MAPPING HEALTH ON THE INTERNET

MAPPING HEALTH ON THE INTERNET

By

JOHN MACLACHLAN, B.Sc. (Hons.)

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Master of Arts

McMaster University

MASTER OF ARTS (2002)
(geography)

McMaster University
Hamilton, Ontario

TITLE: Mapping Health on the Internet

AUTHOR: John Maclachlan, B.Sc. (McMaster University)

SUPERVISOR: Dr. M. Jerrett

Number of pages: ix, 153

ABSTRACT

This thesis is situated in the converging fields of geographic information systems (GIS), multimedia applications on the World Wide Web, and public health surveillance. Recent technological advances in multimedia and GIS software have enabled public health organizations to distribute health maps over the Web, potentially allowing remote users to perform queries and analytical operations. Both static maps and interactive analysis may improve public health through targeted health surveillance and subsequent interventions. The thesis reviews privacy, technological, and cartographic issues related to distributing health data over the Web and evaluates 30 existing sites, allowing for a representative sample of the current state of Web health mapping and suggestions for future health-GIS sites. Most of the sites reviewed were technically sound, but offered little or no opportunity for users to design maps and carry out analyses. The few interactive sites now online allowed for simple queries. Based on ongoing research, we will demonstrate a prototype site that allows for increased interactive use by remote users. Research shows public health officials are concerned about costs involved in incorporating new technology, such as GIS, into their work. The prototype website was created using free software to minimize costs and focuses on asthma, socioeconomic and air pollution data from Hamilton, Ontario. Upon completion the prototype site was tested by target users during a focus group at McMaster University. Survey results from the focus group reiterate findings from the literature. The majority of respondents are interested in incorporating GIS, openGIS and spatial data, but they are concerned of the costs involved

with new software. Complete results from the survey administered and how the results led to changes in the original prototype are documented.

ACKNOWLEDGEMENTS

First of all I need to thank Dr. Mike Jerrett, my supervisor. Throughout the last few years he has helped me improve my writing and researching skills dramatically. I have had the opportunity to present my research findings to audiences in Montreal and Los Angeles to good reviews. At times I may have not seemed like I appreciated the assault my work took with his red pen, but I really did.

Many people at McMaster also need acknowledged for helping get through this. Tom Abernathy and the staff at CHWPIN assisted in organizing the focus group and were always available for a helpful idea. Martin Bunch helped keep this project on track with helpful edits in the early going. Sue Vajoczki was always around for me to bounce an idea off. The office staff, Nancy Brand, Janice Wade, and Ann Wallace made sure I was always well fed and kept in good supply of mints and chocolates. Mike Law who gave me my first real lessons in computer programming (and only rolled his eyes about 1,000 times). The GISHEAL clan of Talar Sahsuvaroglu, Dana Wilson, Chris Giovis, Jason Morrison, and Mike Buzzelli (who just comes in to use all our paper) helped me.... ummm.... alright lets be honest the GISHEAL clan does little to help anyone get work done, but they are fun. I could go on forever like this so let me just say thanks to Ron Buliung, Pat Deluca, Carolyn Eyles (who also helped keep me fed), Steve Manske, and Deane Maynard.

My family who really has no idea what I am doing at University but could not be any more encouraging. I am sure at times they just think it is all crazy but they never tell me that.

Heather Stuart encouraged me to start my Masters degree and has been there every step of the way with me. I am eternally grateful for her help.

That said there is another group of people that distinguished themselves as true friends to me throughout the last two years. Just as Heather helped me out, Amy Williamson, Deb Moore, Jackie Young, John Lea, Patty Meyer, Krista Chomicki and Emmanuelle Arnaude all had there part in helping through the last two years.

Lastly, I want to dedicate this thesis to my Nana, Marie Owenson. She past away December 9th, 2001 after a 9 month stay in the hospital. I visited her everyday I could and while I missed out on a lot of sleep and the odd deadline I don't regret one visit. She always encouraged me and was there if I needed help of any kind. I dedicate this thesis to her.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Research Problem

Along with the prolific growth of health information on the Internet (Jadad 1996), a powerful new technology has emerged, geographic information systems, or GIS. GIS is a spatial analysis system for the organization, storage, transformation, retrieval, analysis, and display of data (Aronoff 1989, Demers 1998) where attribute location is considered important (e.g. the incidence of a specific health condition or disease in relation to a pollution source). Viewing health data in map or cartographic formats can assist health surveillance professionals and researchers with identifying possible associations between disease and risk factors related to disease pathogenesis. There are many examples where GIS has been used successfully in health studies (see Moore and Carpenter 1999 for a recent review). The next step in the evolution of the GIS and health partnership involves the Internet. Allowing GIS to be accessed and used over the Internet without requiring the user to download software is a new technology referred to as 'openGIS.' A recent national user needs assessment conducted for the National Health Infostructure indicated a high demand among users from multiple areas of health surveillance for more advanced internet mapping capabilities (Gosselin et al 2001). To date, however, integration of GIS and health on an Internet platform (openGIS) has been limited.

1.2 Research Context

This research was funded by the national Center of Excellence: Geomatics for Informed Decisions (GEOIDE). The GEOIDE organization supports research programs with strong potential for fostering economic and social development in a networking environment. The objectives of this research stem from the recent national user needs assessment. The research will break down into four major objectives:

I. Review the literature to determine trends and issues in openGIS?

II. Document the current state of openGIS technology

III. Developing the data frame and technological prototype.

IV. Assessment of the prototype by target users.

How can a prototype site address these needs and improve current applications? This question led to the series of thesis objectives. The review of the literature and assessment of current applications is a compilation of many tasks. It involves accumulating information from numerous sources and compiling the information to identify trends. The literature review contributes to this research by: (1) reviewing literature on health mapping issues and applications on the Internet; (2) identifying important issues related to health web design, especially those related to privacy protection, cartographic presentation of the data, and technological limitations; (3) reviewing existing health surveillance websites to determine the current state of the applications; and (4) determining the needs of users based on recent assessments conducted for the National Health Infostructure and consultation with local health officials.

Tasks also stem from the development of the data frame and technological prototype. From this objective this thesis aspires to (1) develop a typology of existing health mapping application to guide the development of a prototype website; (2) create a computer interface capable of displaying and analyzing health data for public health professionals and lay audience; and (3) develop a prototype website for respiratory and environmental health data.

Within the last objective, 'Assessment of the prototype by target users,' the following tasks were achieved after the completion of the prototype: (1) allow health professionals access to the prototype website with the intention of receiving feedback through surveys, and (2) review feedback and alter prototype to meet the suggestions.

1.3 Spatial and Survey Data

Data for the prototype site include results from a survey concerned with adult respiratory health in Hamilton, Ontario (Manfreda et al. 2001), pollution data collected from monitoring stations operated by the Ontario Ministry of the Environment, hospital admissions data from the Central West Health Planning Information Network (CWHPIN), and results from the 1996 Canadian census.

Testing of the prototype website by health professionals and other relevant parties occurred November 2nd, 2001 in an organized focus group. The twenty focus group attendees represented a range of public health interventions/services, such as direct health services and knowledge development. Also represented were people from the GEOIDE project and multimedia experts. All attendees were asked to fill out a survey gauging the prototype's effectiveness and their familiarity with GIS technologies. The data collected was used to adjust the prototype to meet user needs.

1.4 Research Contribution

Only a handful of studies have been done in relation to developing GIS effectively on an Internet platform (Stevenson et al 2000, Sieber et al 1995, Plewe 1997, Peterson 2000) and even fewer have dealt with health specifically. This research compliments previous work by applying techniques gathered by surveying many websites that use openGIS while taking into account ideas from scholarly publications. With recent trends in GIS leaning towards openGIS McKee noted "Downloading and installing software and data is the old way. The new way is using remote software and data via the web." (McKee 2001 pg. 6)

The background research for the prototype website covered many academic disciplines. All questions on how to produce an effective openGIS system were grouped into the major categories of confidentiality, cartography, and technology. Ultimately this research will enhance understanding of the interface between technological development of openGIS and the societal use of information generated for public health surveillance.

1.5 Chapter outline

This thesis is comprised of 6 chapters. The following chapter (*Research Context, Literature Review, and Current Application Review*) examines the literature related to openGIS by addressing unsolved broader issues concerning cartography, privacy and technology. In addition to examining the literature, the four main distribution models of openGIS are explained and illustrated.

Chapter 3 (*Data and Methods*) addresses the current application review which will identify what is currently available in openGIS through a critical appraisal of representative websites. Criteria for the appraisal were derived from the literature review. This is followed by detailing the development of the openGIS prototype and how it fits into a larger, ongoing, website project. This is followed by methods of data collection, both survey and geostatistical. The methodology involved in the creation of the prototype, the computer languages used, and decisions involving functionality are justified. Lastly, this chapter includes the development of a survey issued to the initial users of the prototype.

Chapter 4 and 5 describe the results from the assessment of current applications, improvements to the prototype, and the survey findings. Within Chapter 4 (*Results*) respondent's perceptions of the prototype website are disclosed. This feedback helped determine what needed improvement on the prototype. The improvements implemented to the prototype are explained and illustrated. In addition, the chapter includes results from the current application review, which showed that few websites effectively use advanced openGIS methods. Chapter 5 (*Discussion*) synthesizes key findings of the thesis and discusses how this research has advanced knowledge in the fields of Geography and Public Health. These key findings are then scrutinized against potential influences to the results.

In the final chapter (*Conclusions*) the major findings are summarized, followed by a discussion of the major substantive, theoretical, and practical contributions of this research. This thesis concludes with recommendations for future research in openGIS.

CHAPTER 2

Research Context, Literature Review, and Current Application Review

2.1 Introduction

This chapter provides context for this research by examining literature related to broader issues of openGIS. Examination of the literature led to three general issues being identified: privacy, technology, and cartography. Issues do, at times, flow from one group to another (e.g., screen resolution affecting cartographic output).

Drawing from literature on an emerging field has limitations. Many of the articles dealing directly with openGIS focus on the technology itself and not on important issues of cartography or privacy. These issues are represented in the literature but in a less direct fashion. With issues of privacy in this increasingly digital world becoming a potential problem, literature is available. However, this is not directly related itself to openGIS. Cartography is beginning to deal with openGIS with books such as *Cartography on the Internet* (Kraack 2000) but many issues, such as colour recognition, are buried within the literature of disciplines other than geography.

2.2 Determining the Health Internet Mapping Audience

Snow's (1855) early use of maps to assess clusters of cholera and gain clues about possible causal factors that led to a deadly outbreak in London established cartography, and spatial analysis, as a powerful tool for analyzing patterns of health and disease (Curtis and Taket 1996). Since that time many advances in cartographic techniques and spatial analysis (Cliff and Haggett 1988) have occurred. Also, GIS has computer-based tools to address broader health issues (Gatrell and Loytonen 1998). With the advent of the Internet and the World Wide Web (WWW) spatial data can be communicated cartographically to numerous remote users from one source. This allows for the spatial query and spatial capabilities of GIS software to be distributed to the remote users.

Representation and analysis of disease incidence data in map form has now established itself as a basic tool within the realm of public health (Lawson 2001). The ability to plot the relative risk of disease cartographically gives public health officials a better idea of where to target programs of disease prevention. There are many examples illustrating where disease mapping has been used, and how to effectively incorporate it into a project. For example, Lawson (2001) highlights the main issues of disease mapping and biostatistics for public health workers and epidemiologists. MacEachren (1998) shows how to effectively visualize georeferenced data to the represent reliability of health statistics. When media attention in New York City focused on the mosquito transmitted West Nile Virus, disease mapping was used to promote public and health professional awareness (Miller 2001). In Sweden, mapping has been used to determine the health-related quality of life by disease and socio-economic group in the general population (Burstrom 2001). Finally climate-disease connections of Rift Valley Fever were explored in Kenya using mapping as the main tool (Anyamba 2001).

In addition to mapping disease, public health officials and epidemiologists map suspected disease predictors. Studies have mapped immunization rates (Cowell 1999), health care access (Phillips 2000), sociodemographic variables (Selvin 1998, Blake 2001), and pollution concentrations (Burnett et al 2001, Jerrett et al 2001, Pikhart H 2000, Pikhart 1997). The mapping of disease predictors in public health assists in predetermining potential problem areas, which inform surveillance and intervention programs.

Survey results collected in January of 1999 by the Information Network of Public Health Officials (INPHO) in the United States of America revealed that while several state departments had years of successful experience in GIS initiatives, several states were just beginning to use GIS their projects (Yasoff 1999). From this survey five key issues surfaced. The first of these is being a demand for web-enabled GIS. OpenGIS has the potential to allow for GIS for users in public health to have a standard platform to show spatially relevant data and provides a platform for the training of public health officials to use GIS and apply spatial analysis through on-line tutorials. A problem with GIS technology relates to time lost learning the technology and data issues which arise in implementation. Instead of using GIS as a tool to solve a problem or put forth a new hypothesis, researchers must spend too much time on learning the technology (Lloyd 2001). OpenGIS is a tool that will reduce time spent learning spatial technologies, enabling researchers to focus on substantive questions.

2.3 Confidentiality

Privacy issues are paramount in using and distributing health data. For this reason representation of health data in openGIS can be problematic. For example, line and polygon data can lead to identification of individuals if counts are small, and the use of point data to represent individuals may violate his or her right to privacy (Dillehay 1994). Increased privacy protection can be achieved, to some extent, through spatial aggregation and masking statistics.

As the popularity of openGIS increases, it will become more important to develop methods that assure individual and household confidentiality in data presentation. Availability of numerous geocoded data sets capable of pinpointing location to the level of any person's home address clearly creates many threats to personal privacy (Yasnoff 1999). Even if individuals are not identified, group disclosure in studies of environmental exposure can lead to financial risk of individuals in an exposed community, including decreased property values and increased insurance costs (Cox 1996). Without clear laws, guidelines, and standards regarding confidentiality and data release, health care providers and consumers may be unwilling to provide needed information (Croner 1996). With the problems of confidentiality it may become necessary to develop an international uniform framework that protects privacy and permits public health practice and research (Melnick 1999).

Geocoding or georeferenceing in GIS relates spatial primitives (points, lines, and polygons) or raster grids to a geographic coordinate system. This allows data to be examined and analyzed spatially, to derive topological relationships among spatial entities and to link aspatial attribute data to geographic locations (e.g. placing disease sufferers geographically using postal codes). This makes possible the exploration of questions with spatial character such as: Where do particular phenomena occur? Are there patterns to be identified? What trends occur at particular locations and their vicinity (e.g. spatial diffusion)?

While georeferencing health data can enhance the decisions made by public health officials, it makes the task of protecting confidentiality more challenging, particularly for data distributed over the Internet. If the data are publicly available, then most statistical agencies would require protection to ensure individual privacy. Although geo-referencing allows health data to be examined in relation to its spatial distribution, the issue of confidentiality arises when the geographic resolution of data is fine enough to identify fewer than four addresses. It is at this level that the data changes from being merely a research tool to a means of targeting and exposing individuals (Alpert and Haynes 1994; Goss 1995).

One solution to this dilemma relies on aggregating personal identities into groups so that the data of individuals cannot be extracted. The South Carolina Department of

Health and Environmental Control (SCDHEC) biostatistics GIS lab developed a statewide health information system capable of satisfying the needs of health researchers while maintaining spatially referenced public health confidentiality standards. The SCDHEC database contains personal health data aggregated to individual points. Confidentiality is maintained by aggregating this individual point data to larger spatial units, that is, polygons corresponding to 1990 census tracts. Census tracts were used because they are large enough to conceal individual identities and contain a volume of socioeconomic data that could be combined with aggregate vital records. For example, the maternal age data from the vital records database could be related to the female population of the census tracts, combining both data sets and allowing, for example, calculations of statistical rates.

Aggregation maintains confidentiality, but it is not without costs. Key problems that may results from data aggregation are the modifiable areal problem (MAUP) and ecologic fallacy (King 1979). The size and shape of areal units chosen is what causes the MAUP. When data is aggregated the areal units of analysis increase in size (i.e., aggregating from individual points to census level, or from census level to city level). Although larger areal units stabilize health indicators, spatial accuracy becomes more ambiguous (Nakaya 1999). Examples of this problem being reviewed are found in Fotheringham (1991) and Mayer (1983). Related to MAUP is the problem of 'ecologic fallacy.'

Ecologic fallacy occurs when studies are done using groups, not individuals (the basis of aggregation). Associations evident at the level of populations may not apply to individuals. For this reason, the ecologic fallacy is an issue that researchers need to be sensitive to in health studies (Kwok 2001). With the majority of data masking techniques based in data aggregation the ecologic fallacy becomes more of a concern.

2.3.1 Data Aggregation Methods

New methods for masking individual data are under development. Armstrong and colleagues (1997) developed a "randomized location coordinates" masking technique. Using this method an individual has an equal probability of being located within a specified distance (i.e. 0.5 km) from their true location (Armstrong 1997). This method is effective in situations where the counts are small. While it is possible to mask individual characteristics by statistical methods like those from the SCDHEC, only data suppression guarantees protection of an individual's identity within a group of people (i.e. a census tract).

Standard methods employed by most statistical agencies are to suppress data on personal characteristics when counts in areal units are below some arbitrarily low number. An example of this is the "random rounding" procedure used by Statistics Canada. For areas where the census sampled 20% of the population all counts are randomly rounded to a multiple of "5". All counts less than "10" are randomly rounded to either "0" or "10" (Statistics Canada 1999). In the geocoded world, some argue there should be an added aspect to the right to privacy. As well as having a right to be left alone, people need to have the right to control their identity (Clarke, 1994). With geographically coded data it is possible to develop profiles of individuals and neighborhoods that may or may not be true. For example, a health insurance company may wish to charge people who live within high pollution areas higher rates than people who live in low pollution areas. The insurance company would assume that every individual in the high pollution area faces higher risk regardless of individual health status. There are numerous other potential examples such as "redlining." (e.g., people are commonly denied loans because they live in areas where socioeconomic status is low) (Li 1999). In other words, individual privacy and identity could be adversely affected by geocoding, and any effort to distribute data through public sources such as websites will have to protect against violations of personal privacy.

2.4 Cartographic Graphic Format Considerations

Maps using the WWW as a medium may offer no more in terms of analytic capacity than a typical paper map or they may be partially or completely interactive. Such maps should follow established cartographic conventions (Keates 1973), but maps presented on the web also pose new challenges.

2.4.1 Types of Maps

Maps posted on the WWW have different design considerations than static paper maps. Although both types of maps can be created with the same GIS software (e.g., ArcView 3.2), web-based maps have unique needs such as simpler color schemes that will display clearly on computer monitors. Maps on the WWW may take one of three forms: static, image, and interactive (Madej 2000). All three types have different documents that must be taken into account when designed. The following hypothetical scenario illustrates each map type.

Consider a municipal community health group that wants to post a website showing the concentration of respiratory disease within census tracts in their community. The group posts a map on their website that shows disease rates within each census tract by applying a colour or shades to each range of concentration. Maps that use polygons with graded colours for display are known as choropleth maps (DeMers 1998). If users can visit the website and view, but not alter, the map, then it is referred to as a "static map." The health group then decides to add socioeconomic attributes (e.g., dwelling value, education), and exposure to pollutants (e.g., sulfur dioxide concentrations). They alter the site to allow users the option of obtaining all the attributes of a specific census tract by "clicking" on the tract with their mouse. Maps that allow users to locate "hot spots" of information are called "image maps." Finally, the group decides that a query system needs to be implemented to allow people to locate census tracts that incorporate a number of criteria. This query system would prompt the user to enter ranges of each variable they consider important and show on the map an area(s) that meet(s) their criteria. For example, a family with an asthmatic child may like to know which areas have low pollution concentrations and affordable dwelling values. This map would be considered "interactive" because the user has input into the map design through a query process.

Static maps and image maps can be served using standard HTML code and java scripts; however, interactive maps require specific software. Interactive maps have the greatest potential for sharing and analyzing health data on the web, but they are also the most complex to design and implement. As noted earlier, an interactive map allows the user to incorporate their own criteria into the map production process. By allowing the user to select query criteria, maps can be produced showing spatial data on what the user wants to see.

Maps created on the web must use colours that are considered 'web safe.' These are colours that appear on any computer screen in their intended hue (McFerdies 2000). Map colours should also promote the detection of patterns and clusters. Colours schemes that can be confused because of impaired vision and colours where appearance is influenced by surrounding colours should be avoided. In cognitive experiments colours for mapping health phenomena were tested. The best pattern recognition occurred using the diverging schemes of red/blue, orange/blue, orange/purple, yellow/purple, and green/purple (Brewer 1997). These colour pairings show differences well and can be read accurately by most map users (Brewer 2000). These colour pairings should show "hot" and "cold" sections of a thematic map, an important consideration for health maps (e.g., risk factors, high and low disease concentrations), represented by contrasting dark colours. Sections of the map where values are at or near the mean should be a neutral colour. This allows for the "hot" and "cold" areas to be easily distinguished within the map (Pickle et al. 1997).

2.5 Technological Issues

All computer hardware is not created equal and software capabilities range vastly with each package. Technological issues can range from problems of resolution to variance in software capabilities.

2.5.1 Resolution

Computer monitors typically have a screen resolution that is only a twentieth of the resolving power of printing (Piccone 1998). This means that some small clusters of variable data that would be visible on a paper map may not appear on the computer screen if presented in an identical fashion. The absolute size of the map will vary depending on the resolution of the computer screen, which can vary from 640 x 480 pixels to over 1280 x 1024 pixels, with the monitor diagonal sizes generally ranging from 13 to 21 inch (Stynes 1997). The size of a map may be important when considering the Smallest Visible Object (SVO) (Hardie 1998) and for intra-map comparison (i.e. comparing areas within a map). Changes in the absolute size of maps complicate the issue of scale. A map viewed on a 21 inch monitor at a scale of 1 cm is equal to 1 km, will have a smaller scale on a 15 inch monitor (1cm being equal to 1.4km) due to the smaller computer screen. This difference in absolute size may be the difference between a map appearing cluttered or not. Along with problems with scale there are issues of distortion. One way in which the distortion of the map may be kept under control is to allow the map to be viewed in a Portable Document Format (PDF) (Madej 2000).

Having maps available for download as PDF files ensures maps will be the same quality as when they were produced (Madej 2000). There are many advantages in using the PDF format. Colours that are used in map creation stay close to their original hues after downloading. The map will maintain a proper page size, which is important in terms of maintaining scale and minimizing distortion. Text will stay in the proper places, regardless if the computer displaying the file does not have those fonts installed or is using an entirely different operating system (e.g., UNIX). Also, downloading of maps can be prohibited in PDF. A disadvantage of the PDF format is that the opening, viewing and printing of the file must be conducted with the Adobe Acrobat Reader, Adobe Photoshop, or Adobe Illustrator. The Acrobat Reader is a plug-in available for most web browsers and a PDF file viewer can be downloaded for free from the Adobe website (Adobe 2000). While PDF offers many advantages, maps in this format limit the user to viewing maps in static form with no interactive capacity.

2.5.2 Software and Hardware

OpenGIS operates on the WWW, not the Internet itself. The Internet consists of millions of computers of various types, physically or remotely connected by many companies and communicating via several languages (called protocols). No single organization owns the entire network, although it is controlled by a series of accepted standards. Not all Internet servers are part of the World Wide Web (Plewe 1997). The WWW is one form of communication on the Internet, alongside other forms such as forward transfer protocol (FTP), e-mail, and chat. Documents on the WWW are formatted in a script called Hypertext Markup Language (HTML) that supports links to other documents, as well as graphics, audio, and video files. It is HTML's ability to function with multiple media forms that makes it so popular. This facilitates rapid movement from one document to another by clicking on 'hot spots.'

Internet mapping software is available commercially or as open-source software. Open-source software programs are independently created and freely distributed. In addition to being free of cost, the user has access to the source code (the underlying computer language compiled by the original programmers.) All developers require is that advances made by the users in the software be made openly available to all other users free of charge (Levy 1999, Dertouzos & Gates 1999). This is considered by some to be the future of computer software development (Greiner 2001, Careless 2000). This concept exists in stark contrast to the secrecy of commercial software development. Commercial software comes as a polished product that can not be altered easily, if at all. There are advantages and disadvantages to each software concept. Open-source software such as MapServer (MapServer 2001) offer more control of the final product. However, for such software to acquire the functionality of a commercially available server such as ESRI's Arc/Info Internet Map Server (ArcIMS) requires knowledge of programming languages such as HTML, JavaScript, Common Gateway Interface (CGI), and Hypertext Preprocessor (PHP) (PHP 2002). There are many different software packages available to begin an interactive mapping project with. The book 'GIS On-line' (Onward Press 1999) lists many of software packages available and the functionality of each.

2.6 OpenGIS Models

The basic elements of openGIS perform much like that of a desktop GIS. Files are read in the same manner and the same types of files are used (e.g., shape and DBF files). The differences exist in the distribution of the operational components of the GIS system and the functionality that ties these components together. With a desktop GIS system all files and functionality reside on the computer, while openGIS applications, components, functions, or operations are allocated in a distributed fashion between the client and the Web-based server. There are three general models in which openGIS can be implemented. All the models have a minimum of one client, one web server and one GIS server (Foresman 1998, Kraack 2001).

