2010

Geographic Approaches to Global Health

A Self-Directed Mini-Course

John Spencer
Jay Stewart
Becky Wilkes

MS-12-56

2010

Geographic Approaches to Global Health

A Self-Directed Mini-Course

John Spencer, Jay Stewart, Becky Wilkes

MS-12-56

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) through Cooperative Agreement GHA-A-00-08-00003-00 and is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with ICF International, John Snow, Inc., Management Sciences for Health, Palladium, and Tulane University. The views expressed in this publication do not necessarily reflect the views of USAID or the United States government. MS-12-56

USAID FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

Course Credits

Collaborating organizations involved in the development of *Geographic Approaches to Global Health* include:

- Taiwo Abimbola, Centers for Disease Control and Prevention
- Maria Au, United States Agency for International Development
- Rachel Blacher, ICF Macro
- Carlos Castillo-Salgado, Johns Hopkins University School for Public Health
- