A GeoPrimer:

Environmental Public Health Tracking

Version 1.0

(A resource for EPHT managers and a tool for their technical staff)

Membership in the Geography and Locational Referencing Subgroup
(Alphabetical by state or agency the person is representing)

Last Name	First Name	Representing	Last Name	First Name	Representing
Falade	Makinde	CA	Fox-Williams	Kathleen	NV
Wolff	Craig	CA			
			Boscoe	Frank	NY
Sagaram	Deepak	CDC	Grady	Sue	NY
Thames	Sandy	CDC	Le	Linh H.	NY
Wall	Patrick	CDC	Talbot	Thomas	NY
Kochukoshy	Cheruvettolil	Ross&Associates	Brundage	Thomas	OR
Tosta	Nancy	Ross&Associates	Counter	Marina	OR
			Cude	Curtis	OR
Duclos	Chris	FL	Everman	Elizabeth	OR
Johnson	David	FL	Garland	Rodney	OR
Kearney	Greg	FL	Murphy	Megan	OR
			West	Mitch	OR
Reynolds	Thomas	Houston			
Rogers	Peggy	Houston	Arbegast	Dan	PA
			Miller	Patricia	PA
Evans	Lloyd	IL	Sieber	Joseph	PA
			Stengle	Robert	PA
Streeter	Robin	Johns Hopkins University	Unruh	Karen	PA
			Xue	Mei	UT
Silverman	Marc	MA			
			Hoskins	Richard	WA
Aranda	Gina	NM			
Bales	Chandra	NM			
Penman	Shawn	NM			

Contents

This primer is meant to be a resource for the managers of environmental public health tracking (EPHT) projects throughout the United States. They can give copies of the GeoPrimer to their information services technical staff as the planning for the local EPHT project is underway. This GeoPrimer does not replace the many fine technical documents about geographic information system (GIS) applications; rather the intent with the primer is to provide material that would become a common language for communicating the needs and functions of GIS in the EPHT project.

1.0   Introduction

Information about geography and location are critically important to environmental public health tracking, primarily because exposure to environmental hazards is often a function of place. Technology that has become commonly available over the last three decades makes it increasingly possible to track the relationship between environmental hazards and health conditions. Hardware such as global positioning systems (GPS) and software for geocoding can be used to generate accurate locations for hazardous materials, specific health conditions, and other features of interest. A geographic information system (GIS) can be used to integrate, analyze, and display the locational data in various ways to establish relationships among variables. The successful use of these tools for environmental public health tracking is dependent on the development and availability of high quality geographic data.

1.1   How to Use this Document

This GeoPrimer was developed as an introduction for planners, managers, and those people implementing environmental public health tracking. It provides a simple overview of many of the terms used in the process of manipulating geographic data, as well as pointers to additional resources. It is not intended to be a comprehensive guide. Included are descriptions of basic geographical data processes and standards that support the use of geospatial tools. Definitions of commonly used terms in geographic data analysis are also provided. The underlined terms in the text are defined in "Commonly Used Terms," section 2.6.

1.2   Background

Public health and environmental agencies make significant use of data to understand relationships, analyze problems, and communicate results and information. Geographic data are critical for performing the following functions:

- identifying the location of environmental hazards relative to locations of diseases,
- monitoring the distribution of pollution over time,
- analyzing disease trends and patterns over time, and
- associating locations of diseases with various populations (including those populations especially at risk).

In geographic terms, these functions can be summarized as the ability to

- map locations – identify where events and features occur and display those locations;
- map quantities – depict quantitative relationships, such as "most," "least," "average;"
- map densities – show characteristics relative to an area occupied (e.g., number of people per acre or pollutants per square mile);
- determine distance – calculate relationships among features in space;
- monitor changes in space – compare events and features at different points in time; and
- conduct spatial analysis – determine relationships among events and features based on geography.