2.6.1 Model 1

The first model is the most common as it is the easiest to implement (Figure 2.1). The client sitting at a personal computer and using a Web Browser submits a request for a map to a Web server, identified by its Uniform Resource Locator (URL) address. The GIS server prepares the requested images and sends the images to the Web server, which in turn, sends the images to the client's Web browser. At this point the client's GIS

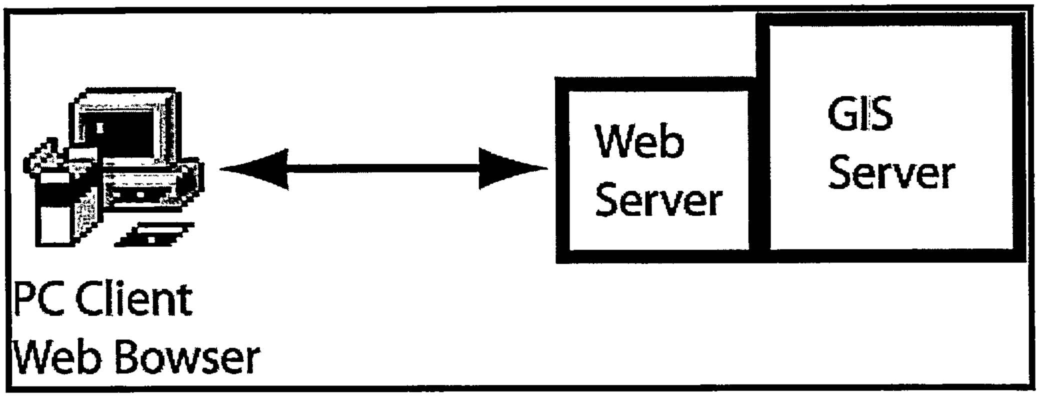


Figure 2.1 – The first model of networking openGIS

software or mapping software can display the image. This is simplest method but also the most limited. Performance speeds are slow and GIS software is needed to view the maps with any functionality. Only one layer at a time may be viewed. It is possible to implement this system using static images as maps, but the images would have no spatial aspect to them. Meaning images have no geographic coordinate system assigned to them.

2.6.2 Model 2

This model of delivering openGIS is similar to Model 1 in the sense that a client sends a request to a Web server, which in turn sends the request to the GIS server (Figure 2.2). The only significant difference is that numerous GIS servers work in unison to allow for larger data sets to be used. With this model multiple servers create a system that works at a faster speed and has the potential to add multiple layers from different computers.

2.6.3 Model 3

Model 3 represents a more advanced openGIS model (Figure 2.3). The client, working with *JAVA* or *ActiveX* browser programs, submits a request to the web server, which forwards the request to the GIS server. The GIS server prepares data requests for images by retrieving selected layers from the database servers using the web or local area networks. The database servers (not necessarily all on one computer) contain all the layers for dynamic maps. With the database servers searching out a client's query using a standard query language (SQL) this system allows users to select and display all the

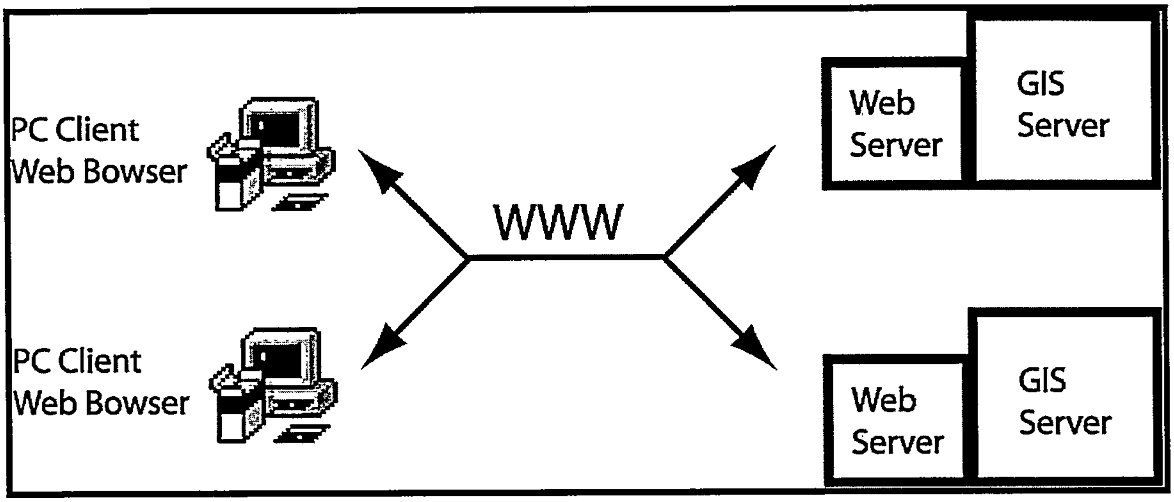


Figure 2.2 – The second model of networking openGIS

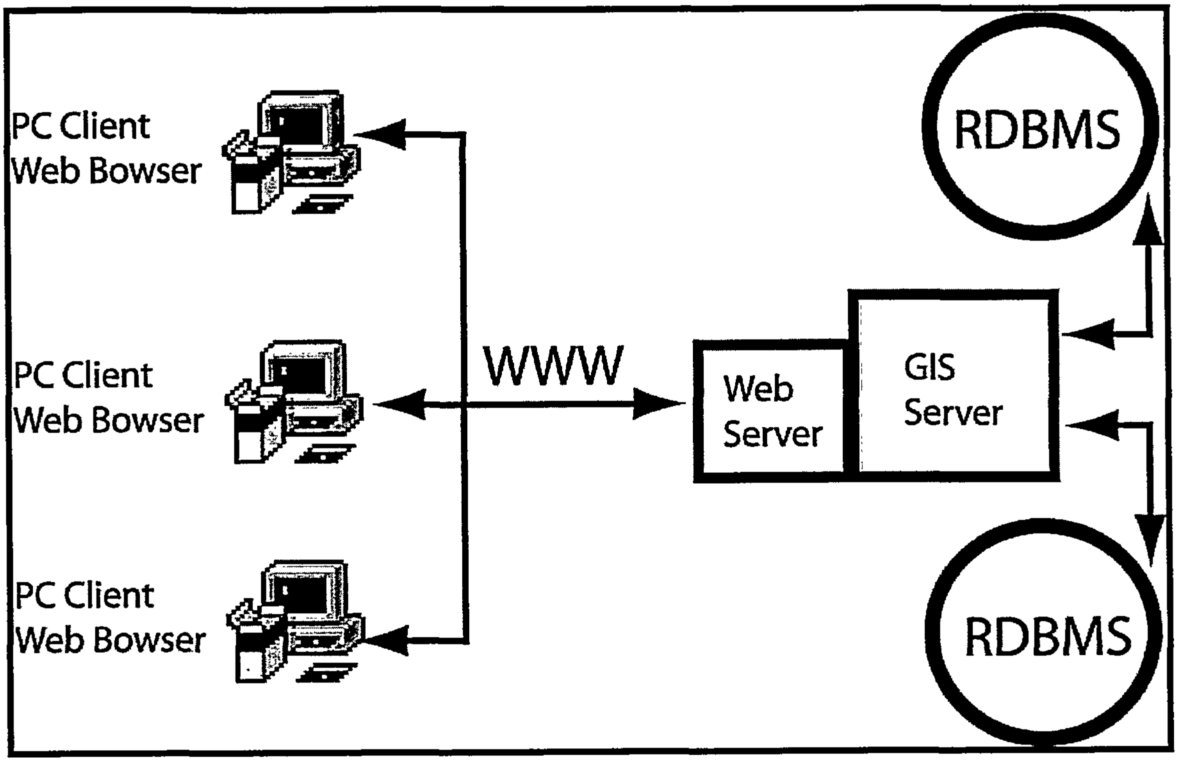


Figure 2.3 – The third model of networking openGIS

layers they want, in a fashion as close as possible to a fully-functional desktop GIS. This model allows for the client to view maps without having GIS software on their desktop. All that is needed by the client to use this model is a Web Browser that supports *JAVA* and *ActiveX*. All the popular brands of Web Browsers such as Netscape Navigator and Internet Explorer give the needed support. The drawback of this model falls on the side of the distributor, as it is the hardest to implement in terms of time and expertise.

2.6.4 Model 4

The fourth model of openGIS is both the most complicated to implement and has the greatest potential. This model has been labeled space online analytic processing, or SOLAP. Figure 2.4 illustrates the clients involved in a SOLAP system. In this model the data is transferred from the database, to the OLAP client through SQL server. Data obtained by the online analytic processing (OLAP) client is then shifted to a Visual Basic Application where it is joined with a GIS program. Users of SOLAP can navigate through the data by adopting two different approaches: one from an OLAP interface and the other from the GIS. The first approach consists in making a selection of information from the OLAP interface and to order the localization of this information on a chart present in the GIS. A data exchange coming from the OLAP and moving towards the GIS is done discretely. Conversely, the second approach consists of viewing the data in GIS and requiring the OLAP to supply the relevant attributes for the areas selected in the GIS. In the same way, a data exchange from GIS towards OLAP is done discretely.

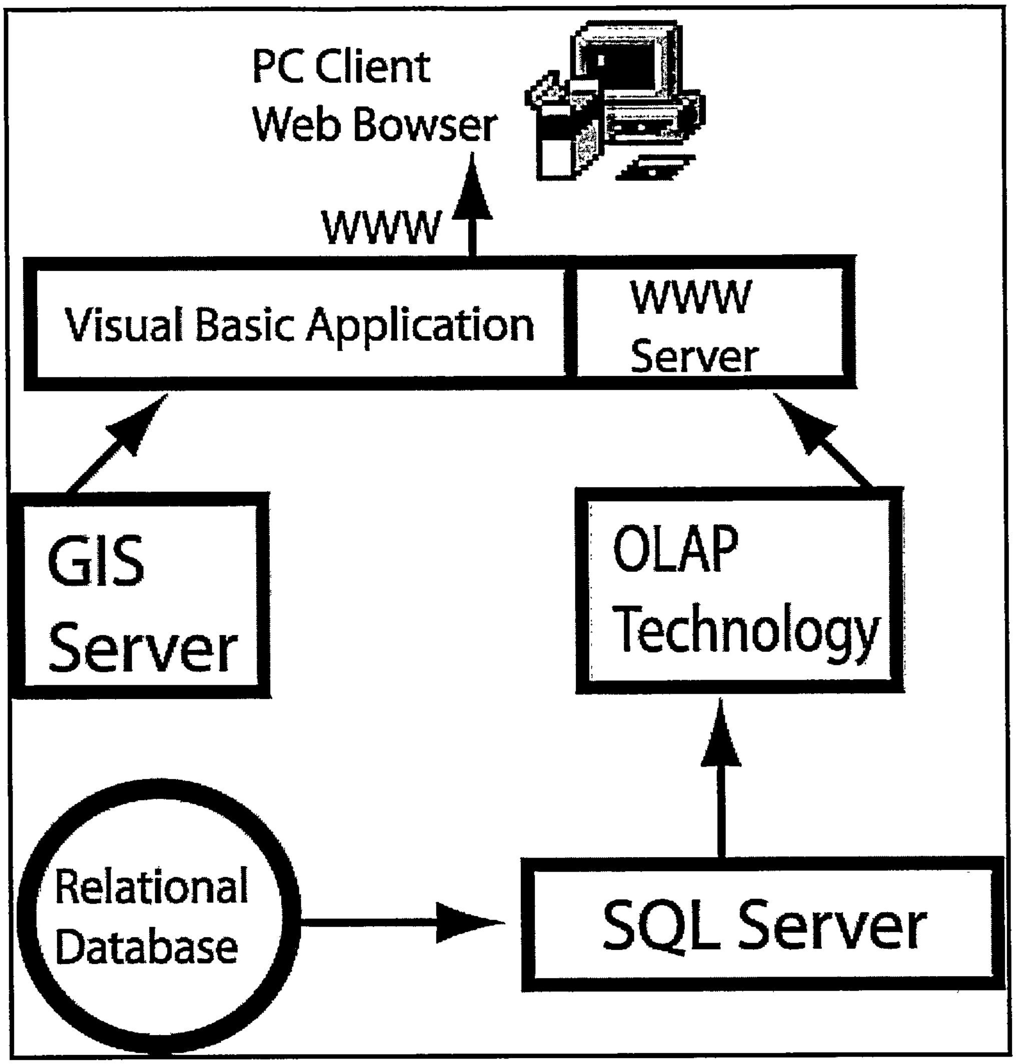


Figure 2.4 - The SOLAP model

Chapter 3

Methods and Data

3.1 Introduction

This chapter focuses on the data and methods used to develop and refine the interactive health mapping website for Hamilton. A full overview of the methods and data consists of four stages: (1) assessing user needs and the current state of the technology, (2) selection and compilation of the data for analysis and the subsequent creation of the prototype, (3) assessment of the prototype by the intended users through a survey and informal feedback from the focus group, and (4) improvements to the prototype based on this assessment. The stages of development can be applied to 'waterfall' method of software development. This idea is further explained in Chapter 5 (*Discussion*).

The research leading to the development of the site stemmed from two studies: (1) a user needs assessment and (2)a review of the current applications. The creation of the prototype is documented to explain how and why decisions were initially made. Specifics of the creation (i.e. the computer code) are referenced in Appendix A.

The user needs assessment of GIS, on-line analytic processing (OLAP), and WWW applications for environmental and public health surveillance supplied important information that guided the initial phases of this research. The study conducted by the National Health Surveillance Infostructure (NHSI) showed public health professionals from all levels of government and the private sector use, or have a desire to use, spatial data in their work (Bedard et al. 2000). Considering the survey results, a review of current applications involved searching for web sites which incorporate openGIS and health data. Upon being located, web sites were subjected to a standard set of criteria to assess the effectiveness. These two studies led to the creation of an openGIS tool for Hamilton, Ontario, Canada.

The tool was created to meet concerns voiced in the users survey and in the literature. The tool sets an example for future openGIS projects. One important concern from the users survey was the quality of data. To ensure data quality, collection methods and other such details are documented. Data used includes asthma rates, key pollutant levels, and 1996 census data all aggregated to the census tract level. Fixed location schools, hospitals, etc, were also included. To test the prototype a focus group was held at McMaster University in partnership with the Central West Health Planning Information Network (CWHPIN) on November 2nd, 2001. About 20 public health professionals from Southern Ontario attended the focus group. The focus group allowed for a demonstration of the prototype and presentations regarding GIS and spatial data.

Attendees were asked to participate in a survey to obtain their opinions on openGIS in general and the prototype in particular.

3.2 Waterfall approach

The waterfall approach views the process of software development as several stages, including requirement specification, software design, implementation, testing and so on. This model has be modified and applied to this thesis. Figure 3.1 illustrates the how the waterfall approach has been applied. The model is broken into four sections, requirement analysis, prototype development, unit testing, and modifications based on testing. When one section of the model is complete it runs into the next section, hence the term 'waterfall.'

Each section of the model incorporates work done for this thesis. Requirement analysis includes the literature review, the NHIS study, and the review of current applications. This work provided the first step in the 'waterfall' process. The second step is the development of the initial prototype. This includes all the research that went into choosing the software and functionality. This resulted in the first version of the prototype, the alpha version. From this step I go to user testing which took place at the focus group. At this step information was gathered about concerns with the prototype. These were validated these against the literature. After the concerns to be addressed are organized the fourth step begins, improving upon the alpha version of the prototype. The improved version is known as the beta version.

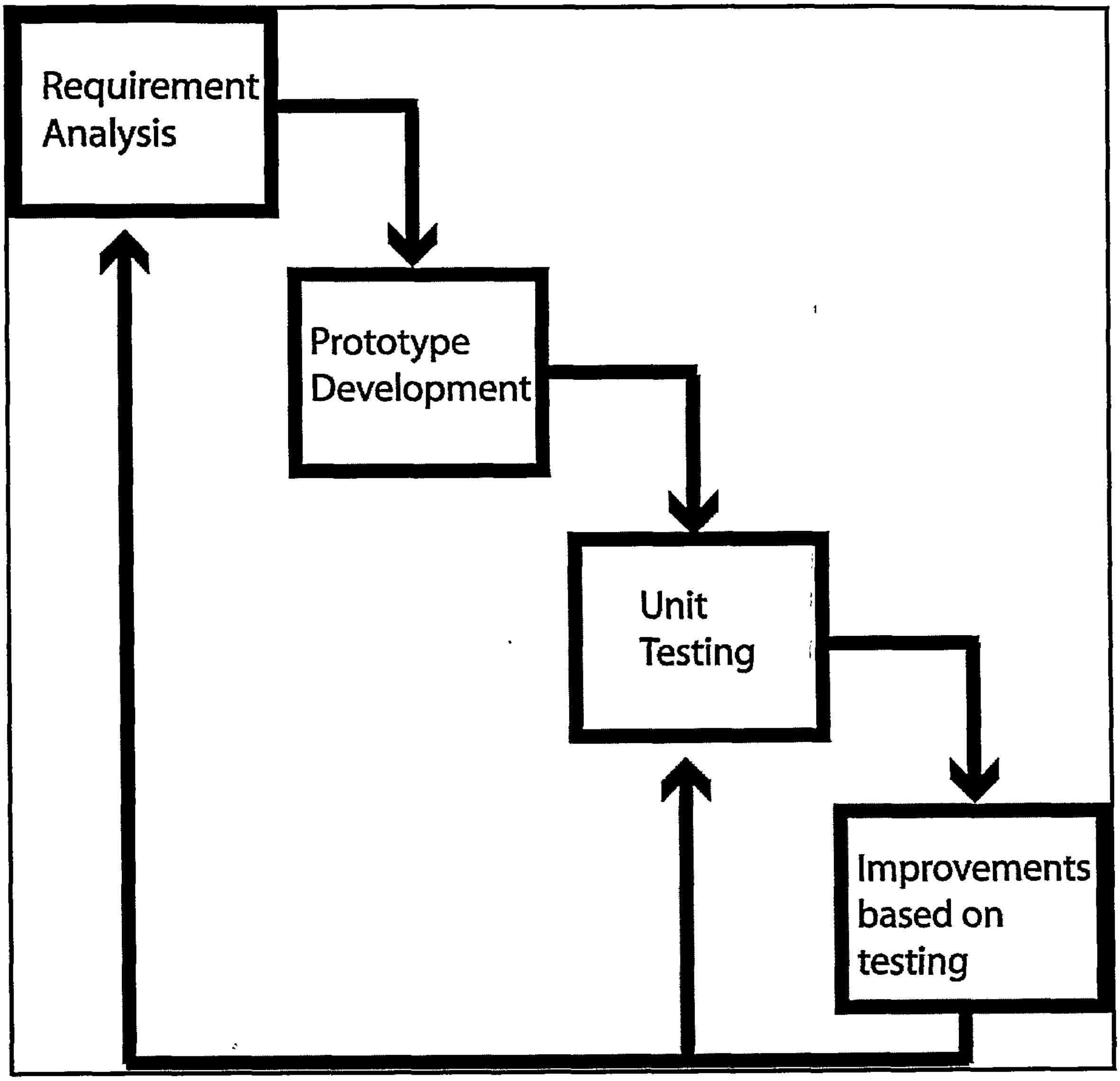


Figure 3.1 - The modified 'waterfall' method of software development

Although the 'waterfall' usually ends with complete software, this project will continue as a 'living' community resource. The four sections to the model described represent the extent of this thesis. An important aspect of any GIS project is determining a reasonable scope of what can be accomplished in the allotted timeframe (Aronoff 2000). Researching, implementing, testing, and improving upon a single openGIS prototype proved a reasonable scope for Master's level thesis. The project will carry on in future years and the model takes this into account. The beta version of the prototype will, in time, be improved upon. The model shows how future versions may be created from unit testing, requirements analysis or a combination of both.

3.3 The User Needs Assessment

The National Health Surveillance Infostructure (NHSI) was launched in 1997 to demonstrate the use of new information technologies and internet-based tools as a way of improving public health surveillance. The NHSI is a series of pilot projects aimed at: (1) improving access to existing databases, (2) facilitating the linkage of databases, (3) providing an affordable evidence-based tool for analysis and presentation of information, and (3) timely access to information.

The user needs assessment carried out by the NHSI (Bedard et al. 2001) contained two parts; (1) a literature review of how the tools of GIS, WWW, and On-line Analytic Processing (OLAP) are used in public health and (2) a national survey to determine how many professionals use these tools and their concerns with them. The benefits of the national survey were twofold. First, it allowed insight into developments that are not yet published. Second, it cross validates user perceptions from the literature review with practical examples and concerns from the field.

The literature review demonstrated that initial applications of GIS technology focused on cartographic visualization of health outcomes and related covariates (Catelan et al. 2000). Current research has shifted to decision-making applications utilizing GIS, OLAP, and WWW technologies. The WWW was the most common tool being used in approximately 80% of applications. Health GIS applications ranked second being involved 20% of the time. Other tools such as OLAP and the medical data warehouse were discussed in a few articles. Another important realization of the literature review was that almost all applications use only one technology. Figure 3.2 gives a graphic overview of the literature review results. The literature acknowledges the multidimensional and temporal limitations of traditional GIS. Certain functions of GIS require complex manipulations and programming which may not always be realistically implemented due to time and fiscal restraints. As a result it is difficult for most health officials to manage and analyze environmental health data in a GIS environment.

The second part of the user needs assessment was conducted as a survey. Thirty public health specialists across Canada participated in a survey focusing on current and future needs of the Canadian health community and requirements driving the use of

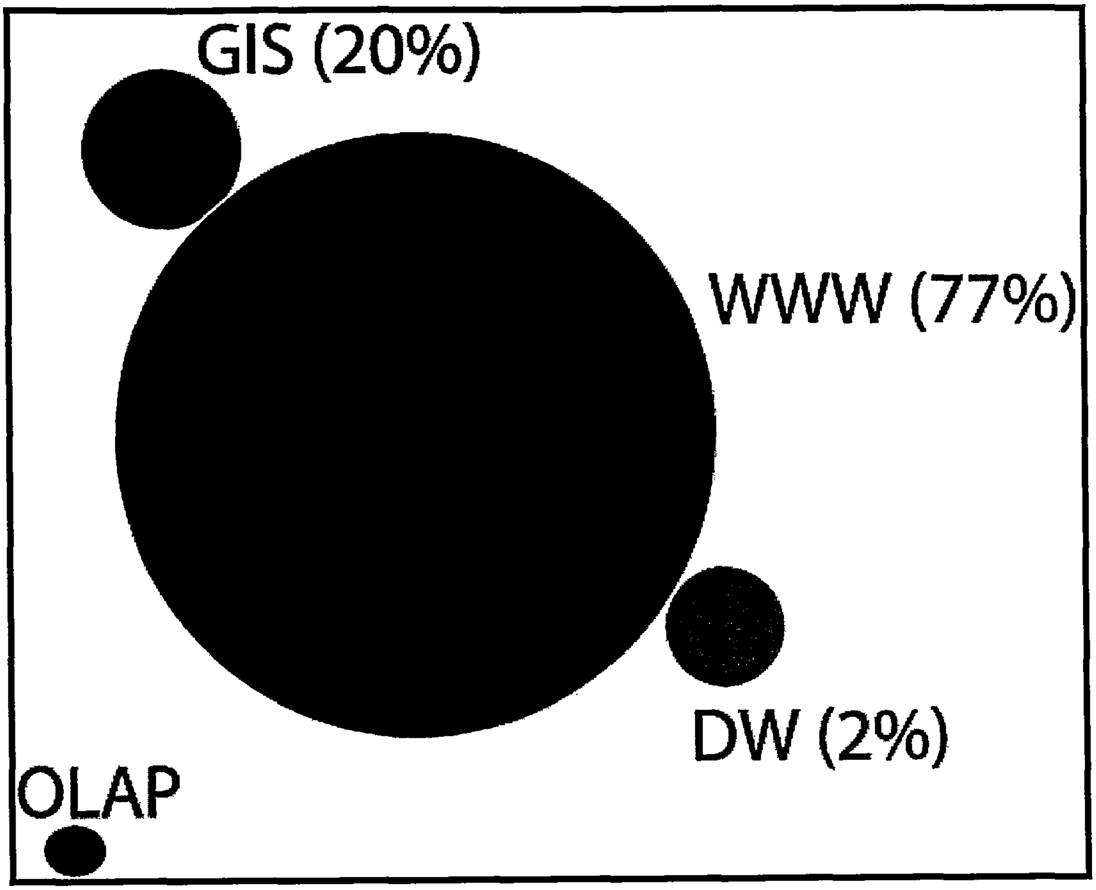


Figure 3.2 – Relative use of GIS, WWW, and Data Warehouse (DW) technologies by Public Health professionals according to the NHIS survey

WWW, GIS and OLAP tools. Table 3.1 gives an overview of the respondent's spatial distributions, organization level and the type of intervention or service dealt with. The interviewees were administered a survey containing 43 open-ended and closed-ended questions between October 21st and November 5th, 1999. Most (80%) interviews were done by phone and had an average duration of 70 minutes. Respondents were informed that all answers would be kept confidential and be aggregated to prevent individual identifications.

The results of the user need assessment survey demonstrate the concerns of data availability/reliability and the generally low level of user-friendliness in most GIS applications. There is high interest among the operational public health professionals for furthering use of GIS (96%), but there is much lower recognition of the need for OLAP (27%). See the detailed results section and Table 3.2 for more results and analysis.

Respondents reported that the use of GIS and digital mapping would bring the greatest value to their organization in the areas of analysis, reporting, problem identification, and decision-making. Less generally the respondents noted that these tools are useful for the identification of patterns (e.g. clustering) in disease, health behaviours, and health care utilisation, and that the tools can also help with further analysis of these patterns.

Table 3.1 – Overview of the distribution, organization level, and type of invention dealt with for respondents of the NHIS survey.

