most important of these as follows: location, demography, economy, and environment.

Location: for the medical health professional, New York’s urban layout presents various navigational challenges. Simply put, at a very basic level, the health professional needs to instantaneously pin-point a required location with the simple entry of address data and a click of a computer mouse in order to retrieve certain data. But location alone says nothing to the health professional as far as a medical profile is concerned. Therefore the raw input of an address receives a contextual element when combined with the following three factors.

Demography: New York presents a multifaceted demographic layout in which the distance of 2 blocks may present a complete transformation of social and ethnic characteristics, all important factors for compiling a more complete medical record relating to an individual and his/her background.

Economy: a person’s economic status and background may have direct bearing on health status. Factors such as access to primary health care, availability of medical insurance, and ability to afford medication determine more completely a person’s overall health picture. For example: Does the patient use hospital emergency rooms as a primary health care option? How does the patient in question pay for medical attention?

Environment: In relation to health exigencies, this variable is perhaps the most important of all. We frequently read of the conflicts over the placement of waste facilities in a community between, on the one hand, residents of various communities throughout New York, and on the other, city government. Why? Because the issue of air quality and its effect on respiratory disease is an issue that even lay-persons understand, since they see for themselves how air quality affects breathing.

In response to these four variables which health professionals continually encounter, this project attempts to integrate the environmental elements of a customized GIS into an EMR system by meeting the following requirements.

Location—providing a map: Using GIS resources, an end user should be able to enter a patient’s address, have it translated into geo-spatial coordinates, and then receive the instantaneous production of a map on a computer screen, pinpointing the location of the address. The map should also allow subsidiary searches which will locate any indexed establishments in the vicinity such as, restaurants, social services, parks, libraries, civic institutions, and the like. Three tables will be used to enhance the map. The tables will be readily available.
through a mouse click, and will assist in constructing a complete medical profile. A descriptive summary of these tables follows:

**Demographic factors:** This table, downloaded from the US Census Bureau, should reflect a summary of principal demographic factors relating specifically to the census tract in which the patient resides.

**Economic factors:** Also attached to the generated map would be a table of economic factors, again relating specifically to the relevant census tract.

**Environmental factors:** The US Department of Environmental Conservation maintains various air quality monitoring stations throughout the New York Metropolitan area [8a]. By calculating the geographic distance from a patient’s geo-location, the GIS should be able to find the nearest monitoring station to the patient’s address. With this information air quality statistics may subsequently also be presented with the map, a useful factor for patients with respiratory problems. The air quality data included in this system is the PM2.5 data collected by the New York State Department of Environmental Conservation (DEC). This data represents particulate matter existing in the air that is smaller than 2.5 microns – very small particles that can affect everything from lung functions to health of the heart [8b, 16]. Hopefully, healthcare providers can use this data to better treat patients and prevent health problems.

For this study, the four factors described above provide the means of integration between GIS and EMR. The following quote from the World Health Organization’s website provides a succinct characterization of a common form of this integration: “Geographic information systems (GIS) provide ideal platforms for the convergence of disease-specific information and their analyses in relation to population settlements, surrounding social and health services and the natural environment.” [17]. Although this quotation is coined in the context of non-urban environments, and probably applies to countries whose development level is lower than that of the US, we may still extrapolate the characterization and apply it to New York City. The specific focus of integration differs from that envisaged by this work, yet it attests to a frequent pairing of the two systems—GIS and health records.

2. Background

A useful definition for a GIS is the following: “a computer system for the input, editing, storage, maintenance, management, retrieval, analysis, synthesis, and output of geographic, or location-based, information. In the most restrictive usage, GIS refers only to hardware and software. In common usage, it includes hardware, software, and data” [2].

Contrast this with a working definition for EMR, one given by the Patient Record Institute (USA): “a repository for patient information within one health-care enterprise (e.g. within one hospital, author’s note) that is supported by direct computer input and integrated with other information sources” [13].