Managers, planners, and implementers of environmental public health tracking efforts can use mapped information to evaluate resource allocations, set priorities, plan interventions and programs, track outcomes of interventions and public health policies, and research environmental health linkages.

2.0 Basic Concepts and Terminology

This section describes fundamental concepts and terms commonly used in locational analyses.

2.1 What are geographic data?

The phrase "geographic data" is generally used to refer to data that are linked to a location on the Earth. Various techniques are available to establish a "linkage." These include locating a feature on a map and manually deriving coordinates or reference points to using advanced technology to link an address to coordinates. The coordinates may be precise (e.g., high-end GPS receiver) or approximate (e.g., centroid of a ZIP codes or census tract). Data may represent a single point, a line (e.g., road or river), or a polygon (e.g., county, building footprint, agricultural field).

Aerial photographs, satellite imagery, and scanned or digitized maps are forms of geographic data. Increasingly, geographic data are digital in form (digital orthophotography). Maps serve as a means to display geographic data. The terms "geospatial" or "spatial" are sometimes used in place of "geographic," and generally have the same meaning.

2.2 What are geographically referenced data?

The terms "locationally referenced" and "geographically referenced" (georeferenced) are technical terms indicating that coordinates or geographic addresses have been developed for a data set. Various systems can be used to establish location, including street address, latitude/longitude, ZIP codes, census tract, city, and county, etc. The more accurately something is locationally referenced, the more accurately it can be integrated with other geographic data sets. Geographically referenced data may be derived by digitizing or scanning maps, processing address lists to develop coordinates (this is referred to as "geocoding"), or using GPS devices to develop coordinates during field sampling.

2.3 What is a geographic information system (GIS)?

A GIS is a computer-based system of hardware, software, and procedures used to manage, manipulate, analyze, model, represent and display georeferenced data. These data layers may be physical, cultural, or economic. A GIS provides the means to address complex problems involving the interpretation and integration of these data in space.

A GIS database includes two main types of data. One type is a spatial database, which contains location data and describes the geography of earth surface features (shape, position), along with the relationships among these features. These features are most often recorded as digital coordinates in the form of points, lines, or polygons. The second type is attributes about the geographic features. These data types are integrated to varying degrees in different GIS packages. The GIS provides the ability to relate the attribute information to the spatial characteristics. Some GIS software now store data about a feature or place, its coordinates, and its attributes in one database.

2.4 What is a GIS used for?

A GIS is used for many purposes, but primarily as a means to geographically relate data so the information can be displayed in a way that enhances understanding. For most people, pollution or diseases (and their relationships to other variables) can be better understood when they are

displayed on a map, as opposed to simply using words. A GIS can be used at many scales, ranging from neighborhood to national, and can display the data in detail or as summaries at a ZIP code or census tract level.

2.5   Who uses GIS?

All federal health and environmental agencies and many state and local agencies are using GIS to assist in performing public health functions. GIS is a key component of environmental health tracking systems because it provides the means to connect data geographically. Initiatives, such as community right-to-know and environmental justice movements, have also led community-based organizations to adopt GIS to map relevant neighborhood data. The ability to analyze geospatial data allows "communities of interest" to better observe, understand, and make decisions for the benefit of their members.

2.6   Commonly Used Terms

The following are commonly used geospatial data and technology terms:

Attribute – a term used to describe a category of information within a data layer. This information is stored in a table that has a column for each *attribute* or category of information (e.g., attributes in a "housing" layer of a dataset might include location, zoning, building age, and square footage).

Data layers – a term used by some GIS software packages to describe the organization of the data within the software. Similar features, such as streets, freeways, and trails; or housing and industrial and commercial buildings; or lakes, streams, and rivers, are organized and stored as a set of data that can be thought of as a "layer" of data. The data layers often represent different overlays on a map. Figure 1 shows examples of three layers: areas or polygons of counties with lines of roads and points of hospitals.