ORGANISATIONS LEVEL					
Number of Respondents	Federal/National	Provincial	Regional/Local	Universities	NGO & Private
30	9	5	10	3	3

TYPE OF INTERVENTION				
Number of Respondents	Direct Health Services	Public Health 2^{nd} & 3^{rd} line	Knowledge Development-Planning	Health impacts, Economics and Social Activities
30	5	12	11	2

REGIONAL DISTRIBUTION OF INTERVIEWEES				
Number of Respondents	Western Canada	Ontario	Quebec	Atlantic Canada
30	6	11	9	4

Table 3.2 – Functionality respondents would like to perform on spatial data

ACTIVITIES AND FUNCTIONS	RESPONSES
To have access and to query the data	100%
To perform detailed ad hoc aggregation and exploration of data	90%
To edit or create new listings, reports, statistical charts, etc.	87%
To edit or create new thematic maps	77%
To perform sophisticated data analysis	70%
To view pre-designed listings, reports, statistical charts, etc	63%
To perform spatial analysis	57%
To perform detailed mapping	53%
To perform automatic detection of data patterns, correlation and trends	47%
To view pre-designed thematic maps	47%
To perform automatic detection of spatial patterns, correlation and trends	40%

Additionally, the tools used by public health professionals are thought to provide important visual support for the reporting/dissemination of information to a wide range of audiences (from political officials to the general public). This ability to visualize health information may bring attention to a problem that a report based solely on numbers may not. Identification of potential health problem areas, and subsequently identifying intervention areas and boundaries with GIS and digital mapping, are also useful.

While the study illustrated the confidence of public health professionals in the ability of GIS and digital mapping to enhance their work, it also showed how GIS is underutilized. A clear indication of this confidence was indicated by 96% of respondents felt spatial information was not being used to its full extent. Many respondents felt that the use of spatial data was just in the beginning stages and will increase in use. There were concerns that may slow the growth. Concerns such as a lack of resources, capacity to train staff, and quality data could be addressed with openGIS. Allowing public health offices to access an openGIS system at little or no cost coupled with a diligent effort towards data sharing would allow for these concerns to be at least partially addressed.

Concerns with data management also became apparent from the study. Users were uneasy with data access, feeling they do not currently have access to enough data sources and do not know where to find them. Even when they do know where the data is located often the cost recovery by federal data providers impedes their access. The lack of data standardisation limits the usefulness of many data sets and causes extra work (e.g., changing the map projection). This extra work makes data more expensive. Respondents expressed concern about the quality of the data received. Most respondents report having insufficient funds to access as many of the data sources as they would like. Finally, currency and reliability of data sets used by respondents is problematic as many reported using older, potentially out-of-date data sets. Table 3.3 summarizes users concerns and issues concerning spatial data and technology.

With major concerns revolving around cost and data quality, the study showed that there is an audience for an openGIS system. An openGIS system has the potential to bring GIS technology to public health professionals at minimal cost. No expertise is needed with a well designed openGIS. Data would be standardized for all those using the system.

3.4 Assessing the Current Technology

Effectively implementing an openGIS system required exploration of the tools in current applications. Initially the search was limited to openGIS sites that dealt specifically with health surveillance. It became quickly apparent that this limited scope would be a problem. Currently, only a few sites deal with health surveillance mapping. We also reviewed sites that could be used to assist health surveillance such as those that supply sociodemographic or pollution data. This allowed for a lager sample of sites reviewed.

Table 3.3 – Users concerns and issues concerning spatial data and technology

PROBLEMS/ ISSUES	% OF RESPONDENTS THAT SAID HIGH/MEDIUM
Data related issues	
Knowing what data exist and finding it	**90%**
Timeliness (ability to get data timely)	**80%**
Quality of data	**73%**
Standards/lack of compatibility	**70%**
Costs of data	**63%**
Access of data	**63%**
Security/confidentiality	**43%**
Other issues	
Training and education of users	**77%**
High costs to sustain IT	**63%**
Technology limitations	**43%**
Absence of demand from clients	**37%**

3.4.1 Searching for openGIS applications

All of the websites were found through the use of numerous Internet search engines, including www.northernlights.com, www.altavista.com, and www.google.com. A study done to document the accessibility of information on the web has shown that no one Internet search engine indexes more than 16% of all websites (Lawrence 1999, Lookoff 2000). There are two major types of search engines 'web crawlers' and 'web portals' (Lookoff 2000). The search engine 'Google' is a web crawler, 'Northernlights' and 'Altavista' are web portals.

A web crawler scours the Internet looking for pages to 'index.' It generally starts with some default web addresses and downloads them. It pulls out all the addresses (links) contained within the page to be searched later. It then indexes all of the words in that page allowing for every word and phrase to be stored in a database. Thus, every word in the page is 'searchable' once it is saved to the database. The process of indexing web pages continues as the addresses pulled out earlier are then searched. The main problems with web crawlers are that not all websites can be reached from the arbitrary point the 'crawling' starts with and searching for common information (such as the keyword 'health') results in millions of web pages matching the criteria.

A web portal is a search engine that organizes information by topic. This produces a more systematic search than web crawlers. It is also a less encompassing search. Only the summary of the web site (approximately 100 words) is searched, not the content. The main problem with the web portal method is defining a web sites broad topic. For example, would a hypothetical web site entitled 'Health mapping on the Internet: an experiment' fall under the category of 'Health care,' 'Internet technologies,' or 'Academic web sites?'

Websites are recognized by search engines after the site has been registered or 'discovered' the search engine. Not all websites need to be individually registered. A website served from an institution which has already been registered (e.g. McMaster University) will automatically be recognized.

It was for these reasons that three different Internet Search engines were used. Within each search engine the words or phrase, cartography, mapping, Internet, 'ArcView Internet Map Server (AVIMS),' 'Internet map server,' health, GIS, 'geographic information systems,' and IMS were all used individually and in various combinations to search for applicable websites. This method of searching was not intended to be exhaustive, but instead to supply a representative sample of what the Internet has to offer in terms of mapping. With well over one billion web pages (Inktomi 2001), an exhaustive search would have been well outside the bounds of this thesis.

3.4.2 Determining the Criteria for Assessment

To evaluate existing websites systematically a table was created (Table 3.4) to display which web sites met evaluation criteria. To do this 21 evaluation criteria were identified which were divided into five subgroups: 'Metadata,' 'Technical Aspects,' 'Data,' 'Cartographic Capabilities,' and 'Data Analysis.' Reasoning behind each criterion is explained in the following paragraphs.

It is possible for anyone with a computer and an Internet connection to create and post a web page for the world to see. Because of this one must be cautious when considering the quality of information available. All information and data posted must be specific about the source. This becomes even more imperative when dealing with health data. For a researcher, physician or public health agent to work with information found on the Internet they would need to know the 'why,' 'what,' 'where,' 'when,' 'who,' and 'how' of the data. All of this should be found in the Metadata. The Metadata is an overall description of the contents of the database (Demers 1997). Referring back to Table 3.4 the criteria with the Metadata subgroup are 'Authorship attribution,' 'Disclosure of funding,' 'Currency of data,' and 'Attribution of sources.' These criteria are used to determine if the web site has credibility. 'Authorship of data' checks if the credentials and affiliations of the author(s) are made clear. The thoughts as to the credibility of the author(s) are personal for each user but the information for that decision must be made available. The source of funding gives clues about where the data collection and analysis came from. 'Disclosure of Funding' allows the user to look for any conflicts of interest. For example, a user may not trust pollution data that was collected and analyzed through funds made available from a steel company. Dates in which the data was collected and posted are necessary for future use in

Table 3.4 - The criteria current openGIS applications were judged by

	Criteria Tested	Websites reviewed																														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	% fit criteria
Metadata	Authorship attribution																															
	Disclosure of funding																															
	Currency of data																															
	Attribution of sources																															
Technical Aspects	Site accessability																															
	Ease of Navigation																															
	Map upload time																															
	Speed of site startup																															
	Query process time																															
Data	Links to other sites																															
	Interactive search																															
	Confidentiality explicit																															
	Downloadable data																															
Cartographic	Scale Capabilities																															
	Colour scheme availability																															
	Graphing																															
Data Analysis	Interactive overlays																															
	Boolean queries																															
	Reclassing data																															
	Pattern recognition																															
	Advanced regression																															

research and this criteria is observed in 'Currency of data.' Finally, all references, sources, copyrights and website 'ownership' need to be available for the user (Silberg 1997). This is checked in 'Attribution of sources.'

The 'technical aspects' section of Chart 1 deals with criteria that show the ease of user interaction. Websites that have long uploading or processing times can be irritating to the user, as can a site which is awkward to navigate (Stynes et al. 1998). To allow for the 'yes/no' answer that was needed to produce the chart, a criteria by which the speed was to be judged had to be determined. For each of the chart sections associated with speed, 'Speed of site start-up,' 'Map uploading time,' and 'query processing time,' a time was taken, using a stopwatch, to measure how long each applicable site took to accomplish each speed associated criteria. The 'Speed of site start-up' criterion tests the web site which houses the openGIS. 'Map uploading time' and 'Query processing time' test the openGIS aspect of the web site. Some of the web sites tested did not have the capability to query the data and these were excluded. After each site was timed, a mean was taken and each site that took less time than the mean met the criteria. The last criteria in this section 'Is data accessible by all?' shows whether or not the site has data that may be accessed by anyone or permission must be granted from the website creator to use or view the data.

Data availability is important to health researchers. The criteria in the 'Data' subset on Table 3.4 look at data within the web sites tested. The ability to download data allows others to add to, and potentially improve upon, work already done. Earlier we established confidentiality as an issue within the realm of openGIS and health. Thus steps to ensure confidentiality should be acknowledged. The final two criteria within the subsets 'Interactive search' and 'Links to other sites' test the web sites ability to help the user find the information they need. A web site with an interactive search function allows the user to search the entire website quickly without going through the potential problem of missing the one piece of information they need. An interactive search is simply a text box that allows the user to type in a word or phrase they would like to search for within the site. Being able to search for information within the site allows the user to get what they need quickly even if the site is not well organized. Having links to other sites is a way for the user to find more information about the web site (e.g., a link to the funding agency), the data (the web site of the researcher who collected it), or further information (e.g., a link to the Center for Disease control).

The last two subsets of criteria, 'Cartographic capabilities' and 'Data analysis,' test the functionality of openGIS. Within 'Cartographic capabilities' the ability to create graphs with the data available for the openGIS is tested. The ability to create graphs is another way to explore the data and look for trends. The 'Colour scheme availability' criterion demonstrates if the colours used by the openGIS can be changed. With colour having an impact on the map reader's perception (Brewer 1998), it has been suggested that giving the user the option of altering the colour scheme may not be a good idea

(Kraack 2001). Having the ability to 'zoom in' and 'zoom out' satisfies the 'scale capabilities' criteria.

The last of the criteria subsets, 'Data Analysis,' tests the abilities of the most advanced openGIS sites. 'Boolean queries' test for the ability to query the data with the users' criteria (e.g. an 'AND' or 'OR' statement) (Demers 1997). The capability of overlaying multiple layers is a basic concept in desktop GIS but an advanced one in openGIS. 'Pattern recognition' and 'advanced regression' are two functions are widely used in spatial analysis (Bailey and Gatrell 1995) and are advantageous functions to health professionals. The last criterion in the subset, 'Reclassing data' is ability to change the number or position of classes on a choropleth map. For example, a data set might have one hundred records and be initially shown on a chloropleth map with four classes. If the user has the ability to reclassify the data the map could be remade using, for example, six classes.

3.5 Data and Data Relationships

In the user needs assessment respondents identified having access to relevant, reliable, and current data on health outcomes and risk factors as a high priority. Thus, the prototype health mapping tool project allows the user to explore relationships between asthma rates, pollution rates, and socioeconomic status. The data used in the project can be broken down into three groups: pollutant data, census data, and health data. In addition to explaining the data this section will include background information explaining why the data is relevant.

3.5.1 Pollution Data

The prototype includes three pollution variables: particulate matter, sulphur dioxide, and nitrogen oxides. Each pollutant has been linked to health problems in recent literature (Wardlaw 1993, Schwartz 1993, Ponka 1991). Although the evidence for a genetic basis for asthma is strong, the environment also plays a major role in inducing and maintaining the asthmatic process (Manfreda 2001). Pollution data was collected at fixed site ambient monitors operated by the Ontario Ministry of the Environment (MOE) throughout Hamilton. Particulate matter was collected at 19 stations, sulphur dioxide at 9 stations, and nitrous oxide at 6 stations (MOE 1998). The data were interpolated to give a continuous surface across the city. Interpolation is the estimation of values of a surface at an unsampled point based on the known values of surrounding points (Bailey and Gatrell 1998). Each pollutant was interpolated using a different type of interpolation. Particulate matter was interpolated across the city using the 'universal kriging' method, sulphur dioxide using the 'spline' method, and nitrous oxides using 'inverse distance weighting.' Each census tract was then assigned a pollutant value based on the pollution value at the population-weighted centroid (Jerrett et al. 2001). The interpolation of the pollution data was not a part of this project. For a complete explanation of the types of interpolation used please refer to Bailey and Gatrell (1998).

3.5.2 Particulate Matter

Airborne particulate matter consists of organic and inorganic solids from a variety of sources which are suspended in the air. They can be formed directly from chemical emissions or secondarily from reactions between gases in the atmosphere (Ontario 1999). Particulate matter is traditionally classified into three broad categories, total suspended particulate (up to 50 micrometers in diameter), PM_{10} (coarse particulate) and $PM_{2.5}$ (fine particulate). The website uses PM_{10} data only for numerous reasons: (1) it is spatially disaggregated, (2) available in the same time frame as the asthma data, and (3) linked to many health effects.

Particulate matter has been associated with numerous detrimental health effects, especially those that are cardio-respiratory in nature. It has been linked to chronic obstructive pulmonary disease and asthma (Schwartz et al. 1993, Ponka 1991, Anderson et al. 1998, Chew et al. 1999, Stieb et al. 1996, Bates 1983, Pope 1989). Evidence suggests that there is no threshold value for health effects to be present, although they are thought to be proportional to pollutant concentration (WHO 1997). The MOE suggests that a:

> Large number of scientific studies carried out more recently in the United States, Canada, Britain and Europe have provided a growing body of evidence linking relatively low concentrations of PM, within the range experienced regularly across Ontario and Canada, to adverse cardio-respiratory impacts, including increased hospitalization and increased premature mortality, in a broader spectrum of the population (Ontario 1999).

Epidemiological studies have estimated higher relative risk for elderly patients with existing respiratory diseases (Elbelt et al. 2000), other subpopulations at risk are currently poorly characterized (CRPAPM 1998). Ambient particulate matter levels found outdoors predominately originate from motor vehicles.

3.5.3 Sulphur Dioxide

Sulphur dioxide (SO_2) is an air pollutant produced through combustion, both naturally through volcanic eruptions, but more importantly through the burning of fossil fuels, such as in vehicles or coal and oil power plants (USDHHS 1998). A report from the MOE (1994) estimates that 54% of sulphur dioxide comes from industrial sources. Sulphur dioxide is commonly associated with several health effects, including bronchioconstricition, elevated morbidity, mortality and asthma (Lippmann 2000).

While exposure to sulphur dioxide is primarily due to outdoor concentrations, personal outdoor activities can lead to higher exposures. Sulphur dioxide is highly soluble and is readily absorbed by the lungs protective mechanisms at normal breathing rates (Health Canada 1998). At higher breathing rates brought on by exercise the ability to remove sulphur dioxide from the air that reaches the lungs is compromised and the gas can penetrate into the lungs where it can potentially cause health problems. Individuals who perform outdoor exercise in areas of high ambient sulphur dioxide concentration are at a greater risk for higher exposures. Of particular concern are asthmatics who undertake outdoor exercise in areas of high concentrations due to their higher susceptibility to the effects of exposure (Lippman 2000).

Exposure to sulphur dioxide is primarily due to outdoor sources as there are few significant indoor sources. Spatial variation in sulphur has been found to be quite large, with levels ranging between 1-5 ug/m3 in rural areas in the United States, compared with levels reaching as high as 6,000 ug/m3 (USDHHS 1998, Ito and Thurston 1993). Diesel fuel, industrial activities that involve the use of fossil fuels, and electrical utilities are major sources that can lead to higher concentrations. The only indoor sources that are of concern are improperly vented kerosene space heathers and occupational exposure for workers in steel processing, refinery and chemical plants (USSDHHS 1998).

3.5.4 Nitrous Oxides

Nitrogen dioxide, and other related nitrogen oxides (NO_x), are air pollutants produced through combustion processes. Evidence indicates that exposure increases respiratory infection rates and limited data indicates that it may heighten responses to airborne allergens (Sunyer 1997). Asthmatics with allergic asthma are a susceptible population to negative health consequences from exposures. At high concentrations lungs are damaged through oxidant injury, lung defence mechanisms are adversely affected and clearing of the infected organisms, particularly bacteria, is reduced. Nitrous oxides are thought to play a role in increasing incidence and severity of respiratory infection and exacerbating pre-existing respiratory problems such as asthma.

Outdoor levels of NOx are predominately due to motor vehicle traffic, fossil fuel burning plants and industrial boilers (Levy et al. 1998). Measured levels have been observed to be much higher in urban areas than rural and the urban levels have been steadily increasing with an increasing number of cars.

3.5.5 Health Data

Asthma data was collected for adults through a survey administered in 1994 (Manfreda 2001). The survey was mailed to randomly selected samples of the general population aged 20-44 years in 6 cites across Canada, selected to represent different environments with respect to climate, air pollution and occupational exposures. One of these sites was Hamilton, with the others being Vancouver, Winnipeg, Montreal, Halifax and Prince Edward Island. Sampling and data collection were evenly distributed throughout the year.

The questionnaire was developed by the International Union against Tuberculosis and Lung Disease as a standard to be used in prevalence studies of asthma in adults (Burney 1987). The validity and reproducibility of this questionnaire were found to be satisfactory (Burney 1994). In the absence of an accepted definition of asthma, the burden of asthma in the population is best represented by the distribution of asthma symptoms and bronchial responsiveness (Manfreda 2001). The prevalence of asthma symptoms, attacks and medication are questioned as opposed to 'Have you ever had, or have, asthma.' Questions concerning symptoms of respiratory problems, non-asthma respiratory symptoms, smoking, occupational exposure to dust, fumes and gases and residential history are included in the questionnaire. Each sample was drawn from a population of at least 150 000 people of all ages. Samples consisted of 3000 – 4000 adults ages 20-44 years in five urban sites and in PEI, where the rural – urban ratio was 1.5:1. In the five cities random digit dialling was used to identify eligible individuals. For each city, a random sample of 18 000 telephone numbers, divided into 12 sub samples, was generated. Interviewers called the random numbers to determine if more than 1 member of the household was eligible. Random digit dialling was not used in PEI because the sampling frame was the entire province and most calls would have been long distance.

Hamilton data were included in the health mapping prototype, although extension to other sites remains a possibility. The results were aggregated into rates at the census tract level. This study also produced smoking rates as an example of risk behaviour. These rates were included because smoking has been linked to the development and aggravation of asthma (Agabiti 1999, Cassino 1999).

3.5.6 Census Data

Census data is available from Statistics Canada. The three census variables implemented in the prototype are 'Dwelling value,' '1996 Population,' and 'Gini Coefficient.' Dwelling value is used as a measure of permanent family income (Jerrett et al. 1997). Studies have shown linkage between asthma rates and low income (Littlejohns

1993, Litonjua 1999, Persky 1998). The gini coefficient gives an indication of the income-distribution in the census tract. Ranges from 0 to 1, where higher numbers represent a higher degree of income inequality within the area. In the literature arguments have been both for (Lockner 2001) and against (Shibuya 2002) income inequality being a fair indicator of health. Earlier research indicated the Gini coefficient had the largest association with all cause mortality compared to 16 other variables. It is included in this project for users who incorporate it into their work. Lastly, the 1996 population was included for two reasons. First, it is a number of interest to the general user. Second, a higher population means the potential for more vehicles on the road. Traffic congestion has been linked to high asthma rates (Weiland 1994, Venn 2001).

3.6 Development of the Prototype

The first step in the development of the prototype openGIS system involved software selection. Because ESRI's Arc/INFO Internet Map Server (ARCIMS) is the leading commercial program available for use at McMaster University, it was initially selected for detailed evaluation. ARCIMS is a closed source software (where the source code is not available). Being a closed source program meant that there was little that could be done to add to existing functionality. This software also has other disadvantages, including: (1) large computational demands, (2) the need for downloading an Internet plug-in to support the system, and (3) high purchase cost. Although this program is available at minimal cost to the researcher through a McMaster University site license, this would not be the case in most organizations. The commercial price for

ARCIMS is $13,500, loaded on one Internet Server and one desktop computer. Additionally, there is an annual licensing fee of $2,700 to use the software. Demonstrating that openGIS can be implemented without using many resources was important for acceptance by public health agencies that lack resources. This led to a search for other suitable software. This research and initial testing led to a decision to use the open-source Mapserver software (MapServer 2002).

The University of Minnesota (UMN) in cooperation with NASA and the Minnesota Department of Natural Resources (MNDNR) originally developed MapServer. The MNDNR and the Minnesota Land Management Information Center (LMIC) made additional enhancements. Current development is funded by the TerraSIP project, a NASA sponsored project between the UMN and a consortium of land management interests. Using MapServer allowed for the development of functionality not available with ARCIMS. MapServer also comes at no cost. This may encourage future use of the tool. MapServer does not require a client side program aside from an Internet browser (called a plug-in in ARCIMS) for maps to be viewed. The ability to view maps without having to download plug-ins may be important to the person working in an office that runs on a network that does not allow downloads without the permission from the system administrator. Lastly, one of the main benefits of developing an application with Mapserver is the availability of the 'list serve.' The 'list serve' is an on-line newsgroup that allows the developer to e-mail a large group of people with any development questions.

3.6.1 Functionality of the Prototype

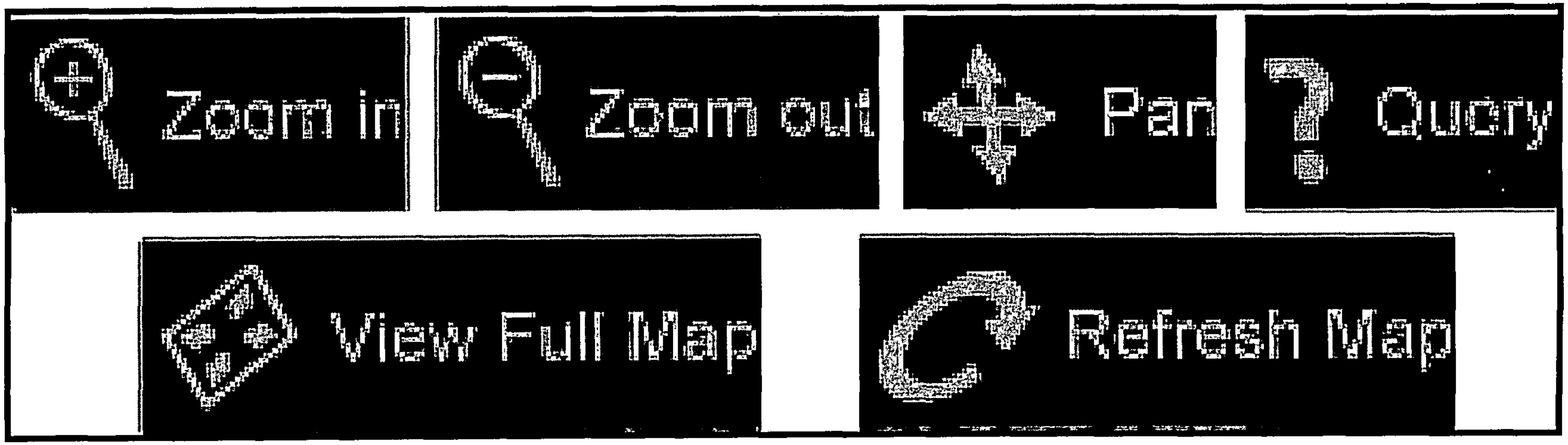
To create an effective openGIS site, we had to develop a site map that included functionality that would be useful to the intended audience. Functionality used to explore the map and data mimic those used in a standard GIS (Plewe 1997). This allows for users with any exposure to GIS to become instantly familiar with the functions. People not having prior GIS exposure may learn the functionality through 'trial and error' or, more effectively, the on-line tutorial. Keeping functionality simple to use, yet advanced enough to contribute to the intended audience, was a major challenge. The NHSI user needs assessment demonstrated that all users would like the ability to explore and query spatial data but worry about employing vast amounts time and resources. This is why functionality is kept as simple as possible, while still allowing for adequate capabilities. Basic functionality used in map exploration are the tools 'zoom in,' 'zoom out,' 'pan,' and 'view full map.' Map exploration tools do nothing to change the look of the map (what data is being viewed), and only change the users' viewpoint. The second step involved adding data exploration tools. Data exploration tools include: (1) several 'layers' of spatial information and the capability to turn them on and off (Longley 2001); (2) the ability to change the background from a uniform colour to a chloropleth map different, predetermined variables; (3) allowing the user to obtain information for a specific area of interest through a query button; (4) being able to input numbers into a query system permitting the user to locate areas on the map that match all the inputted criteria; and (5) develop a tool allowing the user to locate the census tract containing an inputted postal code.

The creation of the prototype involved many different programming languages to obtain the desired functionality. Appendix A contains all the computer code used to create the prototype. In the following paragraphs only the basics of how the prototype was developed will be discussed.