Examining samples of current literature, it turns out that the integration of GIS with health systems is not new at all and may be traced back as far as 1854 when an English physician, Dr. John Snow, used a map to trace the outbreak and spread of cholera in London [7]. The previously mentioned World Health Organization [14] has a complete web site devoted to the use of GIS in the public health arena. Some of the objectives cited on the website are the following:

- “Determining geographic distribution of diseases
- Analyzing spatial and temporal trends
- Mapping populations at risk
- Stratifying risk factors”

Here in the US, the National Center for Health Statistics [8] also maintains a huge web site devoted to the use of GIS to track all manner of public health statistics, featuring maps of mortality, many different datasets, and in general a wide variety of GIS resources. There is therefore a healthy precedent for the integration of maps and mapping data with data generated for medical records.

A literature search on GIS integration with health systems produces results heavily skewed towards the epidemiological (the incidence, distribution, and control of disease in a population) use of GIS in large health systems, especially public health. For instance, Richards, Croner, et al [12] provide a comprehensive examination of trends, advantages, disadvantages, and future possibilities for the integration of GIS with public health systems. In their account, these systems harness GIS technology principally for disease mapping, for example: the mapping of areas which present high-risk disease factors, e.g. lead poisoning as it affects the newly born; the pinpointing of communities in which injuries due to accidents are unusually elevated; the mapping of areas evidencing a high level of cancer cases. A trenchant current example of disease mapping occurs in the use of Google maps to document the spread of the 2008 flu. Internet users can click on a “flu trend” map of the US and receive an advisory of the level of flu activity in a particular state.

An organization called MEDIAS-France implemented a GIS earlier this decade to help track areas of large mosquito populations to help prevent mosquito-borne epidemics in Africa [5]. In an attempt to assist with lead paint remediation, local governments in one area of Illinois took data showing older, low-income housing, and previous lead-poisoning cases. Subsequently, they used a
GIS system to plot and analyze this data, creating maps showing areas of possible high lead concentrations [5].

In a project more closely related to our work, researchers at New York University have used GIS software to analyze the location of industrial and manufacturing zones in relation to schools in the Bronx. The purpose of the study was to try to find links between air quality in these areas and asthma in school-age children [6]. These projects were all recent and used GIS software produced by Environmental Systems Research Institute, Inc.

Phillips, Kinman, Schnitzer, et al. [11] analyze the use of GIS technology to survey health care access in a rural community of Missouri. Germane to the objectives of the current project, the article delineates the use of a GIS to geo-code the addresses of the survey participants and link such geo-coding to the boundaries of US census tracts, in the same manner envisaged by our project. Even more interesting is the description of datasets used in the survey: “…3314 patient records of the CHC for 1998. Each record contains the home address, number of visits, age, sex, payment method, and household income.” In effect this is an EMR system without the electronic aspect. In this particular referenced application, GIS and medical data systems are integrated but only for a specific investigative purpose.

In contrast to these other uses of GIS, this work seeks to integrate a GIS into an EMR system as a permanent and practical feature which would provide more complete medical profiles. Planning this effort required the evaluation of other GIS applications implementing the Google Maps API. One such interesting example is the website, Geocaching - The Official Global GPS Cache Hunt Site [3]. This site uses Google Maps technology along with its own “layers” to provide information to its viewers. For the current work, future goals include integrating features used by the Geocaching website, such as placing data directly in informational pop-up “bubbles” on a rendered map. In a fashion similar to the Geocaching site’s approach, this study positions location-specific-map data next to a rendered map.

3. Project Relevance

The focus of this work is to embrace the end-user’s standpoint; therefore, the system should have easily utilizable functions, practical relevance for day-to-day use, and seamless integration with the EMR system. In effect, the GIS should provide a practical tool easily utilizable by all strata of the medical profession. Clearly then, a different approach is necessary for the integration of GIS and EMR systems since the requirement is neither epidemiological nor related to an investigative survey. The link between the two systems is the patient, and the actual point of intersection is very basic—the patient’s address.

4. Methodology and Material

As shown in Figure 1, the implemented GIS system devolves through a number of dynamic ASP.NET web pages interacting with two databases. The system is initially triggered by the input of a patient’s address passed from an EMR. Examining Figure 1 from an analytic standpoint, the process unfolds as a series of increasingly specific data events.