Figure 1: Depiction of Data Layers
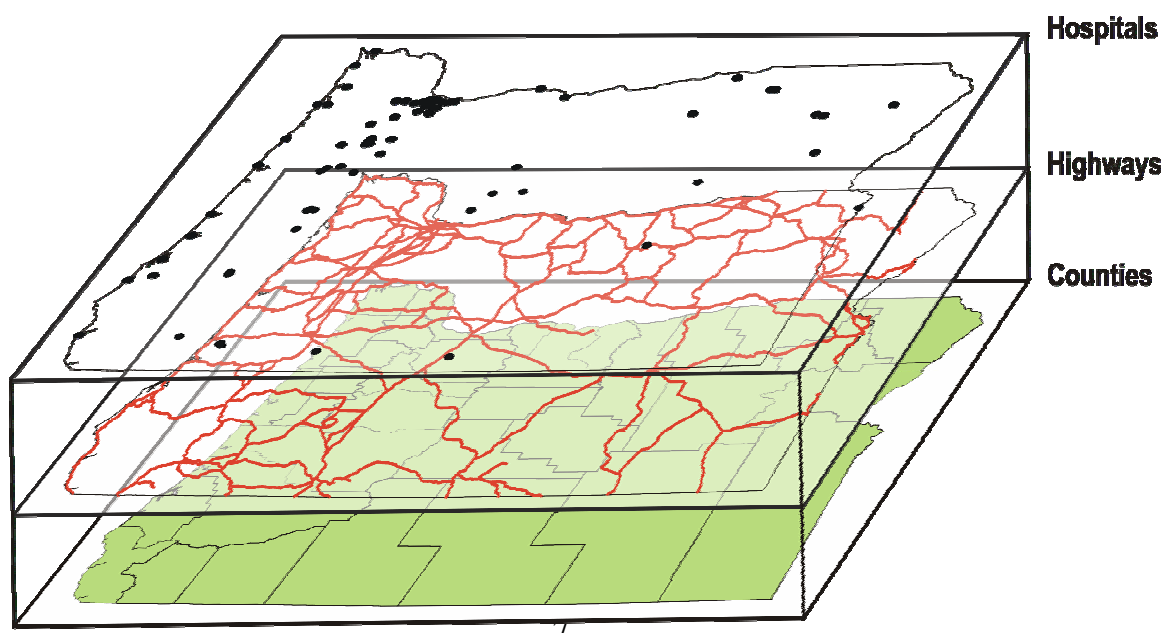
Digital orthophotography – a base layer of digital photography that has been corrected so that streets, buildings, and other features are shown in their true map position. Corrections are usually made using elevation reference points. Most of the digital orthophotography produced in the public domain is represented at a scale of 1:12,000. These represent one quarter of a United States Geological Survey (USGS) 7.5-minute quadrangle, and are referred to as digital ortho quarter quads (DOQQs) or digital ortho quads (DOQ).

Geocoding – the process of determining the coordinates of a specific location based on its street address or its existence within a known region (e.g., ZIP code, census tract). Coordinates can be assigned as x and y coordinates or, most frequently for GIS purposes, as longitude and latitude coordinates.

Geospatial analysis – the process of manipulating data based on location. Examples of questions that can be analyzed include:

- What is the distribution of asthma cases?
- What hazards are present within some distance (e.g., 5 miles) of these asthma cases?
- What is the distribution of birth defects within this region?
- Where are the water wells that do not meet water quality standards?

Geospatial tools – any of the tools and technologies commonly used in the course of conducting geospatial analyses. These may include GIS, GPS, remote sensing, and others.

GIS (Geographic Information System) – a system of computer software, hardware (plotters, digitizers, servers, etc.), geographically referenced data, and personnel trained in the use of the software that supports the ability to manipulate, analyze, and display data tied to locations.