Mapserver 3.3 allows for some of the basic functionality to be created with ease. Using the 'Mapserver Wizard,' which is downloaded from the MapServer web site (Mapserver 2001), spatial data can be uploaded into an openGIS system. In this basic openGIS functionality is limited to turning themes on and off and zooming in or out. To solve this problem of limited functionality and the unattractive interface created by the Mapserver wizard a new web page had to be created to house the openGIS. This was created using Dreamweaver software from Macromedia (Marcomedia 2001).

MapServer creates maps within a Java Applet. A Java Applet is a program written in the Java programming language that can be included in an HTML page, much in the same way an image is included. When a Java technology-enabled browser is used to view a page that contains an applet, the applet's code is transferred to the remote system and executed by the browser's Java Virtual Machine (JVM) (java.sun.com 2002). MapServer gives the option of embedding the legend, a North Arrow, and the scale bar within the applet. Our website used all three. With Dreamweaver software the Java Applet can be embedded in an HTML page, allowing for new functionality and improved appearance.

Initially the 'zoom in' and 'zoom out' tools worked when the user clicked ('click' or 'clicked' will always refer to the user pushing the mouse button) on a check box. To make this function look more like a traditional GIS the 'zoom in' and 'zoom out' functions were assigned to graphics (Figure 3.3) produced with Adobe Illustrator software (Adobe 2002). Additional graphics were produced for the 'pan,' 'query,' 'refresh,' and 'view full map' functions (Figure 3.3). The 'pan' function was created using the computer language JavaScript. Code was written to command the computer to re-center the map on the point of the users choosing. Additionally, 'pan bars' were created to surround the map enabling the user to shift the map 10 kilometres in any of eight directions. The 'refresh map' and 'view full map' functions use very similar code. Again using the JavaScript language, the code created for 'refresh map' allows the user to implement the changes he or she may have made in terms of turning off or on layers. When using the 'refresh map' function, all changes in map visuals take place in the same scale of the current map. The 'view full map' function runs on the same code as 'refresh


Zoom in
Zoom out
Pan
Query
View Full Map
Refresh Map

map.' It does refresh the map but it also returns the user to the initial map extent. With the ability to explore the map implemented the next step was to produce functionality to explore the data. The first step was to create a query function.

'Query' functionality available through the MapServer software. The 'query' button allows the user to select an area or object on the map. Once identified, the user obtains specific information about the area (e.g. asthma rates and income). The user may choose which layer of data they would like to query but that option was not implemented here. With the prototype the user may only query the census tract data. After 'clicking' the query button a census tract of interest may be 'clicked' on. After the census tract is chosen a new window appears. The new window contains information about the census tract, displaying the population, gini coefficient, average dwelling value, and the concentrations of the following pollutants, PM_{10}, SO_2, and NO_x. Because users will not know the range of values for the data, the ranges are included for the population, gini coefficient, and dwelling value. For the pollutants graphs are shown placing the census tracts pollutant value as a percentile of all others in the city (Figure 3.4). This functionality was also created with JavaScript. The next step in the evolution of the 'query' command was to give the user the option of locating census tracts matching their criteria. The housing of the query system was created using HTML. The outline allows the user to input a range for the following variables 'Asthma Rates,' 'Particulate Matter,' 'Sulphur Dioxide,' 'Nitrous Oxides,' 'Smoking Rates,' 'Gini Coefficient,' '1996 Population' and 'Dwelling Value.' The user may enter ranges for up to three different

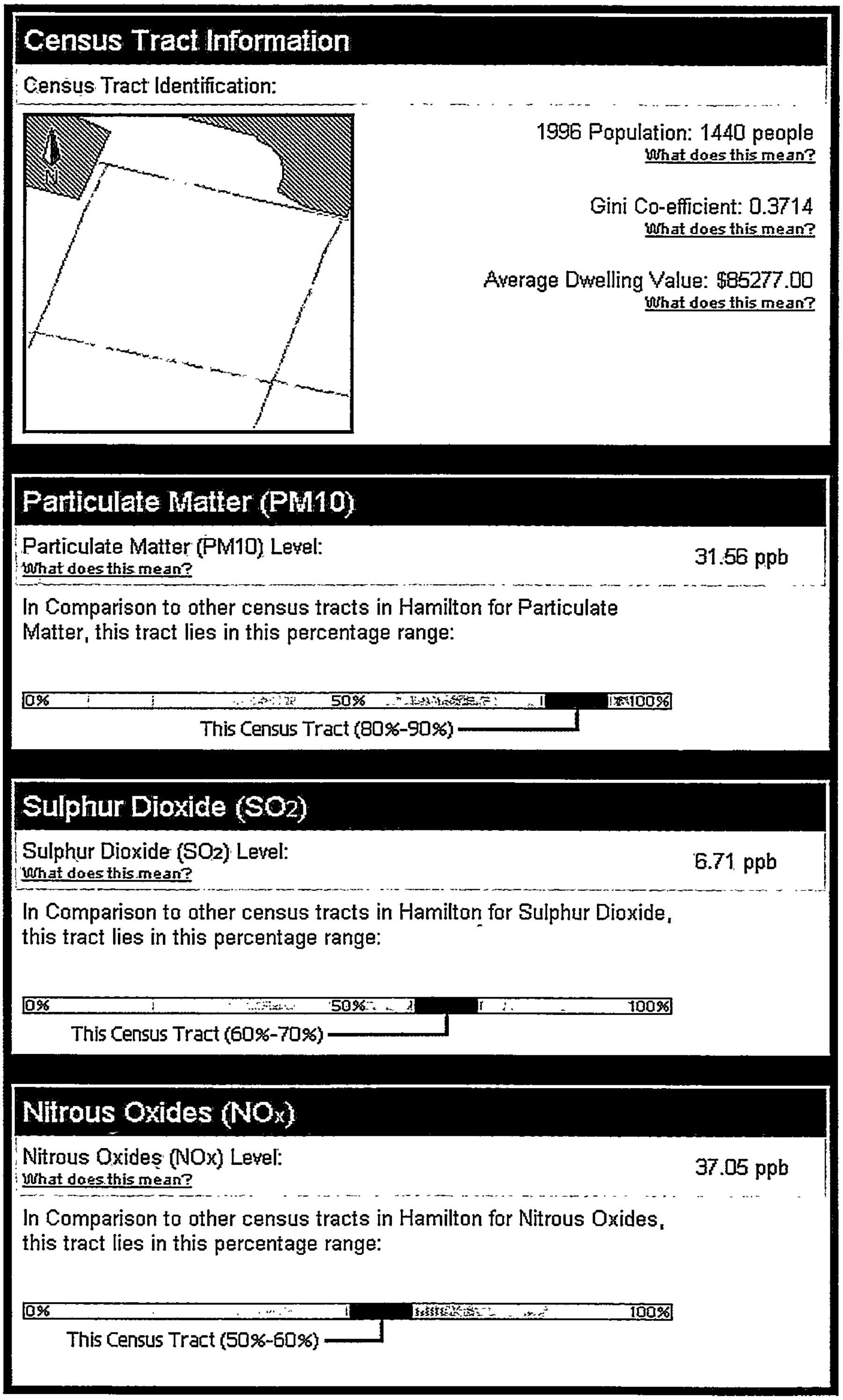


Figure 3.4 - Page that comes up when a census tract is queried

variables. With the variables entered the user must refresh the map. When the map is refreshed a JavaScript takes the entered values from the HTML code and transfers it to the Java Applet. The Java Applet then produces a new map which highlights census tracts meeting the users criteria. This tool allows the user to explore relationships in the data. For example, they could explore whether smoking rates are high in areas of low dwelling value or if asthma rates have correlation to high levels PM_{10}. Figure 3.5 illustrates this example. With the system set up to allow for the querying of the census data other layers of data needed to be added. Adding other layers of data is a standard process with MapServer. The software allows the user to turn on or off themes using check boxes. By clicking the box, a checkmark will appear and when the map is refreshed all layers with checkmarks will become visible on the map. The data for each layer is always available for the map but until the checkbox is activated the layer remains invisible. There are six layers available: the Niagara Escarpment, Arterial Roadways, Greenspace, the Industrial Area, Hospitals and Schools. Three additional layers compromise the base map, the census tract layer, Lake Ontario, and Hamilton Harbour. From a design standpoint, minimizing the number of checkboxes was desirable to keep the interface as unimposing as possible. There is no reason the census tract layer would need to be turned off as it contains all the important information in that layers associated database. The Lake and Harbour will never clutter the map as they do not overlap the census layer and they assist the user in locating an area. Two of the layers, 'Schools' and 'Hospitals' may only be viewed at high zoom. When the schools were first added as

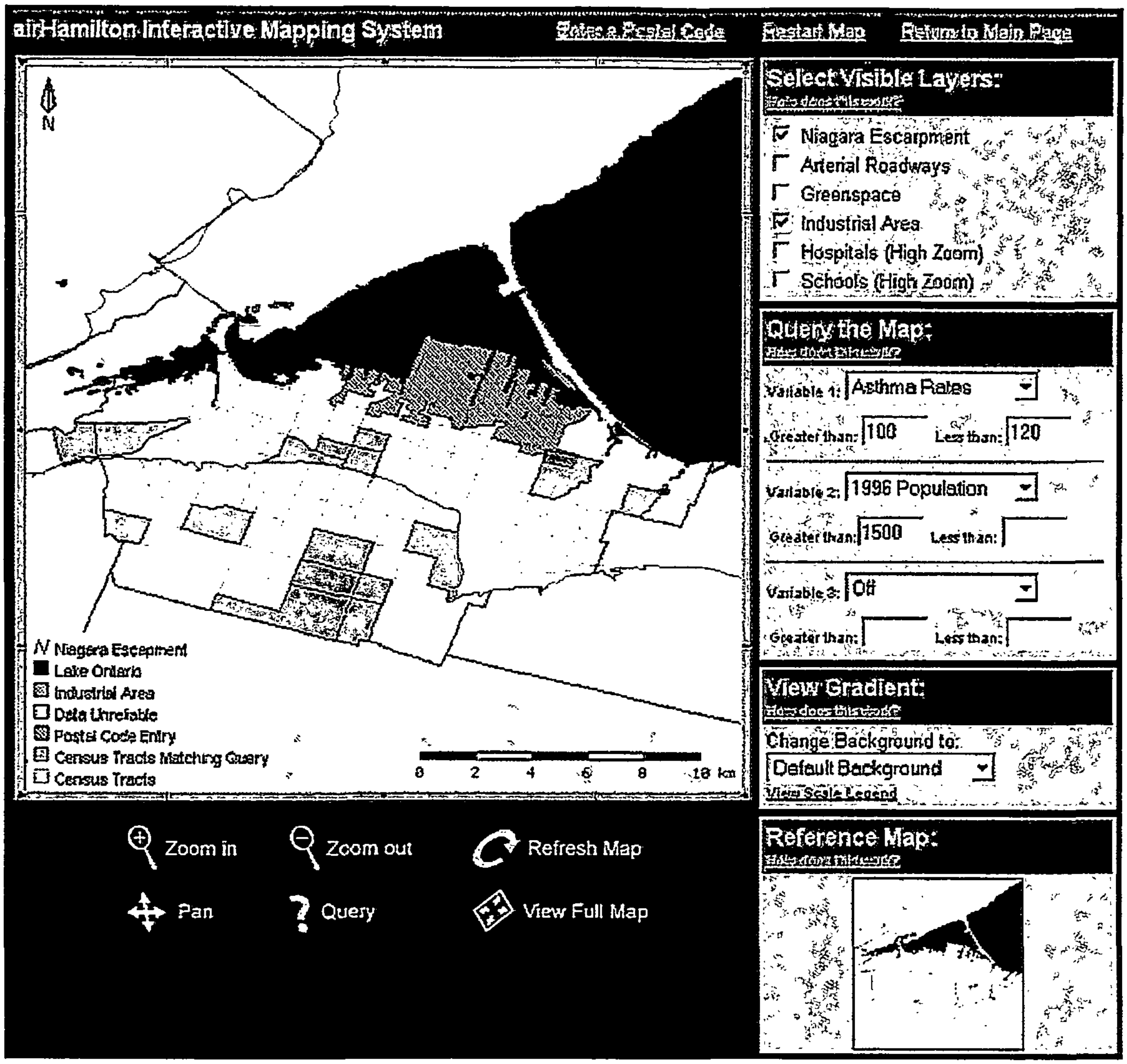


Figure 3.5 - Results from a multilevel query

a layer it was possible to view it at any scale. This cluttered the map. The schools are point pattern data and are represented as a small flag created as a 16-pixel image. The size of the image is static (i.e., it does not change with map scale.) When the schools were visible at full extent there were 145 images shown. This made it impossible to see anything on the maps other than a cluster of images and labels. By setting the maximum extent at which the schools can be viewed, cluttering can be avoided. All schools are shown with the school name so people may locate a school of interest. Hospitals are represented as polygons allowing the user to know the shape of the building. The polygon files are not static and change with the scale of the map. At full extent the hospitals are not visible, but they are made visible at high zoom (scale 1:2000). Setting the maximum viewable extent of the layers is an option to the programmer when the data layers are added.

3.6.2 Increasing the User Friendliness

The first step to increase user friendliness was to develop comprehensive instructions for use of the system. Accompanying all tools is a 'How does this work' link. This link opens a new window containing information about how the tool works. With the 'Query the Map' function a window opens showing the ranges of the datasets. The window that opens when the user queries a specific census tract contains numerous 'What does this mean?' links. These links accompany the values of all the reported variables and when clicked a new window opens to give the user a definition of the variable.

The second step in user friendliness improvement was to create dichromatic maps the user could view within the openGIS environment. To accomplish this, the asthma rate and dwelling value data was reclassified into seven classes using the natural breaks method in ArcView 3.2. The value representing the boundary value for each break was then transferred manually into Mapserver. The new boundary levels were then used to create a gradient map (Figure 3.6) to be viewed within the Java Applet. When the user chooses to view a gradient map the background layer changes to meet that choice. This new map acts just as the original census layer did (it can be queried and layers may be draped on it). Having a gradient map gives the users the opportunity to look for patterns in the data.

The last step to improve user friendliness was the 'postal code look-up.' The postal code look up allows the user to locate the census tract that houses a postal code of interest. The user must input the postal code before the MapServer generates the Java Applet. With the postal code chosen MapServer creates a red hatch on the matching census tract (Figure 3.7). To create this functionality all the postal codes in Hamilton had to be matched to a census tract. This was done using the Postal Code Conversion File (PCCF) from Statistics Canada. Using ArcView 3.2, a table was created matching all postal codes to their census tract. This table was uploaded to the computer acting as the WWW

Server. When a user enters a postal code the computer identifies elements of the database that should be highlighted. This aspect of the functionality was programmed in the computer language PHP. The census tract encasing the inputted postal code is then highlighted. When the prototype enters the public domain the ability to locate a postal code will empower users unfamiliar with cartography to obtain information about a neighbourhood of interest (e.g., the neighbourhood of the school their child attends). People may want information about their neighbourhood, or their child's school and this tool would assist them in achieving that.

The next step in improving the user-friendliness was development of an on-line tutorial for the user. The on-line tutorial was created, using Flash software (Macromedia 2002) by Amy Williamson, McMaster University as part of her multimedia thesis (Williamson 2002). The tutorial allows the user to get an automated demonstration of how each button works and what it does.

3.7 The Focus Group

To test the prototype and administer a questionnaire to obtain user feedback, a focus group was held November 2nd, 2001 at McMaster University. The focus group was held in partnership with the Central West Health Planning Information Network (CWHPIN), one of five Health Intelligence Units (HIU's) established by the Ontario Ministry of Health and Long-Term Care. CWHPIN is a partnership made up of seven public health

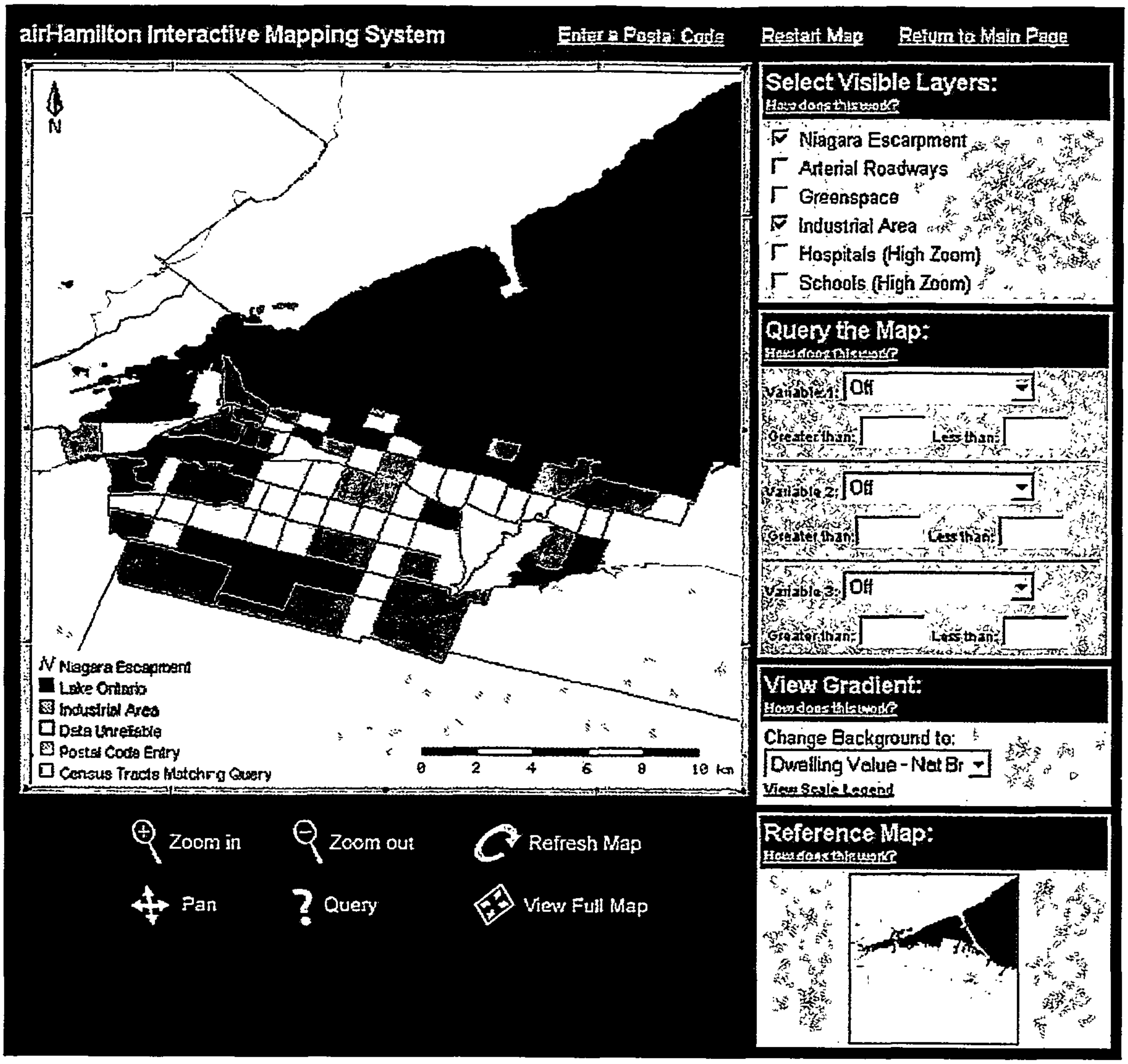


Figure 3.6 - Example of a gradient map showing average dwelling value throughout Hamilton

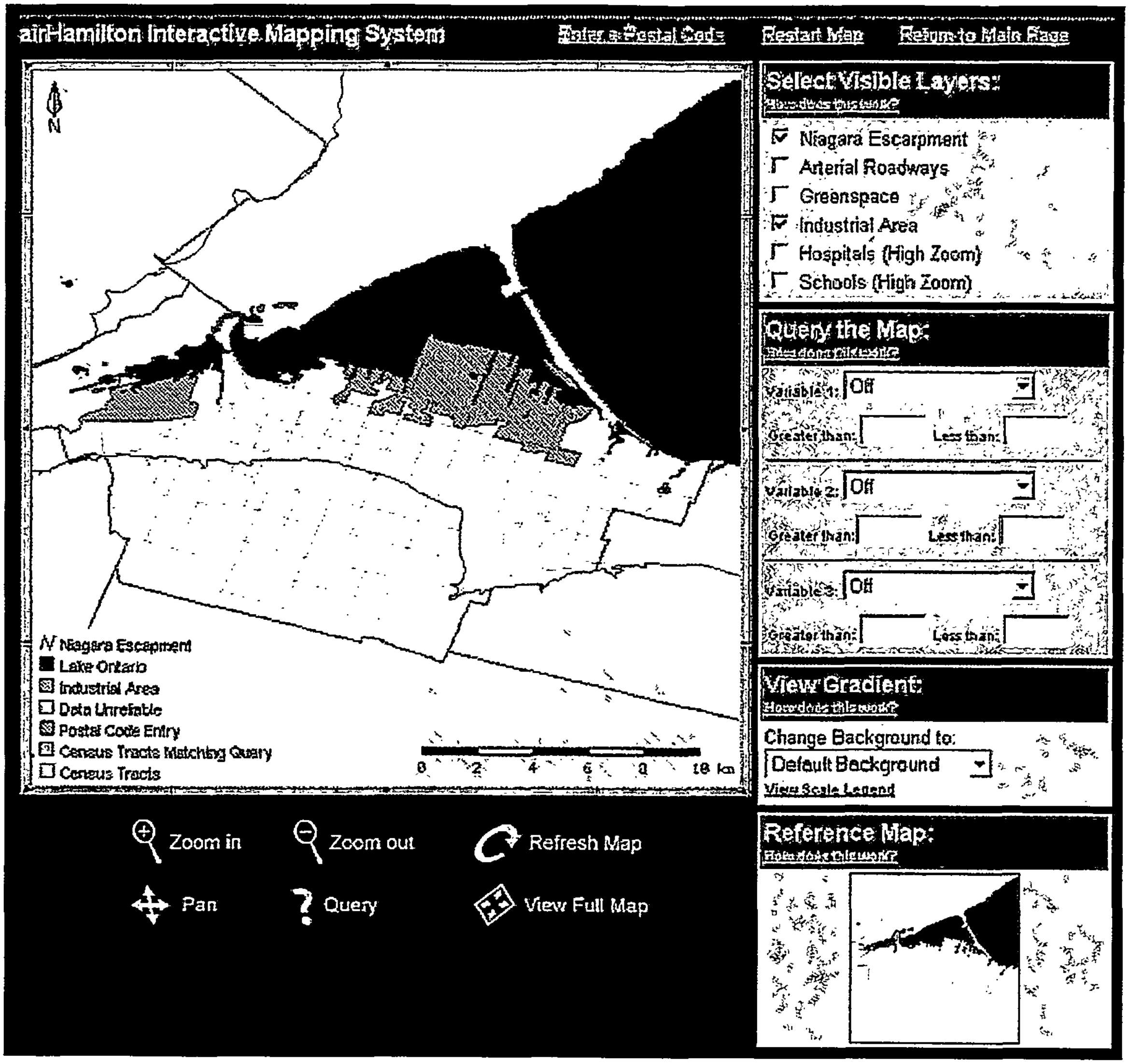


Figure 3.7 - Results from inputting a postal code

units and four district health councils, located in the districts of Brant, Haldimand-Norfolk, Halton, Hamilton-Wentworth, Niagara, Waterloo and Wellington-Dufferin and the McMaster University Faculty of Health Sciences (CWHPIN 2002). The director of CWHPIN, Dr. Tom Abernathy, asked their thoughts of openGIS in June 2001. It became apparent that while all seemed to have at least some interest, few people actually had an idea of the capabilities. With this in mind partners from CWHPIN were invited to attend the focus group.

The focus group took place at McMaster University within the Anne and Neil McArthur Humanities Multimedia Wing. It was chosen because it contains numerous, powerful computers. The lab contained 18 Silicon Graphic computers containing Pentium III, 800 MHz motherboards and 19 inch flat screen monitors. These computers are more than adequate to run the prototype.

Of the 27 people invited to the focus group 19 people attended. The focus group started at 9:00 am. For a full schedule refer to Appendix B. The focus group began with an address from Dr. Abernathy followed by presentations from Dr. Michael Jerrett, School Of Geography and Geology, McMaster University and Dr. Yvan Bedard, Department of Geomatics, Laval University. The intent of the presentations was to familiarize respondents with spatial analysis in Public Health, openGIS, and other applications. Following the presentations a demonstration of each the prototype and the on-line tutorial were shown to the attendees. With the presentations and demonstrations completed the focus group participants had the opportunity to explore the openGIS prototype. While the participants were testing the site the presenters stayed to answer any question people had about the presentations or the prototype itself. Following the site exploration, participants were asked to complete a survey with questions concerning their computer knowledge, openGIS, the prototype, and the focus group. The survey was designed with the help of Dr. Steve Manske of the University of Waterloo, and Sarah Wakefield from McMaster University. Survey results helped to give insight to how people felt about more than just the prototype. As well as determining feelings toward the prototype the survey was designed to establish base knowledge of GIS and openGIS. Appendix C contains the survey.

CHAPTER 4

Results

4.1 Introduction

This chapter presents results from the review of current applications of openGIS technology and the survey conducted at the focus group for public health professionals. It also documents changes made to the prototype due to those results are documented. As a consequence this chapter addresses three of the four research objectives of the thesis: (1) to document the current state of openGIS technology; (2) to assess the openGIS prototype created to explore health, pollution, and socio-economic data for Hamilton, Ontario; and (3) to improve the prototype based on findings from the focus group. As previously described in chapter 2, the review of current openGIS technology was conducted by evaluating a sample of current web sites against set criteria. This allows for applications to be compared against each other and assessed individually. The survey conducted at the focus group assessed the attendees background knowledge of GIS, thoughts on openGIS, thoughts concerning the prototype they tested, and the focus group itself.