![Figure 1: Web Page Sequence.](image)

For instance, once a street address has been passed to the page “street_conv.aspx” (#2), this page queries the SQL server. Using the street address, a stored procedure in the SQL server converts the address to its corresponding census tract. Both census tract and street address are then passed to the page “geocode.aspx” (#3). At this point this page, using an Ajax routine, in turn extracts latitude and longitude coordinates. Thus at each stage geographic data provides the input which triggers dynamic responses. As to the two databases, although they operate in completely disjunctive spheres, they do interact indirectly through the data passed back and forth by the web pages.

The geographic aspect of the system does not limit spatial data (in other words all data that could be used to produce or enhance a map) to purely mapping functions. Take for instance the web page, “street_conv.aspx”. This page is called up when a patient’s address is passed to it. To map that address, the system has all it needs; however, for other aspects of
functionality, one important piece of information is missing: the census tract [Figure 2, 14].
(The US Census bureau organizes the geography of the US into units that begin from the very broadest, e.g. Nation, and progress into steadily smaller and more detailed units. The census tract is one of the smaller units within this geographic scheme.)

![Figure 2: US Census Bureau Geography](image)

The census tract is useful for the extraction of US Census Bureau datasets, so in processing a patient, the first step is to convert address data to census tract. Although one may navigate to the Census Bureau site and access their converter, this would be a cumbersome procedure. Since the goal of the project GIS is end-user simplicity, we created a custom SQL stored procedure to avoid the need to access an outside website. The conversion is performed by passing variables to a SQL stored procedure which itself runs against a series of census geography tables (see Appendix A, Figure 3 for an explanation of this procedure) [15].

A second stored procedure called from the page “pm25.aspx” uses the new GEOGRAPHY data type in SQL 2008 to calculate the closest air-quality monitoring station to the address in question. Instead of calculating by street mileage, all addresses are converted to latitude-longitude coordinates, with an ensuing calculus using these units. The page “geocode.aspx” retrieves latitude/longitude coordinates from the patient’s address using a Javascript function. The stored procedure is initiated by clicking the <Pollution Levels> link on the geocode.aspx page. In the database, a table, AQ_stations, contains GEOGRAPHY data type coordinates (AQCoordinates) for the 3 monitoring stations in Brooklyn. A SELECT is run using the geo-spatial method STDistance. It calculates the distance between the patient’s address coordinates and the coordinates of each of the stations. It selects the shortest distance and chooses the station based on that parameter. PM2.5 data is then pulled up for that particular station.

The use of these two stored procedures to manipulate geographical data illustrates the need to think of a GIS in an enhanced manner, i.e. the mapping data may be used in other ways exotic to purely map production.

The linchpin of the system is the ASP.NET page “geocode.aspx” which connects the two databases in the following manner:

1. It receives the patient address, submits it to a Javascript routine, and ultimately produces a map from the Google Ajax mapping API. [4]
2. It receives the Census tract number which enables a call to the correct tables for production of demographic and economic data.

One of the challenges that is encountered while working with GIS is ensuring that data is accurate and consistent. When comparing two sets of coordinates, for example, it is important that the two sets are using the same coordinate systems and that measurements are of the same scale. Coordinate systems, more formally known as geodetic datum, measure a specific point’s location on Earth relative to a static starting point. Arguably the most commonly used geodetic system today is WGS 84 (World Geodetic System, developed in 1984). WGS 84 is the system currently used by the Global Positioning System, and is also used by Google Maps [1, 9] These credentials make the system perfect for the scope of this project. In addition, Microsoft SQL Server 2008 uses WGS 84 as its default spatial coordinate system for the Geography data type, as noted by the inclusion of the parameter ‘4326’ in the declarations of variables of that data type in the code for this system.

The Google Ajax API is a service which accesses the vast Google mapping database and enables anyone to produce a map on a web page. One simply signs up for an API key, and once approved, the key is furnished with a sample script—one that may be placed within the <body> section of an HTML page and which will produce a map as soon as the page is loaded. For this project we modified the Google script to allow the passing of an address from a form on the page. Since the Google Ajax API uses asynchronous Ajax technology, the page receives only a partial refresh, in which the map is called up and situated in a pre-designated region of the web page.