GPS (Global Positioning System) – a means to determine a position on Earth, in any weather. The system includes a minimum of 24 GPS satellites that orbit at 11,000 nautical miles above the Earth and are continuously monitored by ground stations located worldwide. The satellites transmit signals that can be detected by anyone with a GPS receiver. With use of the receiver, a location can be determined with great precision. The locational accuracy of GPS can vary from 100 to 10 meters for most equipment. With military-approved equipment or equipment and software systems that correct GPS signal errors (differential corrected GPS), accuracy can be pinpointed to within 1 meter.

Locational referencing (georeferencing) – is the process of assigning a feature of interest in the landscape (e.g., hospital, hazardous waste site) a set of coordinates that then support the ability to analyze those features using software such as GIS. Geocoding is the process most commonly used to locationally reference attribute data.

Metadata – are "data about data." They describe the content, quality, condition, and other characteristics of data. Metadata help to locate and understand data. See www.fgdc.gov for more information.

Points, lines, and polygons – geographic features that are manipulated in a GIS are generally represented as points, lines, or polygons. Different features lend themselves to different representations. For example, roads and rivers are frequently lines (very wide rivers may be thought of as an area or polygon). ZIP codes or census tracts are areas or polygons, and hospitals are considered points (unless the analysis is done on a very small area, in which case the perimeter of the hospital might be thought of as a polygon).

Remote sensing – the use of cameras, digital cameras/scanners, and other data capture techniques to record the characteristics of features from a distance. Remote sensing commonly refers to data captured by airplane or satellite. See examples at http://www.terraserver.com.

Scale – the ratio of the size of something displayed on a map to its true size on the Earth (i.e., the relationship between distance on the map and distance on the ground). A map scale usually is given as a fraction or a ratio—1/10,000 or 1:10,000. The U.S. Geological Survey produces topographic maps at 1:24,000 scale (1 inch equals 24,000 inches or 2,000 feet). The larger the scale of the map (the closer it gets to a 1:1 depiction), the more accurately something can be depicted. Large-scale maps can be reduced in scale, without loss of accuracy. Small-scale maps, such a 1:500,000 state map, cannot be enlarged without introducing errors (see http://erg.usgs.gov/isb/pubs/factsheets/fs01502.html).

3.0   Environmental Public Health Tracking and Geospatial Analysis

3.1   Accurately Locating Environmental and Health Data

The ability to conduct environmental public health tracking is, in many ways, dependent on the ability to establish linkages between hazards in the environment, exposures to these hazards, and health conditions. Tracking is therefore dependent on developing relationships between hazards, exposure, and health outcomes. These relationships are frequently established on the basis of common time frames and common locations (e.g., individuals with specific health conditions have been within some proximity of certain environmental hazards within a specific time frame).

Establishing accurate linkages is dependent on the availability of accurate geographic and temporal (time) data. Most health and much environmental data are referenced by street address. These must be converted to coordinates that can be incorporated and used with GIS software through a process of geocoding. Geocoding is discussed in further detail below. Within the health community, given the availability of addresses, geocoding is the most common approach for accurate geographic encoding of health data. Other approaches to establishing locational data include the use of global positioning system (GPS) devices, scanning paper maps, digitizing maps (both on digitizers and on-screen), and importing already digitized data or remotely sensed imagery.

3.2   Geocoding

3.2.1   Basic Concepts

Address geocoding is a common approach for preparing environmental and health data for geographic analyses with tools such as GIS. Geocoding requires two primary data sets:

- A database that includes addresses to be geocoded (usually these are typical street addresses: 123 Main Street, Anytown, State 00936.) This is referred to as the "address file" in the discussion that follows.
- A georeferenced database of street locations with attributes (e.g., street names, address ranges, street types, etc.). This is referred to as the "road database" in the discussion that follows.

Additionally, software to perform the geocoding is needed. If the software is not available locally, geocoding services are available for purchase on the Web. Users of commercial services should ensure that the data they are geocoding are covered by appropriate privacy and confidentiality agreements.