The presentation of the assessment of current applications in openGIS is organized around five major sets of criteria, Metadata, Technical Considerations, Data, Cartographic Capabilities, Cartography, and Data Analysis. The survey results are organized to demonstrate the user's general knowledge of GIS, openGIS and their thoughts regarding the prototype. Their openGIS knowledge, thoughts and concerns towards the ethics of openGIS are also documented. The literature has revealed concerns regarding public access, differential access, and confidentiality (Crampton 1995, Curry 1995, Onsurd 1995, Sheppard 1995). Results of the survey will determine if public health professionals in Southern Ontario share these ethical concerns.

4.2 State of Current Technology

The results of the assessment of the current technology are summarized in Table 4.1. The Table shows the criteria in the subgroups described in Chapter 3, Metadata, Technical aspects, Data, Cartographic Capabilities, and Data Analysis. The web sites reviewed are numbered one through thirty. Table 4.2 shows the name of the organization along with the web address corresponding to each number. At the end of each row is the percentage of web sites that fit the criteria. The percentage only includes websites applicable to that particular criteria. For example, if a web site does not allow the user to query the data then query processing time would be a non-applicable category and assigned an N/A. The Table is organized by placing the subgroup meeting the highest percentage of total criteria at the top with the least satisfied subgroup at the bottom. Within each subgroup individual criteria are organized in this same way. For example the Metadata subset contained four criteria for a total of 120 possible fits (4 criteria multiplied by 30 websites). With 92 of the possible 120 criteria matching, the matching rate is 77%. Each subset is broken down by pointing out important trends.

Table 4.1 – Results from assessment of current openGIS technology

	Criteria Tested	Websites reviewed																														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	% fit criteria
Metadata	Authorship attribution	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	97%
	Disclosure of funding		X			X	X		X		X		X		X	X	X	X	X	X		X	X	X		X	X	X	X	X	X	70%
	Currency of data	X	X	X		X		X	X	X			X			X	X	X	X	X	X		X	X		X	X	X	X	X	X	73%
	Attribution of sources		X	X		X		X	X	X	X		X			X	X	X	X	X	X	X	X	X			X	X	X			67%
Technical Aspects	Site accessibility	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	97%
	Ease of Navigation		X	X		X	X		X		X		X	X		X	X	X		X	X	X		X	X	X	X	X				63%
	Map upload time	X	X				X	X	X	X	X	X	X				X				X		X	X						X	X	50%
	Speed of site startup	X		X		X		X	X				X	X			X			X	X	X			X	X				X	X	50%
	Query process time	X	N/A			N/A	N/A	N/A	X	N/A	N/A	N/A	N/A		N/A	N/A	N/A	N/A	N/A		N/A	N/A	N/A	N/A	N/A	X	N/A	N/A	N/A	N/A	N/A	43%
Data	Links to other sites	X	X	X				X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	87%
	Interactive search				X			X	X		X		X	X		X	X		X	X	X			X			X	X	X			50%
	Confidentiality explicit		X											X						X				X			X	X				20%
	Downloadable data			X		X								X		X				X				X								20%
Cartographic	Scale Capabilities			X		X			X		X	X	X	X	X	X	X	X	X	X	X	X			X	X	X				X	63%
	Colour scheme availability			X																												7%
	Graphing																						X									3%
Data Analysis	Interactive overlays					X			X			X		X		X		X		X	X					X						30%
	Boolean queries	X		X	X				X					X						X						X						23%
	Reclassing data								X																					X		7%
	Pattern recognition																															0%
	Advanced regression																															0%

Table 4.2 – The websites assessed as current applications. The number in the left hand column corresponds to the numbers on the top Table 4.1. The full web address for each application is located in the reference section.

1	The University of Leeds
2	Statistics Canada
3	The American Cancer Atlas
4	Nevada Ecosystem Project
5	United States Census
6	CIET
7	Arizona State University
8	The City of San Diego
9	NBCi
10	Microsoft Terraserver
11	Ecotourism
12	California Resource Evaluation System
13	Friends of the Earth
14	Information Center for the Environment
15	Environmental Protection Agency
16	National Geographic
17	The city of Sacramento
18	Nursing Home Information
19	The World Health Organization
20	The Artic Mapping System
21	United Nations Environmental Program
22	Austrian Federal Environmental Agency
23	Children's Report Card
24	Humboldt Bay
25	USA Department of Environmental, Food and Rural Affairs
26	Health Canada -Disease Surveillance On-line
27	Health Canada - FLU watch
28	Environment Canada - Species at Risk
29	Census Mapping ESRI
30	US. Center for disease control

4.2.1 Metadata

Metadata is data that describes geospatial data. Metadata tells you the 'who, what, and where' of a geospatial dataset (Intergraph 2002). Of the criteria subsets this was the most complete with 77% of total criteria matched. 'Authorship attribution' is the criterion which had the highest matching rate at 97%. This was expected as most who create a website want to be recognized for funding and promotion purposes. The majority of reviewed sites disclosed their funding (70%) and the currency of data (73%). Within this subset the least matched criterion is attribution of sources for data or the literature (67%). Incomplete metadata means the site is of limited use to external researchers because not knowing the data collection and implementation history inhibits future use of the data. This finding emphasizes the need for clear and complete metadata in the development of future sites.

4.2.2 Technical Aspects

Technical aspects matched 65% of the total criteria. The 'Site Accessibility' criterion was met by 97% of sites tested although this result is somewhat misleading. A site failed the criterion if any aspect was password protected. For example, if a password was needed to download data this would warrant the site failing the criterion. While few sites had aspects that were inaccessible, there were sites which were completely inaccessible. The completely inaccessible sites were not included in this study because other criteria could not be tested. 'Ease of Navigation' is the only criterion that was dependent on personal opinion. It was deemed that 63% of the sites could be navigated easily.

Within the 'Technical Aspects' subset the three criteria which are measured by time are present. The 'Speed of Site Start Up,' 'Map Uploading Time' and 'Query Process Time' criteria all matched 50% of the time. Each site criterion was tested from the same computer at approximately the same time of day (Pentium 3 600Mhz between 3 – 6 pm). After all were tested the average was taken and those below the average matched the criteria. Numerous sites did not have query capabilities. Consequently these cases are defined as not applicable (N/A). The average speed of start up was 6.7 seconds, map uploading time 7.4 seconds, and query processing time 10.7 seconds.

4.2.3 Data

The Data subset matched 44% of the total criteria. The majority of sites (87%) have links to other sites. Exactly half (50%) contained the ability to search themselves. The ability to search the entire site interactively which houses the openGIS is not a popular function. How issues of confidentiality are dealt with were rarely explicit (20%) and the ability for data to be downloaded (20%) was just as rare.

4.2.4 Cartographic Capabilities

The criterion for 'Cartographic Capabilities' had a matching rate 24%. The 'Cartographic Capabilities' subset is a technological intensive set of criteria. The ability to produce graphs (3%) and alter colour schemes (7%) are not typical openGIS functions. About 63% of sites had capacity for scale adjustments (i.e., zooming in).

4.2.5 Data Analysis

The 'Data Analysis' subset matched 12% of the total criteria. Only three criteria, 'Boolean Queries' and 'Interactive Overlays,' were matched with any regularity, at 23% and 30% respectfully. Of the final three criterions 'Reclassing Data', 'Pattern Recognition', and 'Advanced regression' only one had any matches. The ability to reclass data was found in 7% of studied sites.

4.2.6 Synthesis of the State of the Technology

Examination of the current state of the technology greatly influenced the design of the health openGIS prototype. The overall state of the technology is that openGIS is underutilized. The majority of openGIS sites are technically sound and contain adequate metadata. These are key issues if the technology is to be used properly, but they are also the easiest to implement. Less then half the sampled sites had the ability to query or overlay data. While it was not a criteria, there was not one sampled openGIS application that allowed for multi-level queries. That is to query the data using multiple criteria to pinpoint areas of interest. This was the most striking gap in functionality that proved to be realistic to implement during this Master's thesis timeframe. More generally, the prototype website attempts to give the user as much information, as quickly as possible. This can be done by changing the 'Background Gradient' or 'Querying the Data.' This is in contrast to the majority of current applications that have static maps and no numerical data.

4.3 The Focus Group Survey

The focus group held on November 2nd, 2001, allowed public health officials to work on a prototype site and voice their views of the current and future direction of openGIS in their profession. Quantitative results in the form of statistics, and qualitative results in the form of comments to open-ended questions from the survey administered that day are reported.

4.3.1 Knowledge of the Technology

Public health departments in North America underutilize available technology, in particular GIS (CDPHE 2001). Coupled with underutilization there is a problem with differential access to the technology (Croner 2000). To test these ideas questions were directed at testing basic knowledge concerning GIS and perceptions of how this technology could help in decision making. Surprisingly most participants believed they had a good understanding of GIS, with 62% of the respondents agreeing or strongly agreeing with that statement. Not surprisingly as the complexity of the technology increases the perceived understanding decreases. When asked if they had a good understanding of Internet Mapping, only 42% of respondents 'agreed' or 'strongly agreed.' The downward trend of understanding continued with only 38% of respondents having a good understanding of OLAP.

While the understanding of GIS and GIS related technologies is not constant among public health professionals, most believe that the technologies could assist them with their work. To the statement 'I think using spatial information in my work could help me do my job better,' 56% of respondents 'strongly agreed' while the other 44% 'agreed.' Having 100% of respondents considering spatial information as a potential benefit is not unexpected. Public health journals often contain papers which show how spatial data contributed to the solving of a problem (e.g. Grossman 1999). Prior to the focus group the annual conference of the Association of Public Health Epidemiologists in Ontario (APHEO) was based on the use of GIS in Public Health. The meeting was titled 'GIS: Tools for Mapping Public Health.' This illustrates how the public health profession is embracing GIS as an analysis and data management tool. The survey allowed the respondent to share their thoughts on the benefits of spatial data in an open ended question. Sentiments about the potential usefulness of the technology were similar with all respondents agreeing that spatial information can be beneficial. The following two examples essentially encompass all the ideas put forward:

> Ned: ...rather than summarizing the information in a table format and thinking in the abstract (i.e. referral patterns) I would be able to overlay several variables of interest to show diagrammatically what the situation is. Patterns would become more obvious, not only because of pictorial representation but also due to the overlay of several indicators.
>
> Maude: There is no doubt GIS would be extremely useful for my work as an epidemiologist for a public health unit. The key for me would be the expense for GIS mapping. I work in a small health unit and therefore cost is a key factor to consider.

Each of the open ended responses contained at least one of the subsequent ideas: visualization, pattern recognition, or cost. The idea of the costs associated with spatial information must be considered important with respondents as they discussed the drawbacks of cost when the question was asking for potential benefits.

Given the visibility of GIS in Public Health fields, strong support for the use of spatial data was expected. It was surprising, however, to find an even higher percentage (67%) 'strongly agreed' that Internet mapping could help them in their work. With spatial information being an integral element of Internet mapping technology one would expect at most only those who 'strongly agreed' spatial information would be beneficial to their work would also be as positive with Internet mapping. With this question there was again an opportunity to share unstructured thoughts. The main ideas revolved around data sharing, data access, ease of use, and the presentation of spatial data to lay audiences:

> Jason: ... a way to visually communicate data and trends to lay people and council members.
>
> Amy: ... Internet mapping would be able to provide us with more information & expand the use of data sets to more effectively provide more detailed information of the community (especially is upload of data sets is a capability of internet mapping). The ability to share data is also of importance and would assist the health department's ability to obtain additional resources that can be incorporated into our work.
>
> Katherine: Benefit of Internet based mapping would be the number of individuals that could be trained in GIS & the simplicity with which this could be done. It would also make mapping a much more accessible resource.

This sample of the typical open-ended response gives an initial indication that openGIS may make GIS and consequent data more assessable. In turn it may contribute to the growth of GIS in Public Health. It is a simple equation; an increase in access equals an increase in use which will transform into growth of new techniques and ideas to improve the tool.

To prompt consideration of the potential benefits of Internet mapping, respondents were asked to rate how strongly they agree or disagree with eight different statements. The percentage of people who 'strongly agree' or 'agree' can be seen in Table 4.3. In all cases, the majority of respondents agreed with the statements. Of the eight statements, the ability to perform advanced statistics had the lowest rate of agreement with only 69%. Many focus group attendees may have little background knowledge in advanced statistics, thus explaining this result. Surprisingly the ability to produce paper maps was the second lowest rated benefit with 75% of users agreeing. Most of the respondents communicated that increased access to maps and the ability to share data was vital and a paper map is one way to accomplish this. Earlier it was established that public health professionals do see benefits with spatial data. This is reiterated in this section, 100% of respondents believed that the ability to explore spatial relationships was a benefit of Internet mapping.

In an attempt to understand the respondent's perceived problems with Internet Mapping technology, they were questioned with eight potential shortcomings (results in Table 4.4). The responses reinforced results of the NHIS user needs survey and those which are found in the literature. Over half of the respondent's all 'agreed' that 'Limited Data', 'Cost Involved', 'Access to Data', and 'Staff Resources' are potential problems. The similarity of all these problems is data or cost. This result is consistent with past research (Gosselin 2000). No other potential problem was 'agreed' upon by more than 33% of respondents. Just over 20% felt that Internet mapping is difficult to use, based on their limited experience with the prototype. Considering 82% of respondents found 'Ease

Table 4.3 – Responses to the survey question 'I see the following potential benefits of Internet Mapping'

	Strongly Agree (%)	Agree (%)	Neither Agree nor Disagree (%)	Disagree (%)	Strongly Disagree (%)
Increased access to maps	68.4	26.3	5.3	0	0
Increased access to data	52.6	36.8	10.5	0	0
Ability to share data	47.4	42.1	10.5	0	0
Ability to produce paper maps	47.4	26.3	26.3	0	0
Ease of use	68.4	15.8	15.8	0	0
Explore spatial relationships	68.4	31.6	0	0	0
Perform cluster analysis	47.4	36.8	15.8	0	0
Perform advanced statistics	26.3	42.1	21.1	5.3	5.3

Table 4.4 - Responses to the survey question 'I see the following potential shortcomings of Internet Mapping'

	Strongly Agree (%)	Agree (%)	Neither Agree nor Disagree (%)	Disagree (%)	Strongly Disagree (%)
Limited data available	15.8	36.8	21.1	21.1	5.3
Difficult to print maps	11.1	50.0	27.8	11.1	0
Difficult to pinpoint an area	5.6	16.7	38.9	33.3	5.6
Difficult to use	5.6	16.7	27.8	44.4	5.6
Limited analytical capabilities	11.1	22.2	38.9	11.1	16.7
Cost involved in startup	5.9	47.1	23.5	17.6	5.9
Access to data	5.6	50.0	22.2	22.2	0
Staff resources	27.8	33.3	16.7	22.2	0

of Use' a potential benefit of Internet mapping, it is reasonable to expect about 18% would see this issue as important. Although only 11% of respondents felt a difficulty to print maps would be a potential problem, there is a likelihood that they did not completely understand how Internet mapping worked. Many Internet users are familiar with copying an image from their web browser to another document by right clicking the mouse. With Internet mapping this can not be done as the maps are produced as a Java Applet. The Java Applet does not allow itself to be copied in that manner. Comments from respondents reinforced the ideas extracted from Table 4.4:

> Jim:...cost will be an important factor...limited amount of data...it will be extremely important to keep all data sources up to date, for example vital statistics should be latest available.

> Chris: The branch in which I work supports a wide range of departments/services. A barrier/limitation in using Internet mapping might be having the time to keep updated/current with any changes to Internet mapping service. Its use would be for specific/limited areas of work.

These comments give another indication that the issues involving data and cost are foremost in on the minds of the respondents. Similar to the national user needs assessment, the institutional barriers of cost and data emerged as more important issues than the technology itself.

In addition to determining the perceived benefits and problems with openGIS technology it is advantageous to knowing the potential uses of technology. By determining what public health professionals may use the technology for, future improvements may be tailored to their needs. The results of this line of questioning clearly indicated how this technology will be used. The majority of respondents (83%)

felt that the technology could be used for data visualization and collaboration among colleagues. Slightly fewer respondents (78%) 'agreed' that the tool would be used for data exploration with spatial overlays and data analysis. These results describe a general trend: that public health professionals would like to use this technology mainly for the purposes of data visualization.

4.3.2 The Prototype Itself

Participant views concerning the prototype itself were also questioned. The first six questions of the section were all based on thoughts and issues concerning the health openGIS prototype. As can be seen in Table 4.5, the question did not deal with the functionality of the prototype. The first three questions, 'I found the on-line tutorial helpful', 'The prototype is easy to use', and 'The prototype was intuitive in its use' allow the respondent to decide if a lot of time would be needed to learn how to use the prototype. The longer it takes the user to learn the program the more it will cost in terms of salary and lost opportunities. Reaction to these questions was positive with 79% of people finding the on-line tutorial helpful and 95% feeling the prototype was intuitive in its use and easy to use.

To test opinions on the aesthetics of the prototype the survey questioned the colour scheme and design. Fully 100% of respondents 'agreed' that the site

Table 4.5 - Responses to the survey question 'I see the following feelings about the prototype itself'

	Strongly Agree (%)	Agree (%)	Neither Agree nor Disagree (%)	Disagree (%)	Strongly Disagree (%)	Don't Know (%)
I found the on-line tutorial helpful	15.8	63.2	10.5	0	0	10.5
Easy to use	36.8	57.9	5.3	0	0	0
Intuitive in its use	26.3	68.4	5.3	0	0	0
Appropriate colour scheme	21.1	63.2	10.5	0	5.3	0
Layout was appropriate	10.5	98.5	0	0	0	0
No problem in full public use	10.5	31.6	31.6	5.3	10.5	10.5

design was appropriate. This bodes well for improvements in the prototype. Adding functionality to the technology is a challenge but to completely alter the site design would be extremely time consuming. Not quite as many 'agreed' that the prototype was using an appropriate colour scheme. While 84% of users agreeing is still a large majority, some had issues with colours chosen. The following two comments taken from the survey relate these feelings:

> Thelma: ...would have liked water to be "softer" blue.
>
> Louise: ...colours are confusing. Green is too much like greenspace, perhaps stick to more traditional colour schemes.

As explained in chapter two, the more traditionally accepted colour schemes such as red to blue (with red representing a high risk area) may be alarming to some and are not easily distinguishable by people with colour blindness. The purple to green scheme used was not randomly picked, but was chosen based on a review of the issue in the literature (Brewer 1999). Using the research to create the colour scheme the program vischeck (Vishchek 2001) was used to test the site. Vischeck's color vision model allows you to simulate how the world looks to people with various sorts of color vision deficiency. Figure 4.1 portrays the prototype's appearance to a person with the following types of colour blindness, protanope, deuteranope, and tritanope. Brettal (1997) determined that protanope and deuteranope, two of the most common types of colour blindness, can be effectively simulated with digitized images. Tritanope is a rare type of colour blindness, but it is also more severe than protanope or deuteranope (Arden 1999). While the appearance of the prototype does change, different layers and classifications are still distinguishable.

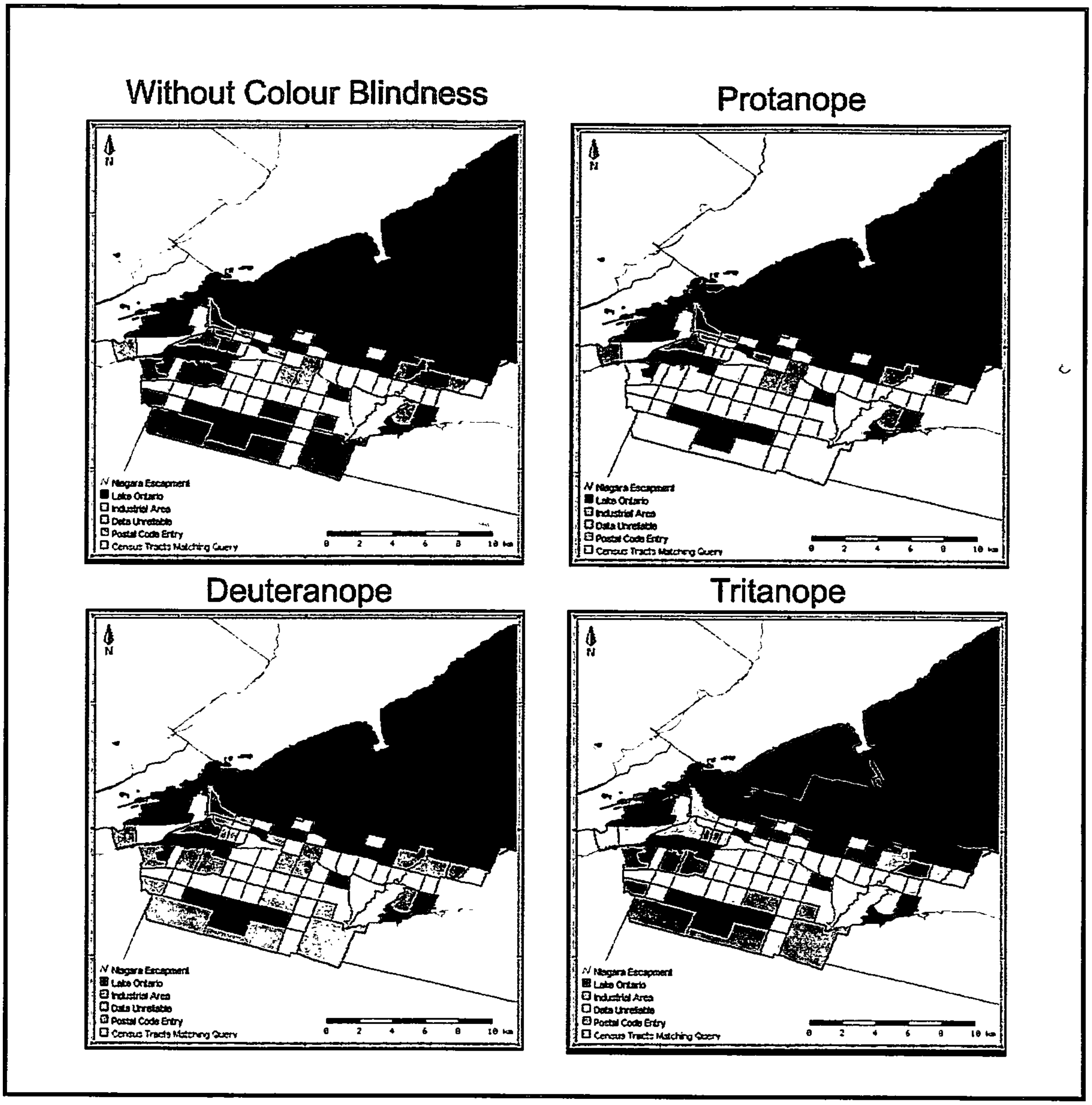


Figure 4.1 - The look of the prototype with various types of colour blindness

The greatest concern with the prototype in this line of questioning was allowing public access of the information. Only 42% of respondents 'agreed' that there would be no problem with allowing full public use of the site. Comments on this subject were sparse but all indicated that public access needed to be approached with care:

> Amanda: Access to the public needs care.
>
> Peter: Public use, confidentiality is an issue.

This is inconsistent with the review of the literature where issues of confidentiality are a major concern. The federal user needs assessment also found only a minority of respondents concerned about confidentiality issues.

The existing functionality of the prototype was well received (Table 4.6). Simple functionality such as overlaying different layers and zooming in and out was liked by 100% of respondents. The ability to query census tracts and perform multi-level queries was slightly less accepted at 95%. The ability to locate a census tract by postal code was the least appreciated functionality (85%) and prompted the following comment from one respondent:

> Jane: Postal code was not valuable for me as a non-resident. It would have been better in my area.

The intent of the question was to gauge the potential usefulness of the tool as it applied to any area of interest but this was not clear. The overwhelming support of the existing functionality demonstrates the importance of supporting site development with exclusive review of existing sites, user needs, and cartographic literature.

Table 4.6 - Responses to the survey question 'I see the following feelings about the existing functionality of the prototype'

	Strongly Agree (%)	Agree (%)	Neither Agree nor Disagree (%)	Disagree (%)	Strongly Disagree (%)
Ability to look up postal code	47.4	36.8	10.5	5.3	0
Overlay different layers	73.7	26.3	0	0	0
Zoom in/ Zoom out	63.2	36.8	0	0	0
Query single census tract	63.2	31.6	5.3	0	0
Multi-level Query	84.2	10.5	5.3	0	0

The next step in the evolution of the prototype will add new functionality. All the 'simple' functionality has already been implemented, so anything additional is a major undertaking. To ensure the desires of the public health professionals is known, a line of questions concerning the thoughts towards hypothetical functionality was issued. The two most 'agreed' upon ideas were the ability to upload data and the ability to produce graphs, both at 95% acceptance. These two ideas garnered much conversation at the focus group and were well represented with the comments:

> Janet:...upload would allow for custom tailoring of variables by individuals, hence increase functionality.
>
> Christine: Ability to produce graphs/charts would add to the functionality and usefulness of using Internet mapping.
>
> Jack: The mapping application should be able to accept a data file – ex. (sic) Specifications provided regarding the structure of the data file, what variables in what order and how labeled.

Having the ability to upload data and view it instantly would be unique to Internet mapping programming. The majority of respondents were interested in all the hypothetical functionality: detect clusters (90%), interpolate values (85%), alter colours (79%), and measure distances (75%). To implement any this functionality is unrealistic within the timeframe of this Master's project. However, it will be important for future projects involving the study of openGIS technology.