To further illustrate the Ajax technology, three separate pages (dem.aspx, csv.aspx, pm25.aspx) may be called up through links from the page “geocode.aspx”. They are subsequently deposited into a <div> element situated next to the map region. On each of these pages there exists nothing but a table filled by data from the SQL 2008 database using the tract number as a criterion.
They are filled by tract-based datasets downloaded from the Census bureau and subsequently formatted—a pair for each tract (economic, and demographic). In the case of “pm25.aspx”, a connection is made to the SQL database, calling a stored procedure to determine the nearest air quality monitoring station, and pulling that data from the proper SQL table.

Once again we emphasize that geographical data is used in a broad-based manner through the mapping of street addresses to their equivalent census tracts. A multitude of information is available from the census bureau using the tract as the geographical designation; the economic and demographic tables are just one example of what is available on the census bureau website. But the crucial technique involved is to convert street addresses to geographic coordinates and census tracts. Once that operation takes place, it opens up a world of dataset possibilities.

**5. Results**

The GIS is modeled as shown in the Data Flow Diagram (Figure 4) and the Activity Diagram (Figure 5). These diagrams show the basic layout of the system as visualized in the beginning stages of project planning.
A second issue concerns access to the New York State DEC PM 2.5 fine particulate monitoring stations located in Kings County. The New York State DEC website is comprehensive in its organization, allowing for the production of many types of reports, however there is no pathway to establishing a continuous feed of the PM 2.5 data such that it can be automatically collected and saved in the project database. All downloads must be accomplished by a manual process in which report type, data type and time period must be specified. A downloadable file is then made available for which the receiving location must be specified. The system’s team has developed a Microsoft C# .NET application that can automatically download properly formatted data every hour. However, it is still unknown how that data will be collected and successively placed in the SQL database.

6. Conclusion

The objectives of this effort have largely been met for the construction of a prototype GIS which may be integrated into an Electronic Medical Record system. The system has been designed to be modular in order to allow the integration of various types of useful data, in addition to what is already presented by the system. It is designed to require little effort on the part of the end-user (medical providers), and does not require a great deal of modifications to existing EMR systems. The GIS is intended to be practical and present localized data on demand.

Presumably most commercial EMR systems may be customized to allow the placement of custom buttons or other objects which can be used to call up the different ASP.NET project web pages. For this GIS, only one is required: a link or button to pass the patient’s address to the page that begins the processing—"street_conv.aspx".

As to the outlook for expansion of the system, there are many avenues. With greater resources, a commercial mapping software program, such as the ArcGIS system developed by ESRI, would allow greater mapping flexibility for the construction of custom maps. For example, census tract maps relating to each patient’s address could be stored in the patient’s database record. Some effort at identifying health hazard locations by address with maps of the areas could also be an additional avenue to a more comprehensive medical picture. The Census Bureau itself is constantly updating and upgrading its resources, therefore improvement of the GIS should focus on making it truly modular so it may respond to changes in census datasets.

Health professionals can and will benefit from this work which can influence the progress of health surveillance, environmental health assessment and the geographic allocation of health resources in Brooklyn while serving as a model for expansion elsewhere.
REFERENCES


[14] US Census Bureau: "Designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions at the time of establishment, census tracts average about 4,000 inhabitants.” http://factfinder.census.gov/home/en/epss/glossary_c.html #census_tract

[15] US Census Bureau: These tables are called Tiger files—“ TIGER ® is an acronym for the Topologically Integrated Geographic Encoding and Referencing (System or database).” Source- http://factfinder.census.gov/home/en/epss/glossary_t.html


Street to Tract stored procedure description: Example: 100 Willow St.

1. Run Willow against field \(1\) FULLNAME (all street names in Brooklyn) for match.

2. Divide Street # by 2 to determine correct side of street. Odd = 'L' or even = 'R'; in this case result is 'R'.

3. Search for closest address range by matching TLID \(2\) field in both “edges” and “addr” tables with following parameters:
   - SIDE \(3\) = 'R'; FULLNAME \(1\) = Willow; 100(street #) \(4\) \(>=\) FROMHN \(4\) \(\leq\) TOHN \(5\) (these are address ranges)

4. Select TRACTCE00 \(6\) FROM faces by matching edges.TFID \(7\) field (this is the boundary of an address unit) to faces.TFID \(8\) and this will retrieve the correct tract for the street address.

Fig 3: Explanation of street to tract conversion: using TIGER files from Census Bureau stored in Project db (they contain detailed geo-info for Brooklyn)