Cities, counties, and local municipalities generally create the most accurate databases of roads and associated address ranges. The activities of local governments (e.g., zoning, land use, transportation planning, and assessing) provide an incentive for developing accurate street locations and names, and for maintaining database currency. If these databases are available, environmental health professionals should consider taking advantage of them. Increasingly, national data sets such as those developed by the U.S. Census Bureau and private companies are relying on locally developed data sets.

3.2.2 Common Errors

Geocoding matches the street name in the address file to a name in the road database. For an exact match, each component of the address must be the same (the prefix, number, name, road type, suffix, etc.). If any of these in the address file are different than in the road database, an error in geocoding may result. If they match, the software assigns a match between the file and the database. In some cases, very accurate road databases store exact address locations that have been developed by using a GPS device on each residence or building along the street, providing a precise address for the geocoding software to match.

Several sources of error can occur during geocoding. These are displayed in Figure 2, below. When geocoding does generate errors, particularly in addresses that cannot be matched, manual efforts for cleanup and matching may be required. Address files should be checked to ensure addresses are accurate. In some cases, it may be necessary to use aerial imagery to identify the location of buildings or residences and to geocode these locations manually on a computer screen. This approach is often used when the importance of an accurate match is critical. In other situations, where matching to a street name is not possible, addresses may be coded to a ZIP code or census tract.

Figure 2: Sources and Fixes for Geocoding Errors

Potential Source of Error	Approach or Solution to Resolve
Addresses may be duplicated. More than one "Main Street" may exist in a state.	Use city names or ZIP codes to refine the possible area for address matching.
Addresses may be misspelled or inaccurately represented. For example, "Maine Street" would not match "Main Street."	Software can be adjusted to overlook these minor differences in spelling (although it creates a risk when these really are different streets).
Address files may include post office boxes rather than addresses. Prefixes or suffixes may be missing altogether. Institutional names (e.g., a nursing home)	No easy fix. Need to find or generate an actual address. Adoption of an address standard that requires that certain fields be filled in could also

or building numbers (e.g., apartment numbers) may not be included.	assist.
"Northwest" in the address file will not match "NW" in the road database.	Develop an "alias table" where the software is told that "Northwest" and "NW" mean the same thing.
Road databases may not be geographically accurate. The accuracy depends primarily on how the road data were collected (e.g., via GPS, digitized from a map, hand drawn).	GPS tends to produce the most accurate geographic coordinates. Small-scale maps (of a state or the nation) are much less accurate geographically than larger scale maps (of a neighborhood or city). The analysis being conducted determines the geographic accuracy needed.
Roads may be missing (e.g., new subdivisions).	Determine the currency of data and review metadata. Examine recent aerial photographs to identify missing features.
Road databases may have incorrect attributes. Street names may not be accurately encoded in the road database (missing or misspelled). Rural route addresses are not typically included in road databases.	Clean up the road database to meet the needs of the analysis.
ZIP code boundaries can change frequently.	Know the dates of both the address files and road databases and ensure they are appropriate timeframes for geocoding.
Geocoding against address ranges can introduce positional errors because the software assumes equal distribution of addresses on a block. This can be an issue in rural areas, where residences are not evenly distributed, or in urban areas that have significantly different lot sizes on a block.	Encoding exact addresses via GPS is one solution.
Geocoding software is based on proprietary approaches using various assumptions to solve address or matching problems. The approaches are not all the same, meaning that different coordinates may result when address files are geocoded with different software packages	Know the vendor and the assumptions being made (algorithms being used) in the software.

3.2.3   Centralized Geocoding Services

Increasingly, geocoding vendors provide centralized services that match addresses against multiple street centerline datasets, exposing their functionality and data holdings over the Internet in standard request/response flows (e.g., Web Services/SOAP). These types of services open the opportunity for testing the ability to geocode an address very early in the reporting process, possibly in real-time. For reporting systems that already use standards-based electronic reporting mechanisms, geocoding services can markedly increase the quality of geocoded addresses and decrease the overall cost of ownership of geocoding functionality. In this manner, address data is geocoded at the same time that the address is key-entered into the reporting system. Real-time edits can be requested of operators who enter incorrect addresses, or more information can be requested of the operator to perform a successful geocode. Much time and energy will be saved, because geocodable datasets on the Network will be available, essentially, pre-geocoded.