4.3.3 Focus group and Future Directions

The last section of the survey questioned the perceived value of the focus group and potential future use of openGIS technology. In both instances the results were favourable. The focus group was deemed useful by 94% of those who attended. To the question 'Assuming that a full set of data pertinent to my geographic region is available, I would use Internet mapping in the next year,' 22% of respondents 'agreed,' with 72% strongly agreeing. This indicates that people do see the value of this technology. Reinforcing this observation is the result showing that 100% of respondents felt confident that they could effectively us an Internet mapping system.

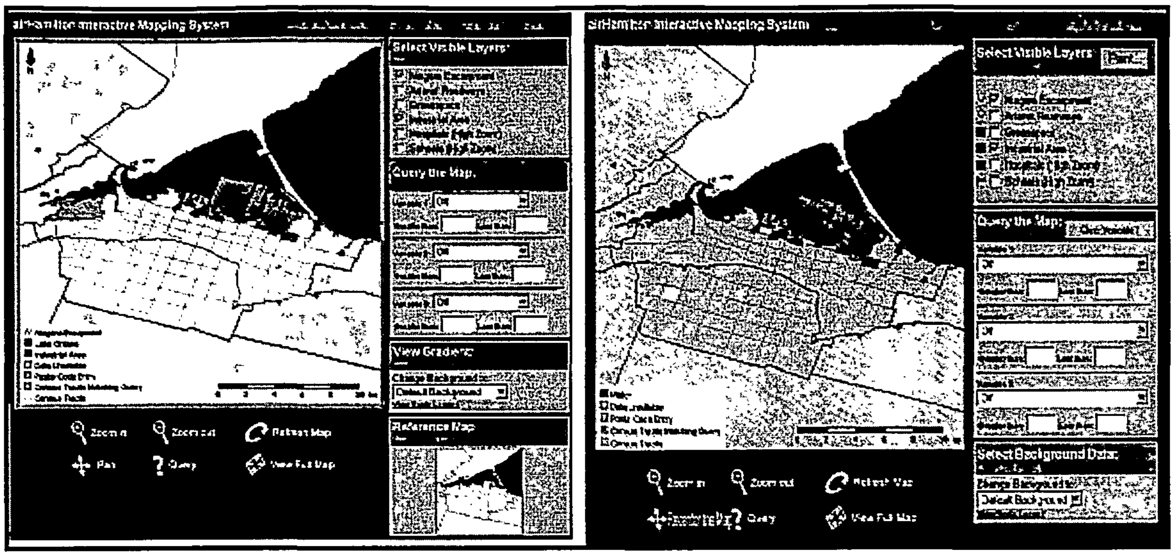
4.4 Improvements to the Prototype

When the focus group was completed and the survey results quantified, the next step in the project was to use those results to improve the existing prototype. Improvements were constrained by those that could be implemented and tested within the fixed term of this Master's program. There was not enough time to implement the larger scale modifications such as adding a data uploading system or allowing for reclassification of data 'on the fly.' It was also important to assess the improvements that were suggested. Many people decided that the purple-to-green colour scheme was not the most aesthetically pleasing and accessible to all, but this was not the reason for that choice. The colours were chosen to be the most effective, not the most pleasing to the eye. There was also the request to allow the user to choose the colour scheme and this is also undesirable. It is considered detrimental to give the user control over the colour of the mapping technology, as they may produce misleading maps (Kraack 2002). The changes that were made increased the intuitiveness of the prototype.

To illustrate the changes made Figure 4.2 shows two screen shots of the prototype showing one from November 2001 and the other from March 2002. The changes are listed underneath the comparison screen shots. The order of the changes is arbitrary and has nothing to do with importance.

The first two changes involved the naming of some functionality. Some users were confused by the term 'pan' which is a common GIS term but is not known in the field of Public Health. To alleviate the confusion of this term it was changed to 'Re-center the map.' This term is used on the Environment Canada 'Species at Risk' Interactive Mapping System (Environment Canada 2002) and Health Canada's 'FLUwatch' mapping system (Health Canada 2002). The next change involved the term 'View Gradient' which was also deemed to be counter intuitive. The 'View Gradient' functionality was changed to 'Select Background Data.'

The next step was to delete superfluous features from the prototype. To keep the program working quickly over the Internet every bit of information and functionality believed irrelevant needed to be erased. Two aspects of the prototype did not justify the extra loading time. While the reference map was aesthetically pleasing and would occasionally allow people to locate themselves within the City it did not add much to the prototype and may have caused confusion. In a project encompassing a larger area such as Ontario the reference map would be beneficial but in the context of a city it was not. The pan bars surrounding the Java applet were also deleted. With the 'Re-center the Map' button there was no justification to keep the pan bars. While they were useful initially, it was a duplication of functionality.



1.	'Pan' changed to 'Recenter the Map'
2.	'View Gradient' changed to select background
3.	Reference map is deleted
4.	Pan bars deleted
5.	Print Map button added
6.	On-line tutorial link added
7.	Clear Variable button added
8.	Escarpment made thicker and changed to black

Figure 4.2 - The look of the original and altered prototype. Included is a list of changes made.

Not all changes required deleting functions. The survey of public health officials showed that many public health professionals would use openGIS technology for data visualization. To assist this a 'print map' button was added. This button takes a snapshot of the Java Applet and sends it to a new window. From this new window the map may be printed using existing technology in Internet Explorer or Netscape Navigator. The map may also be copied from the window to another document (e.g. added to a PowerPoint presentation). The functionality behind the button was encoded using Java and HTML.

A key addition was adding a link to the on-line tutorial. Initially the only way to view the tutorial was by returning to the home page. It was inaccessible from the mapping page. With 85% of users finding the tutorial useful there was justification to make it more accessible.

The last button added to the prototype was 'Clear Variable' within the 'Query the Map' functionality. The need for this button came not of the survey but as an observation. When entering a variable to query the map, the selected variables stay visible after query is completed. This will not change as users may want to remind themselves of the exact query they entered. What changes is the user has the option of clearing the variable with a click of a button as opposed to erasing each field. This button allows easier use of the query function. The button was not the only change made to make the query function easier to use. Data ranges for each variable were added beside them in the selection box. For example, in the variable 1 drop down box beside smoking rates are the numbers 130-420. This allows the user to quickly know the range of the variables. Initially the ranges were available in the ‘How does this work’ link only. The link remains for two reasons. First, to show how the query function works and, second, to give the units for each variable. There is no way to fit the variable units within the drop down boxes without extensive programming, and thus, the link must remain.

The ‘query a census tract’ functionality of the prototype was also tweaked. The web page that opens as a result of querying a census tract gives information about that census tract. A concern brought up at the focus group was that upon querying a specific census tract, not all the information concerning the area are disclosed. Originally the pollutant data, along with the Gini Coefficient, Dwelling Value and population was available to the user. This has been changed, now each available variable is shown.

Originally there were only two backgrounds available within the ‘View Gradient,’ dwelling value and asthma rates. The original backgrounds were classified using the natural breaks method into seven classes. Users questioned why only those two variables were available. To address this, each variable was added as a background, including: sulphur dioxide, particulate matter, nitrous oxide, 1996 population, smoking rates and gini coefficient, was added as a background. The classification method was changed to quintiles and the number of classes dropped from seven to five. The classification method was changed because few public health professionals understood natural breaks. The number of classes changed from seven to five to simplify the results visually. With five classes there are two above and below the middle (neutral) classification. This is faster to load than seven classifications while showing the same trends.

The last group of changes deal with the look of the prototype. The 'Select Background Data' box was moved up to be in line with the top of the Java Applet. This places the boxes housing functionality in order from smallest to largest (see Figure 4.2).

The line representing the Niagara Escarpment was changed in colour and size. The colour was changed from red to black. The thickness of the line was altered from two pixels to four. The increase in line thickness makes the escarpment more predominate.

One of the verbal suggestions taken from the focus group was to make the shading which occurs when a census tract matches an inputted query hatched as opposed to solid. When solid the background data cannot be scene. This is unimportant when the default background is on, but when the background is a selected variable from 'Select Background Data' it takes away from the map. To solve this problem the prototype was encoded to have the census tracts matching the query show as an orange hatch. The orange hatching can be clearly seen over the default background and all other background colours.

The final change to be noted involves the rearrangement of the map legend. This change stems from the complaint of the 'View Scale Legend' link within the 'View Gradient.' This link was used to give the user the ranges for each classification. The ranges were presented in this form as opposed to being placed in the map legend. When in the map legend, the legend was large and overlapped the census tracts. With no way to control the size of the legend the ranges for each class were moved. People found the 'View Scale Legend' link awkward. When the legend was opened it worked well, but when the user clicked again on the map the legend disappeared. To combat this problem the legend, was completely rearranged. Instead of having each visible layer in the map legend the symbol or colour representing each visible layer was moved to the 'Select Visible layer' box. This leaves the map legend with the background layer, census tracts matching query and postal code entry. With the legend shorter and the number of classifications changed from seven to five, the ranges could be included in the map legend. This allows for the 'View Scale Legend' to be erased.

4.5 Summary of Results

The review of current applications illustrated how openGIS has been limited, in most cases, to simply showing a static map. Methods to explore the data through queries, graphs, or statistics are extremely limited. It was this overall lack of ability to explore the data that led to the creation of the health openGIS prototype.

The focus group survey reiterated trends uncovered in the literature review. The survey showed Public Health professionals want to further incorporate spatial data and GIS into their work. Focus group attendees felt openGIS as a tool has the potential to bring affordable GIS to their jobs. A full 100% felt confident they could use openGIS effectively. The survey also gave people the opportunity to critique the health openGIS prototype. Improvements made to the prototype were based on concerns from the focus group.

Chapter 5

Discussion

5.1 Introduction

The strength of this thesis originates from a series of ideas that build upon each other. This process is similar to the 'waterfall' model of software development put forward by Royce (1970) and further explained by Sommerville (1992). The initial literature review along with the NHIS's user needs assessment identified GIS and spatial data as an important topic in the world of public health. That same study showed that the major concerns are data (quality, source, etc) and cost (implementation, training, upkeep, etc). With articles in trade journals such as Geoworld speaking of the revolution of openGIS (McKee 2001, Culpepper 1998, Plewe 1997) peaking interest in exploring this new technology in the world of public health. Investigating the current state of the technology in relation to health information gave insight into trends and shortcomings of openGIS. From this information the task of creating an openGIS prototype for public health officials began. To ensure the validity of the prototype the focus group which gathered on November 2nd, 2001 to praise and criticize the prototype. This chapter will discuss the results from all components of the thesis including the key findings and how they advanced knowledge. To ensure the validity of the results potential influences on data collection are also discussed.

5.2 Key Findings

The range of findings in this thesis reflect the waterfall methodology (Figure 3.1). At each of the major steps there was a major finding(s) which pushed the idea forward. The literature review illustrated how commonly GIS and spatial data is in public health projects. More specifically, the unpublished NHIS study illustrates and reinforces many of the findings from the literature based on a purposeful sample of Canadian public health professionals. With the majority of journal articles being American in origin, the Canadian study shows that similar ideas apply in both countries. Three key findings emerged from the literature review: (1) the desire of public health officials to incorporate spatial data into their work, (2) perceived problems with the cost of implementing GIS, and (3) limitations inherent in current geocoded health data.

Review of the current openGIS applications is included in this first step (Figure 3.1). Spatial data and GIS are desired by public health officials but the cost of equipment and data is an issue. OpenGIS technology has the potential to solve these problems if used correctly. In reviewing the current openGIS applications the key findings quickly became apparent. The surprisingly difficult task of locating openGIS applications awarded few examples using health data. Sites reviewed generally had adequate navigation capacity and other issues of technology, but lacked analytical capabilities. The trend is to create simple systems with little functionality. There are potential reasons for this. The applications may have been created as an immediate response to user needs. Results from both the NHIS survey and the focus group illustrate that Public Health professionals want openGIS for basic data access, sharing, and visualization, and the possibility that public health professionals are still looking for these basic features. Another explanation lies in technological lag that may have meant sites reviewed were the product of earlier technologies than current efforts.

The second step of this project was the development of the openGIS prototype. The main finding within this step was that openGIS applications could be created and distributed with minimal cost. No need exists to purchase expensive software or hardware and computers as open-source software works as well or, in some instances, better. The application, created to explore respiratory, pollution, and socioeconomic contained more functionality then anything currently available to the public. Commonly, openGIS applications run on commercial software such as Mapinfo, Geomedia or ARCIMS (Gosselin 2000).

The third step of this project was the focus group carried out on November 2nd, 2001, at McMaster University. The key findings at this step relate to general ideas concerning openGIS and the application created for the thesis. Survey results reaffirmed that public health officials have a desire to initiate or further their use of spatial data and GIS. Results also show that public health professionals in southern Ontario are particularly concerned with the cost involved with the new technology. Cost would not occur with new software but with training and development. This may be a consequence of (1) a budget conscience provincial government residing power since 1996, (2) past difficulties in obtaining spatial data, (3) prior knowledge of the high cost of GIS software, or (4) a history of problems with cumbersome GIS software.

Survey results also showed a lack of concern for issues of confidentiality. This is in contrast to indications in the literature. As discussed in Chapter 2 numerous confidentiality concerns plague spatial data. Problems such as the ecologic fallacy and redlining are important considerations when using spatial health data. Respondents unfamiliar with spatial data and GIS may have had minimal exposure to these issues before, and this may have led to an under-reporting of concerns about this issue. If spatial information is going to be used effectively in the public health profession critical issues of confidentiality must be addressed. This is the responsibility of the openGIS application provider. Information concerning issues of security and confidentiality has to be made available to the user.

When prompted to assess the usefulness of potential future functionality the majority of focus group respondents saw promise with the uploading of data and graphing the data. The ability to upload data, join it to an existing table, and display it as a map has been achieved by Jon Park and Gavin Goess as part of their undergraduate thesis project in computer science (Goess and Park 2002). The uploading functionality is currently being tested with the prototype. The potential for graphing capabilities is still in the research stage. An open-source software called 'Open Web Tools Chart Engine'

(OWTCE) has the power to create graphs on-line within an openGIS environment (DM solutions 2002). This software will be implemented in the future as the site evolves.

The fourth step of the design model involved implementing suggestions from the focus group. Here the suggestions from the focus group were evaluated and judged. Suggestions which were possible to implement within the required timeframe altered the initial prototype. It was critical that suggestions were assessed properly to ensure the feasibility of implementation within the timeframe of this thesis.

5.3 Advancement of Knowledge

The thesis findings have advanced knowledge in four key areas: (1) sythesizing the current state-of-the-art in openGIS through literature and website review; (2) visualizing spatial health data over an Internet medium; (3) increasing the effectiveness and functionality of openGIS; (4) gauging the needs of public health officials in relation to spatial data; and (5) amalgamating ideas from different academic disciplines with ideas from trade journals such as 'GeoWorld' and instructional text such as 'Web Publishing Unleashed'.

OpenGIS, if easy to use, may rapidly expand the use of GIS and spatial analysis in Public Health. Expanding the use of GIS in Public Health may increase productivity and decision making. The openGIS prototype represents a major advancement in intra-urban health mapping and analysis. No current openGIS applications on the Internet allow for a multi-level query of health information. With the functionality of this prototype, it runs much like an 'out of the box' GIS program. The development of the fully automated on-line tutorial by Amy Williamson as part of her thesis (Williamson 2002) makes it easier for the new user to learn how to use the program without having to take much time out of their schedule. One of the most important developments of this thesis is the discovery of the potential for developing an openGIS program at little cost. The fact that this program runs on completely free software on older computers cheaper hardware solves some of the cost problems. It has been shown that the high price of clumsy programs such as ARCIMS hinder the ability of a public health department to implement the program.

Visualizing health maps over an Internet medium has not been explicitly documented in any text or journal article. Techniques used for this project were gathered from a wide range of sources. Dent (1993) and Monmonier (1990) argued that choropleth maps are the best design for mapping statistical information. Brewer (1997) illustrated through psychosomatic experiments how certain colour schemes are more effective than others. Mason (1975) and Pickle (1996) offered examples of how health data has been portrayed cartographically in the past. All this information of effectively portraying health information is directed at using a paper medium. Different considerations are needed when using the Internet. Books such as, 'HTML for Dummies,' and 'The Java Reference' explain the importance of using web safe colours. These are colours that remain true when viewed on different monitors. Bringing together all these sources and applying them to an Internet medium is a unique contribution.

The continued assessment of user needs, particularly of public health officials, has further advanced knowledge on directions openGIS technology should take. The McMaster focus group survey not only asked general questions of GIS and openGIS, but it also questioned the prototype and how it should evolve. This survey does more than reiterate the results from the NHIS study. This study gave the opportunity for respondents to explore a current, working openGIS application and then gave an idea of how to improve upon its functionality. This direct input will improve future applications of the technology.

This thesis integrates knowledge from numerous areas, both academic and practical. The literature review references academic articles from geography, public health, statistics, and psychology (colour perception). Numerous references come from trade journals such as GeoWorld. These journals often explain trends in the technology, which are important to consider when making initial decisions about what technologies to use. Many texts were used throughout this thesis to assist in the programming of the prototype. This thesis is not intended to be a computer science project. While the prototype took months to develop into a working application and has been evolving ever since most of the computer work was done through 'trial and error.' With all the software used being open-source, code to improve the prototype was, at times, searched for on the web. The code found is free to use by all and provided a distinct advantage in development. Integrating ideas from numerous disciplines is another accomplishment, and is in line with a geographic tradition of interdisciplinarity.

5.4 Possible Influences of Results

There are possible influences to the results of this thesis. The influences revolve around the survey results from the focus group. Participants of the group were invited and had the opportunity to decline the invitation. Those who attended probably had an interest in GIS technologies. The facilities where the presentations and prototype testing occurred employ state of the art Silicon Graphic PIII 800 MHz machines with 19 inch flat screen monitors that allow the prototype to work faster and give a good, overall impression of the event. Focus group participants also benefited from a presentation showing how the prototype functions and from roving tutors who assisted users. This may explain why 100% of participants felt the prototype was easy to use. Participants also saw separate presentations from Dr. Michael Jerrett and Dr. Yvan Bedard on the subjects of spatial data and GIS. These presentations may have helped participants feel that spatial data and GIS would be useful to their work.

6.2 Contributions to GIS

This work has contributed to the ability to bring GIS to a greater pool of users. By demanding that openGIS technology can be achieved with minimal costs, the opportunity of an increased user pool is evident. The advancements in functionality are encouraging. Ultimately the goal is to have an openGIS program that has complete functionality, similar to that of commercial GIS software. Given current software constraints it is important to interpret the needs of the target audience and create the functionality they need.

6.3 Future research

One main task of any thesis is to set a path for potential future research. The work completed here lends itself to many more questions. Key areas for future research can be grouped into three areas: (1) questions based on current openGIS applications and data; (2) how openGIS is used by public health officials, once implemented; and (3) technological advancement.

6.3.1 Current Applications and Data

We have established that there are few health openGIS sites and even fewer with analytical capabilities are currently running. Throughout the thesis this has been portrayed as a weakness in the current technology. The possibility exists that the simplicity of many openGIS web sites are an implicit response to user needs because users generally want data access, sharing, and visualization. This preference for the simple openGIS system may be a result of the general unfamiliarity with openGIS, with their background training, or the nature of their work (i.e. substance, time constraints). Future studies could target why and how current openGIS applications have been created. Current applications have been created with functionality directed at the needs of a certain audience or without the realization of the full potential of the technology.

Numerous times throughout working on this thesis, the question of data availability and quality emerged. Not only within the NHIS survey and the focus group survey, but in numerous conversations with people that have dealt with obtaining spatial data in Canada. The data and cost issues permeate this survey and the national one, indicating a systematic limitation in the provision and organization of health data in Canada. Recent improvements in the ability to obtain data include the public health portal. A service created by Health Canada which gives public health professionals a common, personalized point of secure access via the Internet for health surveillance information products and services. Even with this initiative, this research indicates a need for major policy reform in data provisions.

6.3.2 Public Reactions to OpenGIS

There is a question of how the public will react to visualized health information. An Internet medium allows for the widespread distribution of maps and data to the public. The ability to create maps classifying a neighbourhood as polluted or showing a school in an area with high asthma prevalence may have a negative impact on certain members of the community. This information could result in angry parents calling school boards and local politicians. It could also lower housing prices in an area, although this is unlikely due to the efficiency of housing markets in pricing externalities (Pearce and Turner 1991). The question lies with the publics 'right to know'. Many groups such as the 'International Right to Know' and the 'Right-To-Know-Network' argue that the public should have information of environmental and social nature available to them (Goldman 1994, Right to Know 2002). This would be an interesting study into the reactions and feelings of people after they see the data. The focus group showed privacy is of limited concern among workers in health fields. To ensure that allowing public access of the openGIS prototype was not in violation of ethics, a request to review the prototype was forwarded to the research ethics officer at McMaster University. The entire ethics board was given a chance to test and review the prototype. The ethics board approved of the prototype (the letter of approval is available in appendix D) and even went so far as to say that to conceal this type of information based on concerns about public reactions would be a mistake:

> A member of the board said that to suppress the research data would be dreadfully unwise. The data map shows that the lowest income levels are generally where the asthma rates are highest. It would be more serious if the ethics board should repress studies that might be controversial. It is important for ethics boards to take a clear stance that this should not be done.

The message that repressing studies simply because they may be controversial is unacceptable corresponds with the right to know movement. Studies showing how the public reacts to the health openGIS site would be beneficial. There are also opportunities to track what people do with the information they receive. If they see their area has high pollution, does this worry them, cause them to become proactive, or does it merely get 'shrugged off?' The answer to this question could form the basis of interesting research that could inform health promotion and environmental health fields. Realizing that this may be a possible study in the future a 'web tracker' has been installed on the web server housing the openGIS prototype. The web tracker keeps a record of how many people visit the site, how web pages are served to the user and what time they visit the site. This data is currently being recorded and will continue to be for the foreseeable future. This data could assist in any future study involving pubic use of openGIS.

6.3.3 Technological Advancements

Most technological advancements will come from ideas not yet formed from future literature reviews and future surveys. Ideas to advance the prototype's functionality are being generated. As mentioned in Chapter 5 *(Discussion)* the ability to upload data has been already been accomplished. Currently though, this ability is limited. New shapefiles can not be uploaded and viewed, but data may be joined to current shapefiles. Joining data to current tables allows the user to query their data against data available to all users. For example, if a public health official knew the rate of lung cancer by census tract and they wanted to query the information in association with the pollution data, they could do so. The next step in data uploading would be allowing the user to upload an entire GIS coverage. This would permit the user to upload, query, and visualize their data without the cost of GIS software.

The ability to perform spatial statistics in an openGIS environment would also be beneficial to users. Currently no openGIS applications enable the user to calculate spatial statistics, such as cluster analysis or regression modeling. Spatial statistics are something both the NHIS and focus group surveys showed to be desirable. This type of functionality would require intense computer programming.

6.4 Conclusions

This thesis project demonstrates the value of interdisciplinary collaboration in geographic inquiry, a long standing tradition in the discipline (Mitchell 1989). The scope of this work included contributions from the Faculty of Humanities Multimedia program (Amy Williamson), the Faculty of Computer Science (Gavin Goess and John Park), and the study of Public Health (CWHPIN). The continuation of ideas towards openGIS and health data will hinge on this continued partnership. Similar to other central results, the aspect of collaboration involves human and institutional challenges rather than technological ones. Thus, while openGIS appears posed to revolutionize geographic information science in general and public health practice in particular, perennial questions involving privacy of data, multi-institution collaboration and human resources will continue to slow this technological revolution.

Bibliography

Adobe. http://www.adobe.com (visited December 12, 2000).

Agabiti, N., S. Mallone, F. Forastiere, G.M. Corbo, S. Ferro, E. Renzoni, P. Sestini, F. Rusconi, G. Ciccone, G. Viegi, E. Chellini and S. Piffer. 1999. "The Impact of Parental Smoking on Asthma and Wheezing. SIDRIA Collaborative Group. Studi Italiani sui Disturbi Respiratori nell'Infanzia e l'Ambiente", Epidemiology. Vol. 10, No. 6, Nov., pp. 692-8.

Alpert, S., and K.E. Haynes. 1994. "Privacy and the Intersection of Geographical Information and Intelligent Transportation Systems", Proceedings of the Conference on Law and Information Policy for Spatial Database, Temple, Arizona. October 28-29, pp. 198-211.

Anderson, H.R., A. Ponce de Leon, J.M. Bland, J.S. Bower, J. Emberlin and Dp. Stracha. 1998. "Air Pollution, Pollens, and Daily Admissions for Asthma in London 1987-1992". Thorax. Vol. 53, pp. 842-848.

Anyamba, A., K.J. Linthicum and C.J. Tucker. 2001. "Climate-disease Connections: Rift Valley Fever in Kenya", Cad Saude Publica. Vol. 17, Suppl. 133-40.

Arden, G., J. Wolf, T. Berninger, C.R. Hogg, R. Tzekov and G.E. Holder. 1999. "S-cone ERGs Elicited by a Simple Technique in Normals and in Tritanopes", Vision Res. Vol. 39, No. 3, Feb., pp. 641-50.

Arizona State University. http://www.bgsu.edu/colleges/library/maps/maps.html (visited December 2000).

Armstrong, M.P., G. Rushton and D.L. Zimmerman. 1999. "Geographically Masking Health Data to Preserve Confidentiality", Statistics in Medicine. Vol. 18, No. 5, pp. 497-525.

Aronoff, S. 1989. Geographical Information Systems: A Management Perspective. Ottawa: WDL Publications.

Artic Mapping system. http://maps.grida.no/arctic/ (visited December 2000).

Austrian Federal Environmental Agency. http://www.ubavie.gv.at/ (visited December 2000).

Bailey, T. and A. Gatrell. 1995. Interactive Spatial Data Analysis. Harlow: Longman.

Bates, D.V. 1983. "Epidemiologic Basis for Photochemical Oxidant Standard", Environ Health Perspect. Vol. 52, Oct, pp. 125-9.

Bédard Y., Gosselin P., Jerrett M., Elliott S., Mowat D., Moore J., Goddard M., Catelan R., Gingras A. and P. Poitras. 2000. "Recent Technological Trends vs Users' Needs in Health Surveillance, a Canadian Study", International Health Information Conference Infocus2000, co-organized by the Canadian Institute for Health Information and by Canada's Health Informatics Association. Vancouver, Canada, June 24-27.

Blake, B.J. and L. Bentov. 2001. "Geographical Mapping of Unmarried Teen Births and Selected Sociodemographic Variables", Public Health Nurs. Vol. 18, No. 1, Jan-Feb, pp. 33-9.

Brettel, H., F.Vienot and J.D. Mollon. 1997. "Computerized Simulation of Color Appearance for Dichromats", J Opt Soc Am A. Vol. 14, No. 10, Oct., pp. 2647-55.

Brewer C., Colour Pairs. http://www.essc.psu.edu/~cbrewer/ColorPairs (visited Feb. 14, 2001).

Brewer, C.A., A.M. MacEachren, L.W. Pickle, and D. J. Herrmann. 1997. "Mapping Mortality: Evaluating Color Schemes for Choropleth Maps", Annals of the Association of American Geographers. Vol. 87, No. 3, pp. 411-438.

Burnett, R., R. Ma, M. Jerrett, M.S. Goldberg, S. Cakmak, C.A. 3rd Pope and D. Krewski. 2001. "The Spatial Association Between Community Air Pollution and Mortality: A New Method of Analyzing Correlated Geographic Cohort Data", Environ Health Perspect. Vol.109, Suppl. 3 Jun., pp. 375-80.

Burney, P.G., J.R. Britton, S. Chinn, A.E. Tattersfield, A.O. Papacosta, M.C. Kelson, et al. 1987. "Descriptive Epidemiology of Bronchial Reactivity in an Adult Population: Results from a Community Study", Thorax. Vol. 42, pp. 38-44.

Burney, P.G. 1994. "The European Community Respiratory Health Survey", Eur Respir J Vol. 7, pp. 954-960.

Burstrom, K., M. Johannesson and F. Diderichsen. 2001. "Health-related Quality of Life by Disease and Socio-economic Group in the General Population in Sweden", Health Policy. Vol. 55, No. 1, Jan. pp 51-69.

California Resource Evaluation System. http://ceres.ca.gov/geo_area/index.html (visited December 2000).

Cassino, C., K. Ito, I. Bader, C. Ciotoli, G. Thurston and J. Reibman. 1999. "Cigarette Smoking and Ozone-associated Emergency Department Use for Asthma by Adults in New York City", Am J Respir Crit Care Med. Vol. 159, No. 6, Jun., pp. 1773-9.

Catelan, R., M. Jerrett, Y. Bedard, S. Elliott and P. Gosselin. 2000. "A Systematic Literature Review and User Needs Assessment of GIS,OLAP and the WWW Applications for Environmental and Public Health Surveillance", International Health Geographics Conferences, National 4-H Council. Chevy Chase, Maryland, USA. March 17-19.

Census Mapping ESRI. http://maps.esri.com/ESRI/mapobjects/tmap/tmap.htm (visited December 2001).

Center for the Environment. http://icemaps.des.ucdavis.edu/icemaps2/ICEMapInit.html (visited December 2000).

Chew, F., D.Y.T. Goh, B.C. Ooi, R. Sahorom, J.K.S. Hui and B.W. Lee. 1999. "Association of Ambient Air Pollution Levels with Acute Asthma Exacerbation among children in Singapore", Allergy. Vol. 54, pp. 320-329.

Children's Report Card. http://www.mnplan.state.mn.us/cgi-bin/datanetweb/chi_dpm/ (visited December 2000).

Chi-Wen Li, K. 1996 "The Private Insurance Industry's Tactics Against Suspected Homosexuals: Redlining Based on Occupation, Residence and Marital Status", American Journal of Law & Medicine. Vol. 22, No. 4, pp. 477-502.

CIET. http://www.ciet.org/welcome.html (visited December 2000).

Clarke, R. 1994. "The Digital Persona and its Application to Data Surveillance", The Information Society. Vol. 10, pp. 77-94.

Cliff, A. and P. Haggett. 1988. Atlas of Disease Distributions: Analytic Approaches to Disease Data. Oxford: Blackwell.

CDPHE (Colorado Department of Public Health). 2001. Weekly newsletter obtained from http://www.cdphe.state.co.us/cdphehom.asp.

Committee on Research Priorities for Airborne Particulate Matter. 1998. Research Priorities for Airborne Particulate Matter: I. Immediate Priorities and a Long Range Research Portfolio. Washington, D.C.: National Academy Press.

Cowell, J.M and M.E. Cowell. 1999. "Immunization Rates: A Community Assessment Approach for Diagnosis", J Sch Nurs. Vol. 15, No. 5, Dec. pp. 40-3.

Cox, L. 1996. "Protecting Confidentiality in Small Population health and Environmental Statistics", Statistics in Medicine. Vol. 15, pp. 1895-1905.

Croner C., J. Sperling, and F. Broome. 1996. "Geographic Information Systems (GIS): New Prespectives in Understanding Human Health and Environmental Relationships", Statistics in Medicine. Vol. 16, pp. 1961-1977.

Croner, C. 2000. "Public health News and Information", USA Center for Disease Control. No. 35, July.

Culpepper, R. "Weave Maps Across the Web", GIS World . Vol. 11, No. 11, pp. 46-53.

Curtis, S.E. and A.R. Taket. 1996. Health and societies: changing perspectives. London: Arnold.

Department of Environmental, Food &Rural Affairs. http://maffgis.maff.gov.uk/default.htm (visited December 2000).

Demers, M. 1996. Fundementals of Geographic Information Systems. New York: John Wiley & Sons.

DeMers, M.N. 1997. Fundamentals of Geographic Information Systems. New York: John Wiley & Sons.

Dent, B.D. 1996 Cartography: Thematic Map Design, 4th Edition. New York: William C. Brown.

Dertouzos, M. and B. Gates. 1999. "Titans Talk Tech", Technology Review. Vol. 4, pp. 23-37.

DM Solutions. http://www.dmsolutions.ca (visited Feb.2002).

Ebelt, S., A. Petkau, S. Vedal, T. Fischer and M. Brauer. 2000. "Exposure of Chronic Obstructive Pulmonary Disease Patients to Particulate Matter: Relationships between Personal and Ambient Air Concentrations". Journal of the Air and Waste Management Association. Vol. 50, pp. 1081-1094.

Ecotourism. http://www.gisca.adelaide.edu.au/cgi-bin/eco/ecogis (visited December 2000).

Environment Canada – Species at Risk. http://www.hc-sc.gc.ca/pphb-dgspsp/dsol-smed/ (visited January 2002).

Environmental Protection Agency. http://www.epa.gov/ceisweb1/ceishome/atlas/learngeog/interactivewebmapping.html (visited December 2000).

Foresman, T. 1999. "Spatial Analysis and Mapping on the Internet", Journal of Public Health Management and Practice. Vol. 5, No. 4.

Fotheringham A., and D. Wong. 1991. "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis", Environment and Planning A. Vol. 23, pp 1025-1044.

Friends of the Earth. http://www.foe.co.uk/campaigns/industry_and_pollution/factorywatch/ 14.Information (visited December 2000).

Gatrell, A.C. and M. Loytonen. 1998. "GIS and Health Research: an Introduction" in GIS and Health, Eds. A.C. Gatrell and M. Loytonen. Taylor and Francis, pp. 3-16.

Goess, G. and J. Park. 2002. Improving Public Health's Ability to Access, Display and Manipulate On-Liine Maps. Computer Science Undergrad Thesis.

Goldman, B. 1994. Not Just Prosperity: Achieving Sustainability with Environmental Justice. National Wildlife Federation. February 1994.

Gosselin, P., Y. Bédard, M. Jerrett, S. Elliott, R. Catelan, P. Poitras and A. Gingras. 2000. "GIS and OLAP in Health Surveillance: Needs Analysis for Successful Integration. Final Report for National Health Surveillance Infostructure, Health Protection Branch, Health Canada", Centre de recherche du Centre Hospitalier de l'Universitaire. Laval, Québec, 10 février, 70 p.

Gosselin, P., Y. Bedard, M. Gerrett, S. Elliot, R. Catelan, P. Poitras, and A.Gingras. 2001. "GIS and OLAP in Health Surveillance: Users = Needs Analysis", A Report for the Nation Health Inforstructure. Health Canada.

Greiner, M. 2001. Is 'enyc' a dirty word? http://www.aardvark.co.nz/daily/2002/0327.shtml (visited March 1 2001).

Grossman-Garber, D.A., D. Bengtson, A. Gold, B. Lord. 1999. "Strengthening Undergraduate Teaching and Learning through Partnerships: Recasting the University of Rhode Island's Academic Culture." American Association of Higher Education Faculty Roles and Rewards Conference.

Hardie, A. 1998. The development and Present State of Web-GIS", Cartography. Vol. 27, pp. 11-26.

Health Canada. 1998. The Health and Environment Handbook for Health Professionals. Ottawa: Minister of Public Works and Government Services Canada.

Health Canada Disease Surveillance On-line. http://www.hc-sc.gc.ca/pphb-dgspsp/dsol-med/ (visited January 2002).

Health Canada FLU Watch. http://www.hc-sc.gc.ca/pphb-dgspsp/fluwatch/index.html (visited January 2002).

Humboldt Bay. http://humboldtbay.cnrs.humboldt.edu/index.html (visited December 2000).

Intergraph. http://www.Intergraph.com (visited April 2002).

Ito, K., G.D. Thurston, C. Hayes, and M. Lippmann. 1993. "Association for London, England Daily Mortality with Particulate Matter, Sulfur Dioxide, and Acidic Aerosol Pollution", Arch. Environ. Health. Vol. 48, pp. 213-220.

Jadad, A.R. and A. Gagliardi. 1998. "Rating Health Information on the Internet: Navigating to Knowledge or to Babel?", JAMA. Vol. 279, No. 8, Feb., pp. 611-4.

Jerrett, M., J. Eyles, D. Cole and S. Reader. 1997. "Environmental Equity in Canada: An Empirical Investigation Into the Income Distribution of Pollution in Ontario", Environment and Planning A. Vol. 29, pp. 1777-1800.

Jerrett, M. 2001. "A GIS -- Environmental Justice Analysis of Particulate Air Pollution in Hamilton Canada", Environment and Planning A. Vol. 33, pp. 955-973.

Keates, J. S. 1993. Cartographic Design and Production. London: Longman Group, Ltd.

King, P. 1979. "Problems of Spatial Analysis in Geographic Epidemiology", Social Science and Medicine. Vol. 13D, pp. 249-252.

Kraak, M.J. and A. Brown, Eds. 2000. Web Cartography - developments and prospects. London: Taylor & Francis.

Lawerence, S.and C.L. Giles. 1999. "Accessibility of Information on the Web", Nature. Vol. 400, No. 21.

Lawson, A.B. 2001. "Disease Map Reconstruction", Stat Med. Vol. 20, No. 14, Jul. 30, pp. 2183-204.

Levy, J., J. Spengler and Y. Yanagisawa. 1998. "Impact of Residential Nitrogen Dioxide Exposure on Personal Exposure: An International Study". Journal of the Air and Waste Management Association. Vol. 48, pp. 553-560.

Levy, S. 1999. "Code Warriors", Newsweek. Vol. 3, pp. 59-63.

Lippmann, M. 2000. "Ozone". in Morton Lippmann, ed. Environmental Toxicants. NewYork: Wiley-Interscience.

Littlejohns, P. and L.D. Macdonald. 1993. "The Relationship Between Severe Asthma and Social Class", Respir Med. Vol. 87, No. 2, Feb., pp.139-43.

Litonjua, A.A., V.J. Carey, S.T. Weiss and D.R. Gold. 1999. "Race, Socioeconomic Factors, and Area of Residence are Associated with Asthma Prevalence", Pediatr Pulmonol. Vol. 28, No. 6, Dec., pp 394-401.

Lloyd, W.J. 2001. "Integrating GIS into the Undergraduate Learning Environment", Journal of Geography. Vol. 100, No. 5, pp.158-163.

Lochner, K., E. Pamuk, D. Makuc, B.P. Kennedy and I. Kawachi. 2001. "State-level Income Inequality and Individual Mortality Risk: A Prospective, Multilevel Study", Am J Public Health. Vol. 91, No. 3, Mar., pp. 385-91.

Lookoff. http://www.lookoff.com (visited March 2, 2002).

MacEachren, A.M., C.A. Brewer and L. Pickle. 1998. "Visualizing Georeferenced Data: Representing Reliability of Health Statistics", Environment and Planning: A. Vol. 30, pp. 1547-1561.

Macromedia. http://www.macromedia.com/ (visited December 12, 2000).

Madej, E. 2000. Cartographic Design. Canada: OnWord Press

Manfreda, J., M.R. Becklake, M.R. Sears, M. Chan-Yeung, H. Dimich-Ward, H.C. Siersted, P. Ernst, L. Sweet, L. Van Til, D.M Bowie, N.R. Anthonisen, and R.B. Tate. 2001. "Prevalence of Asthma Symptoms Among Adults Aged 20-44 Years in Canada". CMAJ. Vol. 164, No. 7, April 3, pp. 995-1001.

Mapserver. http://mapserver.gis.umn.edu/ (visited April 5. 2001).

Mayer, J. 1983. "The Role of Spatial Analysis and Geographic Data in the Detection of Disease Causation", Social Science and Medicine. Vol. 16, pp. 1213-1221.

McFedries, P.1999. The Complete Idots Guide to Creating a Web Page. United States: MacMillian.

McKee, L. 2001. "Geography Connects Cyberspace with the Real World", GeoWorld. Vol. 14 Febuary.

Melnick, D. and B. Rouse. 2001. Portrait of Health in the United States, Major Statistical Trends: 1st Ed. New York: Bernan Press.

Microsoft Terrasever. http://terraserver.homeadvisor.msn.com/default.asp (visited December 2000).

Miller, J.R. 2001. "The control of Mosquito-borne Diseases in New York City", J Urban Health. Vol. 78, No. 2, June, pp. 359-66.

Monmonier, M.S. 1980. "The Hopeless Pursuit of Purification in Cartographic Communication, A Comparison of Graphic Arts and Perceptual Distortions of Graytone Symbols", Cartographica. Vol. 17, pp. 24-39.

Mitchell, B. 1989. Geigraphy and Resource Analysis. Ney Yowk, NY. John Wiley and Sons.

Moore, D.A. and T.E. Carpenter. 1999. "Spatial Analytical Methods and Geographic Information Systems: Use in Health Research and Epidemiology", Epidemiol Rev. Vol. 21, No. 2, pp. 143-61.

Nakaya T. 2000. "An Information Statistical Approach to the Modifiable Areal Unit Problem", Environment and Planning A. Vol. 32, pp. 91-109.

National Geographic. http://www.nationalgeographic.com/maps/ (visited December 2000).

NBCi. http://home.nbci.com/search/map (visited December 2000).

Nevada Ecosystem Project. http://www.crseo.ucsb.edu/biogeog/snep.html (visited December 2000).

Nursing Home info. http://www.nursinghomeinfo.com/search.html (visited December 2000).

Ontario Ministry of the Environment (OMOE). 1999. A Compendium of Current Knowledge on Fine Particulate Matter in Ontario. Prepared for The Ontario Smog Plan Steering Committee (as released to the CRESTech/NERAM Expert Panel for review).

Persky, V.W., J. Slezak, A. Contreras, L. Becker, E. Hernandez, V. Ramakrishnan and J. Piorkowski. 1998. "Relationships of Race and Socioeconomic Status with Prevalence, Severity and Symptoms of Asthma in Chicago School Children", Ann Allergy Asthma Immunol. Vol. 81, No. 3, Sep., pp. 266-71.

Peterson, M.P. 1997. "Trends in Internet MapUse", ICC 97 Proceedings. Vol. 3, pp. 1635 – 1642.

Phillips, R.L. Jr., E.L. Kinman, P.G. Schnitzer, E.J. Lindbloom and B. Ewigman. 2000. "Using Geographic Information Systems to Understand Health Care Access", Arch Fam Med. Vol. 9, No. 10, Nov-Dec., pp. 971-8.

PHP. www.php.com. (visited January 3, 2002).

Pickle, L.W., et al. 1996. Atlas of United States Mortality. Hyattsville, Md: National Center for Health Statistics, Centers for Disease Control and Prevention.

Pikhart, H., M. Bobak, P. Gorynski, B. Wojtyniak, J. Danova, M.A. Celko, B. Kriz, D. Briggs and P. Elliott. 2001. " Outdoor Sulphur Dioxide and Respiratory Symptoms in Czech and Polish School Children: A Small-Area Study (SAVIAH). Small-Area Varioation in Air Pollution and Health", Int Arch Occup Environ Health. Vol. 74, No. 8, Oct, pp. 574-8.

Pikhart, H., V. Prikazsky, M. Bobak, B. Kriz, M. Celko, J. Danova, K. Pyrl and J. Pretel. 1997. "Association Between Ambient Air Concentrations of Nitrogen Dioxide and Respiratory Symptoms in Children in Prague, Czech Republic. Preliminatry Results from the Czech Part of the SAVIAH Study. Small Area Variation in Air Pollution and Health", Cent Eur J Public Health. Vol. 5, No. 2, Jun., pp. 82-5.

Pearce, D. and Turner, K.1991. Economics of Natural Resources and the Environment. Baltimore: The John Hopkins University Press

Plewe, B. 1997. GIS Online: Information, Retrieval, Mapping and the Internet. Santa Fe: OnWord Press.

Plewe, B. "So You Want to Build an Online GIS?", GIS World. Vol. 10. No. 11, pp. 58-62

Ponka, A. 1991. "Asthma and Low Level Air Pollution in Helsinki", Archives of Environmental Health. Vol. 46, pp. 262-269.

Right To Know. http://www.rtk.net/ (visited March 3 2002)

Royce, W.W. 1970. Managing the Development of Large Software Systems. San Francisco: WESTCON.

Schwartz, J., D. Slater, T. Larson, W. Pierson and J. Koenig. 1993. "Particulate Air Pollution and Hospital Emergency Room Visits for asthma in Seattle", Am Rev Respir Dis. Vol. 147, pp. 826-831.

Selvin, S., D.W. Merrill, C. Erdmann, M. White and K. Ragland. 1998. "Breast Cancer Detection: Maps of 2 San Francisco Bay Area Counties." Am J Public Health. Vol. 88, No. 8, Aug., pp. 1186-92.

Shibuya, K., H. Hashimoto and E. Yano. 2002. "Individual Income, Income Distribution and Self Rated Health in Japan: Cross Sectional Analysis of Nationally Representative Sample", BMJ. Vol. 324, No. 7328, Jan., pp. 16-9.

Sieber, A and L. Wiggins. 1995. "Tour the World Wide Web: A Look at Three GIS Sites", GIS World. Vol. 8, No. 6, pp. 70-75.

Silberg,W.M, G.D. Lundberg and R.A. Musacchio. 1997. "Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware [editorial]", JAMA. Vol. 277, No. 15, pp. 1244-1245.

Sommerville, I. 1992. Software Engineering. England: Adison Wesley Publishing Company.

Statistics Canada. http://www.statcan.ca/ (visited November 15th, 2000).

Statistics Canada. http://ceps.statcan.ca/english/profil/PlaceSearchForm1.cfm (visited December 2000).

Stevenson, J. and W. Cartwright. 2000. "A Toolbox for Publishing Maps on the World Wide Web", Cartography. Vol. 29, No.2, December.

Stieb, D.M., R.T. Burnett, R.C. Beveridge and J.R. Brook. 1996. "Association Between Ozone and Asthma Emergency Department Visits in Saint John, New Brunswick, Canada", Environ Health Perspect. Vol. 104, No. 12, Dec, pp. 1354-60.

Stynes, K., Publishing Cartography on the Web. http://www.geog.le.ac.uk/argus/ICAK.Stynes (visited Mar. 1, 2001).

Sunyer, J. 1997. "Outdoor Air II: Nitrogen Dioxide". in Steenland, Kyle and David Savitz, eds. Topics in Environmental Epidemiology. Oxford: Oxford University Press.

The American Cancer Atlas. http://cancer.gov/atlasplus/maps.html (visited December 2000).

The City of San Diego. http://cart.sandag.cog.ca.us/demog/default.htm (visited December 2000).

The University of Leeds. http://www.ccg.leeds.ac.uk/mce/ (visited December 2000).

The World Health Organization. http://www.who.int/whosis/ (visited December 2000).

United Nations Environment Programme. http://www.unep-wcmc.org/index.html?http://www.wcmc.org.uk/CITES/redirect.htm~main (visited December 2000).

United States Census. http://tiger.census.gov/cgi-bin/mapbrowse-tbl/ (visited December 2000).

U. S. Department of Health and Human Services (USDHHS). 1998. Toxicological Profile for Sulfur Dioxide. Atlanta: Agency for Toxic Substances and Disease Registry.

USA Center for Disease Control. http://maps.esri.com/scripts/esrimap.dll?name=mmwr&cmd=start (visited December 2001).

Venn, A.J., S.A. Lewis, M. Cooper, R. Hubbard and J. Britton. 2001. "Living Near a Main Road and the Risk of Wheezing Illness in Children", Am J Respir Crit Care Med. Vol. 164, No. 12, Dec., pp. 2177-80.

Vischeck. http://ww.vischeck.com (visited Jan 2002).

Weiland, S.K., K.A. Mundt, A. Ruckmann and U. Keil. 1994. "Self-reported Wheezing and Allergic Rhinitis in Children and Traffic Density on Street of Residence", Ann Epidemiol. Vol. 4, No. 3, May, pp. 243-7.

Williamson, A. 2002. An Interactive Flash Tutorial for the Healthgis Interactive Mapping Website. Multimedia Undergrad Thesis.

World Health Organization (WHO). 1997. "Assessment of Exposure to Indoor Air Pollutants", World Health Organization Regional office for Europe. Copenhagen: WHO Regional Publications: European Series, No. 78.

Yasoff, W. and E. Sondik. 1999. "Geographic Information Systems (GIS) in Public Health Practice in the New Millennium", Journal of Public Health Managements and Practice. Vol. 5, No. 4, July.