3.3   Geospatial Analyses

One of the primary purposes of geocoding or otherwise geographically capturing environmental and health data is to be able to perform analyses with those data. There are many different types of geospatial analysis possible with GIS software. Examples are described below.

- Analyses that involve examining one data set in relation to other data, as in determining the population served within the radius of 1 mile (or any distance) of a hospital.
- A buffer analysis to assess the number of potential hazards within some proximity of a school.
- Analysis to determine if a point, line, or area dataset is spatially coincident with another point, line, or area dataset. For example, agricultural data such as crop type or pesticide use, might be combined (overlaid) with census tract boundaries. The field level agricultural data are then apportioned to the tracts based on a weighting of their occurrence within the tract. This provides a tract-level estimate of the agricultural data.
- Analysis to examine data in a form different from how they were originally collected (modeling/interpolation). For example, air quality monitoring data collected at a point location might be interpolated over a surface that includes a regional estimate of air quality.
- Analysis to calculate the distances between various features, such as the distance from freeways to severe cases of asthma, or the proximity to nuclear power stations to cases of childhood leukemia. (There are numerous variations on distance calculations that can be performed with a GIS, including nearest neighbor and distance weighted calculations.)
- Network routing analysis. This may entail developing shortest routes for ambulance pathways in a road network or determining potential traffic exposures when driving from home to work. Environmental analyses may also be conducted to assess the risks associated with a pollution spill in a waterway as it flows downstream.

It is imperative that the GIS staff work closely with epidemiologists and public health researchers to make certain that the environmental hazard, exposure and health outcomes data can be brought together in a way that is meaningful. The accuracy of analyses is a function of many factors. The quality and quantity of the input data must be considered. A low-resolution dataset with few observations might be better represented in a spatial linkage operation that aggregates over a large region. A definition for a health event metric (e.g., rate of disease over a period) and an environmental hazard event metric (e.g., contamination level over a period) must be identified before a geospatial operation can be performed and the reliability of the linkage must be considered. If the accuracy is dependable, subsequent statistical analyses should determine whether a relationship exists.

Temporal issues can complicate the analysis process. Conditions that change over time must be assessed and means to accurately represent the changes considered. Frequently, an environmental health study involves exposure and latency periods that consist of aperiodic hazards (e.g., high levels of air pollution on some days), hazards that change in concentration over time (e.g., groundwater contamination), and individuals who move or travel into and out of exposure to the hazard. Representing these factors spatially and assessing how they affect the outcome of an analysis can be very challenging due to lack of data availability, software limitations, and the need for technical and health expertise integration.

3.4   Geospatial Display (Maps)

The results of these types of analyses may be displayed in a variety of forms. A surface or map that shows risk of "exposure" could be generated, or "hot spots" might be depicted. The ability to display graphically the results of geospatial analyses is one of the most powerful aspects of the use of GIS. An example of geospatial analysis is shown in the next section.

4.0   An Example of Geospatial Analysis

Geospatial analysis of birth outcomes and asthma incidence data with respect to traffic volume metrics in Alameda County, California

The California Environmental Health Tracking Program is examining residential address-level indicators of traffic exposure in its Alameda County pilot project. The address-geocoded health events of interest are four asthma indicators (emergency room visits, outpatient clinic visits, symptom medication purchases, and maintenance medication purchases) and two reproductive outcome indicators (term low-birth-weight births and preterm births). The hazard events of interest are average annual daily traffic volumes along major roadways (freeways and arterials). Traffic volume is characterized around the health events through buffer analysis. Reports include the following metrics within each buffer:

1. Distance/direction to and volume of nearest roadway within buffer
2. Distance/direction to and volume of highest volume roadway within buffer
3. Sum of all roadway volumes within buffer

For each traffic volume metric reported, a distance-weighted volume metric is also computed. This estimates the exponential dissipation of contaminants at a constant distance from the health event to the street segment. An established web service returns a response that includes each of the traffic volume metrics computed for that point location within the specified buffer.