APPENDIX A: Computer Code for the MapServer Java Applett

```
map
  extent -79.954891 43.138133 -79.725663 43.367328
  fontset "fonts/fontlist.txt"
  imagecolor 255 255 255
  name "imaps"
  shapepath "data"
  size 500 500
  status on
  symbolset "symbols/all.sym"
  units dd

  # 1 scalebar #none
  scalebar
    STATUS EMBED
    LABEL
      SIZE small
      COLOR 0 0 0
      OUTLINECOLOR 255 255 255
    END
    STYLE 0
    INTERVALS 5
    SIZE 150 5
    COLOR 0 0 0
    BACKGROUNDCOLOR 255 255 255
    OUTLINECOLOR 0 0 0
    UNITS KILOMETERS
    POSITION LR
    TRANSPARENT TRUE
    postlabelcache true
  END # Scalebar
  # 2 layer #none

  layer
    classitem "SUMBY3"
    data "thisisit"
    group "background"
    labelcache off
    postlabelcache true
    status on
    type polygon
    # 1 class #none
    class
      backgroundcolor 128 128 128
      color 255 200 204
      outlinecolor 185 139 58
      name "Census Tracts"
    end
  end

  layer
    classitem "DWELLING"
    data "thisisit"
```

```
    group "dwellingnatbr"
    labelcache off
    postlabelcache true
    status on
    type polygon
    # 1 class #none
    class
      backgroundcolor 128 128 128
      color 57 0 115
      outlinecolor 0 0 0
      expression ([DWELLING] <= 104872)
Name "Less than $104,872"
 end
    class
      backgroundcolor 128 128 128
      color 107 53 161
      outlinecolor 0 0 0
        expression (([DWELLING] > 104872) && ([DWELLING] <= 124355))
Name "$104,872 to $124,355"
   end
    class
      backgroundcolor 128 128 128
      color 255 255 255
      outlinecolor 0 0 0
      expression (([DWELLING] > 124355) && ([DWELLING] <= 140237))
Name "$124,355 to $140,237"
end
    class
      backgroundcolor 128 128 128
      color 144 199 133
      outlinecolor 0 0 0
      expression (([DWELLING] > 140237) && ([DWELLING] <= 160074))
Name "$140237 to $160,074"
end
    class
      backgroundcolor 128 128 128
      color 15 90 0
      outlinecolor 0 0 0
      expression ([DWELLING] > 160074)
Name "Greater than $160,074"
end
  end

  layer
    classitem "SUMBY3"
    data "thisisit"
    group "asthmanatbr"
    labelcache off
    postlabelcache true
    status on
    type polygon
    # 1 class #none
    class
      backgroundcolor 128 128 128
```

```
      color 15 90 0
      outlinecolor 0 0 0
      expression ([SUMBY3] <= 110.613)
Name "Less than 110.613"
    end
    class
      backgroundcolor 128 128 128
      color 144 199 133
      outlinecolor 0 0 0
      expression (([SUMBY3] > 110.613) && ([SUMBY3] <= 115.942))
Name "110.613 to 115.942"
    end
    class
      backgroundcolor 128 128 128
      color 255 255 255
      outlinecolor 0 0 0
      expression (([SUMBY3] > 115.942) && ([SUMBY3] <= 121.372))
Name "115.942 to 121.372"
    end
    class
      backgroundcolor 128 128 128
      color 107 53 161
      outlinecolor 0 0 0
      expression (([SUMBY3] > 121.372) && ([SUMBY3] <= 126.263))
Name "121.372 to 126.263"
    end
    class
      backgroundcolor 128 128 128
      color 57 0 115
      outlinecolor 0 0 0
      expression ([SUMBY3] > 126.263)
name "Greater that 126.263"
    end
  end

  layer
    classitem "Pm10"
    data "thisisit"
    group "Pm10quintile"
    labelcache on
    postlabelcache true
    status on
    type polygon
    # 1 class #none
    class
      backgroundcolor 128 128 128
      color 15 90 0
      outlinecolor 0 0 0
      expression ([PM10] <= 26.57)
Name "Less than 26.57"
    end
    class
      backgroundcolor 128 128 128
      color 144 199 133
```

```
      outlinecolor 0 0 0
      expression (([PM10] > 26.57) && ([PM10] <= 27.39))
Name "26.57 to 27.39"
end
    class
      backgroundcolor 128 128 128
      color 255 255 255
      outlinecolor 0 0 0
      expression (([PM10] > 27.39) && ([PM10] <= 28.14))
Name "27.39 to 28.14"
    end
    class
      backgroundcolor 128 128 128
      color 107 53 161
      outlinecolor 0 0 0
      expression (([PM10] > 28.14) && ([PM10] <= 28.71))
Name "28.14 to 28.71"
    end
    class
      backgroundcolor 128 128 128
      color 57 0 115
      outlinecolor 0 0 0
      expression ([PM10] > 28.71)
Name "Greater than 28.71"
end
  end

  layer
    classitem "SO2"
    data "thisisit"
    group "SO2quintile"
    labelcache on
    postlabelcache true
    status on
    type polygon
    # 1 class #none
    class
      backgroundcolor 128 128 128
      color 15 90 0
      outlinecolor 0 0 0
      expression ([SO2] <= 5.87)
Name "Less than 5.87"
    end
    class
      backgroundcolor 128 128 128
      color 144 199 133
      outlinecolor 0 0 0
      expression (([SO2] > 5.87) && ([SO2] <= 6.01))
Name "5.87 to 6.01"
    end
    class
      backgroundcolor 128 128 128
      color 255 255 255
      outlinecolor 0 0 0
```

```
      expression (([SO2] > 6.01) && ([SO2] <= 6.16))
Name "6.01 to 6.16"
    end
    class
      backgroundcolor 128 128 128
      color 107 53 161
      outlinecolor 0 0 0
      expression (([SO2] > 6.16) && ([SO2] <= 6.5))
name "6.16 to 6.5"
    end
    class
      backgroundcolor 128 128 128
      color 57 0 115
      outlinecolor 0 0 0
      expression ([SO2] > 6.5)
Name "Greater that 6.5"
   end
  end

  layer
    classitem "NOX"
    data "thisisit"
    group "NOXquintile"
    labelcache on
    postlabelcache true
    status on
    type polygon
    # 1 class #none
    class
      backgroundcolor 128 128 128
      color 15 90 0
      outlinecolor 0 0 0
      expression ([NOX] <= 32.29)
name "Less that 32.29"
    end
    class
      backgroundcolor 128 128 128
      color 144 199 133
      outlinecolor 0 0 0
      expression (([NOX] > 32.29) && ([NOX] <= 36.06))
Name "32.29 to 36.06"
end
    class
      backgroundcolor 128 128 128
      color 255 255 255
      outlinecolor 0 0 0
      expression (([NOX] > 36.06) && ([NOX] <= 37.68))
Name "36.06 to 37.68"
end
    class
      backgroundcolor 128 128 128
      color 107 53 161
      outlinecolor 0 0 0
      expression (([NOX] > 37.68) && ([NOX] <= 39.37))
```

```
Name "37.68 to 39.37"
end
    class
      backgroundcolor 128 128 128
      color 57 0 115
      outlinecolor 0 0 0
      expression ([NOX] > 39.37)
Name "Greater than 39.37"
  end
  end

  layer
    classitem "POP96"
    data "thisisit"
    group "POP96quintile"
    labelcache on
    postlabelcache true
    status on
    type polygon
    # 1 class #none
    class
      backgroundcolor 128 128 128
      color 15 90 0
      outlinecolor 0 0 0
      expression ([POP96] <= 780)
Name "Less than 780 people"
    end
    class
      backgroundcolor 128 128 128
      color 144 199 133
      outlinecolor 0 0 0
      expression (([POP96] > 780) && ([POP96] <= 1160))
Name "780 to 1160"
    end
    class
      backgroundcolor 128 128 128
      color 255 255 255
      outlinecolor 0 0 0
      expression (([POP96] > 1160) && ([POP96] <= 1390))
Name "1160 to 1390"
    end
    class
      backgroundcolor 128 128 128
      color 107 53 161
      outlinecolor 0 0 0
      expression (([POP96] > 1390) && ([POP96] <= 1790))
Name "1390 to 1790"
    end
    class
      backgroundcolor 128 128 128
      color 57 0 115
      outlinecolor 0 0 0
      expression ([POP96] > 1790)
Name "More than 1790 people"
```

```
      end
    end

    layer
      classitem "GINI"
      data "thisisit"
      group "GINIquintile"
      labelcache on
      postlabelcache true
      status on
      type polygon
      # 1 class #none
      class
        backgroundcolor 128 128 128
        color 15 90 0
        outlinecolor 0 0 0
        expression ([GINI] <= 0.316)
  Name "Less than 0.316"
      end
      class
        backgroundcolor 128 128 128
        color 144 199 133
        outlinecolor 0 0 0
        expression (([GINI] > 0.316) && ([GINI] <= 0.348))
  Name "0.316 to 0.348"
  end
      class
        backgroundcolor 128 128 128
        color 255 255 255
        outlinecolor 0 0 0
        expression (([GINI] > 0.348) && ([GINI] <= 0.371))
  Name "0.348 to 0.371"
      end
      class
        backgroundcolor 128 128 128
        color 107 53 161
        outlinecolor 0 0 0
        expression (([GINI] > 0.371) && ([GINI] <= 0.393))
  Name "0.371 to 0.393"
  end
      class
        backgroundcolor 128 128 128
        color 57 0 115
        outlinecolor 0 0 0
        expression ([GINI] > 0.393)
  Name "Greater than 0.393"
      end
    end

    layer
      classitem "SMOKE"
      data "thisisit"
      group "SMOKEquintile"
      labelcache on
```

```
    postlabelcache true
    status on
    type polygon
    # 1 class #none
    class
      backgroundcolor 128 128 128
      color 15 90 0
      outlinecolor 0 0 0
      expression ([SMOKE] <= 310)
Name "Less than 310"
    end
    class
      backgroundcolor 128 128 128
      color 144 199 133
      outlinecolor 0 0 0
      expression (([SMOKE] > 310) && ([SMOKE] <= 350))
Name "310 to 350"
end
    class
      backgroundcolor 128 128 128
      color 255 255 255
      outlinecolor 0 0 0
      expression (([SMOKE] > 350) && ([SMOKE] <= 390))
Name "350 to 390"
    end
    class
      backgroundcolor 128 128 128
      color 107 53 161
      outlinecolor 0 0 0
      expression (([SMOKE] > 390) && ([SMOKE] <= 410))
Name "390 to 410"
    end
    class
      backgroundcolor 128 128 128
      color 57 0 115
      outlinecolor 0 0 0
      expression ([SMOKE] > 410)
Name "greater than 410"
    end
  end

  # 3 layer thisisit
  layer
    classitem "SUMBY3"
    data "thisisit"
    footer "thisisit_footer.html"
    group "Asthma Rates"
    header "thisisit_header.html"
    labelcache off
    name "thisisit"
    postlabelcache true
    queryitem "SUMBY3"
    status on
    type polygon
```

```
  # 1 query #none
  query
    template "thisisit_query.html"
  end
  # 2 class {Census Tracts Matching Query}
  class
    backgroundcolor -1 -1 -1
    color 255 0 0
    name "Census Tracts Matching Query"
    outlinecolor 255 150 46
    symbol "hatch_orange"
  end
end

# 4 layer postcode
layer
  classitem "SUMBY3"
  data "thisisit"
  group "background"
  labelcache off
  name "postcode"
  postlabelcache true
  status on
  type polygon
  # 1 class {Postal Code Entry}
  class
    backgroundcolor -1 -1 -1
    color 255 0 0
    expression "([AREANAME] = 5370048)"
    name "Postal Code Entry"
    outlinecolor 255 0 0
    symbol "hatch_red"
  end
end

# 5 layer peripheral_ham
layer
  classitem "AREA"
  data "peripheral_ham"
  group "peripheral_ham"
  labelcache off
  name "peripheral_ham"
  postlabelcache true
  queryitem "AREA"
  status on
  type polygon
  # 1 class {Data Unreliable}
  class
    backgroundcolor 128 128 128
    color 245 245 245
    name "Data Unreliable"
    outlinecolor 0 0 0
    symbol "hatch"
  end
```

```
end

# 5 layer peripheral_ham
layer
  classitem "AREA"
  data "industrial"
  group "Industrial Area"
  labelcache off
  name "industrial"
  postlabelcache true
  queryitem "AREA"
  status on
  type polygon
  # 1 class
  class
    backgroundcolor 128 128 128
    color 245 245 245
 ###   name "Industrial Area"
    outlinecolor 0 0 0
    symbol "hatch_black"
  end
end

# 6 layer harbour_latlong2
layer
  classitem "AREA"
  data "harbour_latlong2"
  group "Hamilton Harbour"
  labelcache off
  name "harbour_latlong2"
  postlabelcache true
  queryitem "AREA"
  status on
  type polygon
  # 1 class #none
  class
    backgroundcolor 128 128 128
    color 0 0 160
    outlinecolor 0 0 0
  end
end
# 7 layer lake_on_latlong
layer
  classitem "AREA"
  data "lake_on_latlong"
  group "lake_on_latlong"
  labelcache off
  name "lake_on_latlong"
  postlabelcache true
  queryitem "AREA"
  status on
  type polygon
  # 1 class {Lake Ontario}
  class
```

```
     backgroundcolor 128 128 128
     color 0 0 160
     name "Water"
     outlinecolor 0 0 0
   end
 end
 # 8 layer greenspace
 layer
   classitem "AREA"
   data "greenspacegeo"
   group "greenspacegeo"
   labelcache off
   name "greenspace"
   postlabelcache true
   queryitem "FNODE_"
   status on
   type polygon
   # 1 class Greenspace
   class
     backgroundcolor 128 128 128
     color 0 51 0
###     name "Greenspace"
     outlinecolor 0 0 0
     symbol "hatch_green"
   end
 end
 # 9 layer escarp_latlong
layer
classitem "FNODE_"
data "escarp_latlong"
group "escarp_latlong"
labelcache off
name "escarp_latlong"
postlabelcache true
queryitem "FNODE_"
status on
type line
   # 1 class {Niagara Escapment}
class
     backgroundcolor 128 128 128
     color 0 0 0
 ###    name "Niagara Escapment"
     outlinecolor 0 0 0
   end


 end
 # 10 layer arterials
 layer
   classitem "FNODE_"
   data "arterialgeo"
   group "arterialgeo"
   labelcache off
   labelitem "NAME"
```

```
  labelmaxscale 41300
  name "arterials"
  postlabelcache true
  queryitem "FNODE_"
  status on
  type line
  # 1 class Arterials
  class
    backgroundcolor 128 128 128
    color 0 0 0
###    name "Arterials"
    outlinecolor 0 0 0
    # 1 label arial
    label
      angle auto
      antialias
      color 0 0 0
      font "arial"
      force false
      maxsize 7
      mindistance 15
      minfeaturesize auto
      minsize 7
      offset 2 2
      outlinecolor 255 255 230
      position uc
      size 7
      type truetype
    end
  end
end

# 11 layer hospitals
layer
  classitem "HOSPITAL_N"
  data "hospitals_geo"
  group "hospitals"
  labelcache off
  labelitem "HOSPITAL_N"
  labelmaxscale 41300
  maxscale 41300
  name "hostpitals"
  postlabelcache true
  status on
  type polygon
  # 1 class Arterials
  class
    backgroundcolor 128 128 128
    color 102 0 51
 ###   name "Hospitals"
    outlinecolor 0 0 0
    # 1 label arial
    label
      angle auto
```

```
      color 0 0 0
      font "arial"
      force false
      maxsize 8
      mindistance 15
      minfeaturesize auto
      minsize 8
      offset 2 2
      outlinecolor 255 255 230
      position lr
      size 8
      type truetype
    end
  end
end

# 11 layer schools
layer
  classitem "SCHOOL_NAM"
  data "schools_geo"
  group "schools"
  labelcache off
  labelitem "SCHOOL_NAM"
  labelmaxscale 20750
  maxscale 20750
  name "schools"
  postlabelcache true
  status on
  type point
  # 1 class schools
  class
    backgroundcolor 128 128 128
    color 255 255 201
 ###   name "Schools"
    outlinecolor 0 0 0
    symbol "school_flag"
    # 1 label arial
    label
      angle auto
      color 0 0 0
      font "arial"
      force false
      mindistance 15
      minfeaturesize auto
      offset 2 2
      outlinecolor 255 255 230
      position lr
      size 7
      type truetype
    end
  end
end

# 11 layer north arrow
```

```
  layer
    name "northarrow"
    postlabelcache true
    status default
    transform off
    type annotation
    # 1 class north arrow
    class
      symbol "north_arrow"
    end
    feature
      points 15 25 end
    end
  end
  # 11 web imaps.html
  web
    footer "imaps_footer.html"
    header "imaps_header.html"
    imagepath "/var/www/html/tmp/"
    imageurl "/tmp/"
    template "imaps.html"
  end
  # 12 querymap hilite
  querymap
    color 255 255 0
    size 200 200
    status on
    style hilite
  end
  # 13 reference graphics/reference.gif
  reference
    color -1 -1 -1
    extent -79.954891 43.138133 -79.725663 43.367328
    image "images/reference.gif"
    outlinecolor 255 0 0
    size 120 120
    status on
  end
  # 14 legend #none
  legend
    keysize 10 10
    status embed
    postlabelcache true
    # 1 label #none
    label
      color 0 0 0
      font "arial"
      size 8
      type truetype
    end
  end
end
```

Appendix B

GIS MAPPING WORKSHOP

NOVEMBER 2, 2001

Draft

McMaster University
Togo Salmon Hall (TSH) 202 A

8:00 am – 1:30 pm

TIME	ACTIVITY
8:00-8:45	Registration
8:45-9:00	Introductions/Welcome
9:00-9:30	Health GIS Discussion
9:30-10:00	Demonstration of Site
10:00-10:15	Coffee break
10:15-11:15	Site Exploration
11:15-11:30	Questions and Answers
11:30-11:45	Administering Survey
11:45-12:15	Site Exploration
12:30-1:30	Lunch, Faculty Club

Appendix C

'Mapping Health on the Internet' Workshop Survey

- The purpose of this survey is to improve the use of Geographical Information Systems (GIS) on the Internet, including our specific site for Public Health created at McMaster University.
- Please circle your choice or write your response as appropriate.
- Please note that we will keep everything you tell us strictly confidential. Any reports will include only grouped responses.

Thank you for your co-operation

Part One: Users Descriptive Information

1. Name____

5. E-mail ____

6. Education (Circle all that apply) Discipline

   a) College Degree ________________________

   b) B.A. ________________________

   c) B.Sc. ________________________

   d) M.A. ________________________

   e) M.Sc. ________________________

   f) M.D. ________________________

   g) PhD ________________________

   f) Other (please specify)________________________

7. What is your Job Title? ________________________

8. What is your age? ________________________

Part Two: Your Current Understanding of GIS Introductory Questions

	STRONGLY AGREE	AGREE	NEITHER AGREE NOR DISAGREE	DISAGREE	STRONGLY DISAGREE
1.I have a good understanding of GIS/Digital Mapping	1	2	3	4	5
2. I have a good understanding of On-Line Analytic Processing (OLAP)	1	2	3	4	5
3. I have a good understanding of Internet mapping	1	2	3	4	5

4. What type of health surveillance or programming activities do you use on a regular basis? (circle all that are appropriate)

a) COMPILE AND MANAGE HEALTH DATA

b) DISEASE CLUSTER ANALYSIS

c) PREPARE HEALTH STATUS REPORTS

d) COLLECT DATA ON REPORTABLE DISEASES

e) ORGANIZE PUBLIC HEALTH PROMOTIONAL PROGRAMS

f) OTHER (please specify)

	STRONGL Y AGREE	AGREE	NEITHER AGREE NOR DISAGREE	DISAGREE	STRONG LY DISAGRE E	DON'T KNOW
5. I think using spatial (geographic) information in my work could help me to do my job better?	1	2	3	4	5	6

Comments:

Part Three: Technology Questions

	STRONGLY AGREE	AGREE	NEITHER AGREE NOR DISAGREE	DISAGREE	STRONGLY DISAGREE	DON'T KNOW
1. I see potential for using Internet mapping in my organization	1	2	3	4	5	6

.

2. Please rate how strongly you agree or disagree with the following statements on to the potential benefits of Internet mapping to your organization.

	STRONGLY AGREE	AGREE	NEITHER AGREE NOR DISAGREE	DISAGREE	STRONGLY DISAGREE	DON'T KNOW
INCREASED ACCESS TO MAPS	1	2	3	4	5	6
INCREASED ACCESS TO DATA	1	2	3	4	5	6
ABILITY TO SHARE DATA AND ANALYSES	1	2	3	4	5	6
ABILITY TO PRODUCE BETTER PAPER MAPS	1	2	3	4	5	6
EASE OF USE	1	2	3	4	5	6
ABILITY TO EXPLORE SPATIAL RELATIONSHIPS	1	2	3	4	5	6
ABILITY TO PERFORM CLUSTER ANALYSIS	1	2	3	4	5	6
ABILITY TO PERFORM ADVANCED STATISTICS (E.G., REGRESSION)	1	2	3	4	5	6

Please describe other possible benefits of Internet mapping to you or your organization.

3. In general how strongly do you agree or disagree with the following statements onthe potential limitations of Internet mapping in your organization.

	STRONGLY AGREE	AGREE	NEITHER AGREE NOR DISAGREE	DISAGREE	STRONGLY DISAGREE	DON'T KNOW
LIMITED AMOUNT OF DATA	1	2	3	4	5	6
DIFFICULT TO PRINT MAPS	1	2	3	4	5	6
DIFFICULT TO PINPOINT SPECIFIC AREA	1	2	3	4	5	6
DIFFICULT TO USE	1	2	3	4	5	6
LIMITED SET OF ANALYSIS CAN BE PERFORMED	1	2	3	4	5	6
COST INVOLVED WITH STARTUP	1	2	3	4	5	6
LIMITED ACCESS TO DATA	1	2	3	4	5	6
REQUIRES STAFF RESOURCES	1	2	3	4	5	6

Please describe other possible limitations that your organization might encounter using Internet mapping:

4. How would you use Internet mapping technology, under the assumption the data was relevant to a project your organization has asked you to undetake?

a) A TOOL FOR COLLABORATION (in support of health surveillance)
b) A TOOL FOR DATA VISUALIZATION
c) A TOOL FOR EXPLORING THE DATA WITH SPATIAL OVERLAYS
d) A TOOL FOR DATA ANALYSIS
e)OTHER____

Comments:

____

____

____

____

____

____

____

____

____

Part Four: "Prototype Mapping Site" Questions

IN THIS SECTION WE WANT FEEDBACK ON THE SITE YOU TESTED TODAY

	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree	Don't Know
1. I found the on-line tutorial helpful	1	2	3	4	5	6
2. The prototype is easy to use	1	2	3	4	5	6
3. The prototype was intuitive in its use	1	2	3	4	5	6
4. The prototype had an appropriate colour scheme	1	2	3	4	5	6
5. The layout (design) of the prototype was appropriate	1	2	3	4	5	6
6. There would be no problem allowing full public use of the prototype	1	2	3	4	5	6

Comments:

6. Please rate the functionality of the prototype. How useful would the following features be for your work?

	VERY USEFUL	USEFUL	NEITHER USEFUL NOR USELESS	NOT USEFU L	USELESS
LOOK UP A POSTAL CODE	1	2	3	4	5
DIFFERENT THEMES OVERLAY	1	2	3	4	5
'ZOOM IN' and 'ZOOM OUT'	1	2	3	4	5
QUERY A SPECIFIC CENSUS TRACT	1	2	3	4	5
USE A MULTI-CRITERIA QUERY TO DETERMINE WHICH CENSUS TRACTS MEET MULTIPLE CRITERIA	1	2	3	4	5

Comments:

7. How useful do you think the following additional prototype functionality would be if implemented?

	VERY USEFUL	USEFUL	NEITHER USEFUL NOR USELESS	NOT USEFU L	USELESS
ABILITY TO DETECT CLUSTERS	1	2	3	4	5
ABILITY TO INTERPOLATE VALUES	1	2	3	4	5
AN ON-LINE RULER (TO MEASURE DISTANCES)	1	2	3	4	5
ABILITY TO ALTER MAP COLOURS	1	2	3	4	5
ABILITY TO UPLOAD AND VIEW YOUR OWN DATA	1	2	3	4	5
ABILITY TO PRODUCE GRAPHS AND CHARTS	1	2	3	4	5

Comments:

8. The next series of items asks you to rate the prototype you worked on today. For each item there are two descriptive words. Choose a number between 1 and 5 to reflect your opinion. Choosing "1" means you believe the service technology fits the first descriptor. A "5" places your opinion with the second descriptor. A "3" means neutral. A "2" or "4" would qualify your rating.

sophisticated	1	2	3	4	5	unsophisticated
appropriate	1	2	3	4	5	inappropriate
connected	1	2	3	4	5	disconnected
relevant	1	2	3	4	5	irrelevant
straight forward	1	2	3	4	5	confusing
personal	1	2	3	4	5	impersonal
timely	1	2	3	4	5	untimely
positive	1	2	3	4	5	negative
valuable	1	2	3	4	5	worthless
believable	1	2	3	4	5	not believable
reliable	1	2	3	4	5	unreliable
acceptable	1	2	3	4	5	unacceptable
clear	1	2	3	4	5	unclear
interesting	1	2	3	4	5	uninteresting
readable	1	2	3	4	5	unreadable
intuitive	1	2	3	4	5	not logical

Part Five: Workshop Questions

	STRONGL Y AGREE	AGREE	NEITHER AGREE NOR DISAGREE	DISAGR EE	STRONGLY DISAGREE	DON' T KNO W
1. I found the workshop helpful.	1	2	3	4	5	6
2. Assuming that a full set of data pertinent to my geographic region is available, I would use Internet Mapping in the next year.	1	2	3	4	5	6
3. I feel confident that I could effectively use an Internet Mapping system.	1	2	3	4	5	6

4. Do you have any additional comments you have concerning the workshop itself?

Appendix D: Ethics Approval

"A professor in the School of Geography and Geology called the Research Ethics Officer for clarification on the need for ethics approval for a webpage site on Asthma and air quality in Hamilton-Wentworth, that would be published online in a month's time. The data on asthma in the region was collected by a research protocol approved by the St. Joseph's Healthcare REB. It appeared that the secondary use of this data was of some ethical concern. There was also some concern that making the data publicly available would have a community impact on such things as real estate prices for example. The data was correlated against census data and other air quality measurements in the region. It would be obvious doing a navigation of the map where the highest risk areas in the region are. The researchers at first wanted to invite the board to view the data at the new Multi-Media computer laboratory, but it was decided to make the website available for the board members to view online and return some feedback. There has been heightened ethical concern about research on asthma since the research death of a healthy volunteer last summer in an asthma clinical study at John Hopkins. In this case, the Chair and the Research Ethics Office saw no need to review the protocol, but would let the board have access to the website for comment.

A member of the board said that to suppress the research data would be dreadfully unwise. The data map shows that the lowest income levels are generally where the asthma rates are highest. It would be more serious if the ethics board should repress studies that might be controversial. It is important for ethics boards to take a clear stance that this should not be done. Another member said that there is something about asthma in the newspapers everyday and that the public is generally well informed. A real ethical concern is when people sign contracts with drug companies and the companies do not want negative results about drugs published. The Research Ethics Board could not keep negative results like that secret. Certain sorts of research might be damaging to communities but this should be upfront rather than at the end. There would be more ethical concerns if the research on the communities was about race or intelligence. The ethics boards sees proposals but seldom sees the results, as that is up to the researchers to make the decisions, as in aboriginal communities. The ethics board doesn't see or rule on that. If you look at the model of participatory research, which is being used quite extensively these days, the researcher communicates with the community right from the beginning. There is some responsibility to see that the participatory agenda is followed with input from the community."

Best regards
Michael

Michael J. Wilson, Research Ethics Officer,
Office of Research Services, McMaster University
GH 306-K, 1280 Main Street West, Hamilton, Ontario L8S 4L9
Tel: (905) 525-9140, ext. 23142, Fax: (905) 540-8019
E-Mail: srebsec@mcmaster.ca,
Web Site: http://www.mcmaster.ca/ors/ethics/mainmenu.htm