5.0   Effective Use of Geospatial Data and Technologies

Many organizations with an interest in using geospatial technologies for conducting environmental-health analyses are challenged by the rapidly changing technology and the need for expertise in use of the software. Their concerns are valid. The use of geospatial technologies requires expertise, data, and a willingness to learn in a rapidly evolving field. Some organizations have embraced GIS as an organizing approach for all of their efforts (e.g., Honolulu – http://www.honoluludpp.org/ResearchStats/) while others use GIS for specific applications (e.g., http://www.metro-region.org/article.cfm?articleid=1055).

5.1   Steps for Implementing Geospatial Technologies

Several steps will aid an organization in successful implementation of geospatial technologies. These are useful to consider, whether the organization is planning to base many activities on geography or has a limited program using simple GIS tools:

1. Conduct a needs assessment – What are the questions the organization is trying to address? What is the purpose for using geospatial technologies?

2. Conduct a resource assessment/inventory – What resources and capabilities does the organization currently possess (e.g., hardware and software, funding, data, and expertise)?

3. Match the needs to the capabilities and assess the "costs" (both in time and dollars) and consider whether the investment is possible. Organizations that have invested in GIS over the last few decades have recognized a significant return on investment after about a decade. Returns on investment in GIS are likely to occur more quickly as increasing volumes of data are available. (The costs of digitizing, scanning, and maintaining data are the most significant expense in any GIS installation.)

4. Establish and implement policies and approaches that ensure the quality of data collected by linking data collection to critical business functions and collecting data via "typical transactions" wherever possible (e.g., collect information about immunizations electronically at the time of immunization).

5. Identify lead personnel – both technical and political – to oversee the efforts. (The second highest costs for GIS are for securing, maintaining, and retaining expertise.)

6. Establish requirements for internal communication and coordination as interest in the technology spreads. This should include requirements for use of standards – such as metadata (see Section 5.2 below).

7. Invest in maintaining the quality and currency of resident data sets (this may include on-going training about the value of high-quality data collection, including coordinates).

5.2   Standards

Data standards are a frequent topic of discussion in the GIS world because most organizations use data collected by others. To do this, the data must be transferred or imported, the quality of the data must be understood, and the data must be represented in a format that can be integrated with the organization's existing databases. There are many different data standards and new standards being developed as the technology changes to employ approaches such as Web services.

Important data standards include standards for data transfer, data documentation (metadata), and data content. There are two major standards setting bodies in the geospatial world: the Federal Geographic Data Committee (FGDC) and the Open Geospatial Consortium (OGC).

The FGDC, formed by the President's Office of Management and Budget, has developed numerous standards of use for the geospatial community, including those interested in environmental public health tracking (see http://www.fgdc.gov/standards/standards.html).

The OGC has numerous accepted standards for GIS data and functions. A few OGC specifications are of special interest to the environmental public health tracking community:

1. Geography Markup Language (GML) – GML is an extensible markup language XML encoding for the transport and storage of geographic information, including both the geometry and properties of geographic features.

2. Simple Features Structured Query Language (SQL) – The Simple Features SQL Specification application programming interfaces (APIs) provide for publishing, storage, access, and simple operations on simple features (point, line, polygon, multi-point, etc.).

3. Web Map Service (WMS) – Provides three operations protocols (GetCapabilities, GetMap, and GetFeatureInfo) in support of the creation and display of registered and superimposed map-like views of information that come simultaneously from multiple sources that are both remote and heterogeneous.

4. Web Feature Service (WFS) – The purpose of the Web Feature Server Interface Specification   is to describe data manipulation operations on OpenGIS® Simple Features (feature instances) such that servers and clients can "communicate" at the feature level.

CDC's Public Health Information Network (PHIN) and National Electronic Disease Surveillance System (NEDSS) have initiated other relevant geospatial standards efforts. The recommended NEDSS standards include the following:

1. The North American Datum of 1983 (NAD83) shall be accepted as the standard datum for the NEDSS GIS component (http://www.towermaps.com/nad.htm).

2. The FGDC Geospatial Metadata Standard shall be used as standard for geospatial metadata (http://www.fgdc.gov/metadata/contstan.html).

3. Coordinates shall be stored with appropriate metadata, including data standard authority, data standard source, road network used for geocoding, roads layer version, date of geocoding, match level of geocoding, and how acquired (address matching using streets layer vs. GPS).

4. The basic standardized format for address shall follow the FGDC Address Data Content Standard (http://www.fgdc.gov/standards/status/sub2_4.html). These are based on the U.S. Postal Service address format for efficient delivery of domestic and international mail. The following additions will be made to this standard:

   - Institutional names – health care facilities, veterans' homes, correctional facilities
   - Jurisdictional – city, town, locality
   - County/Federal Information Processing Standards (FIPS) Codes
   - Unit/multi-building complexes – basement, penthouse
   - Road network – against which address was geocoded
   - Metadata catalogue ID

5.3 Collaboration and Access to Data

This section provides a short list of some sources of collaboration and access to GIS data. This list is not comprehensive. Many private, local, or federal entities also provide access.

> **FGDC Clearinghouse** (http://www.fgdc.gov/clearinghouse/clearinghouse.html) – The Federal Geographic Data Committee Clearinghouse list links to help you learn more about the Clearinghouse, who the members are, and how you can participate.

> **Geospatial OneStop** (http://www.geodata.gov) **–** Part of the Geospatial One-Stop E-Gov initiative to provide access to geospatial data and information.

> **Geography Network** (http://www.geographynetwork.com) – Environmental Systems Research Institute's (ESRI's) Geography Network directs customers to many free and paid ArcIMS services. Data from these services can be used from desktop ArcGIS products and ArcGIS server applications.

> **TerraServer** (http://terraserver.com) and **TerraService** (http://terraserver-usa.com) – TerraServer has a browser client allowing access to USGS DOQs and digital raster graphics. Applications can access the same data through the TerraService web service that implements the OGC WMS interface.

> **USGS** (http://gisdata.usgs.gov) – The National Center for Earth Resources Observation and Science's (EROS's) data center hosts multiple map services both in ArcIMS format and OGC WMS or WFS formats. The national map, national elevation data, and national land cover data are available through these services.

5.4 Additional Information

Cayo MR, Talbot TO. 2003. Positional error in automated geocoding of residential addresses. International Journal of Health Geographics 2:10. Available at http://www.ij-healthgeographics.com/content/2/1/10.

Buckley DJ. 1997. The GIS Primer. Available at http://www.innovativegis.com/basis/primer/primer.html.

Goodchild MF, Kemp KK, editors. 1990. NCGIA core curriculum in GIS. Santa Barbara: National Center for Geographic Information and Analysis, University of California. Available via http://www.ncgia.ucsb.edu/pubs/core.html.

Longley, PA, Goodchild MF, Maguire DJ, Rhind DW. 2005. Geographic information systems and science. 2nd ed. Hoboken, New Jersey: John Wiley and Sons.

Open GIS Consortium, Inc. 1999. OpenGIS simple features specification for SQL, revision 1.1. OpenGIS Project Document 99-049. 78 pp. Available at http://www.opengeospatial.org/docs/99-049.pdf.

Wiggins L, editor. 2002. Using Geographic Information Systems technology in the collection, analysis, and presentation of cancer registry data: a handbook of basic practices. Springfield, Illinois: North American Association of Central Cancer Registries. 68 pp. Available at http://www.naaccr.org/filesystem/pdf/GIS%20handbook%206-3-03.pdf.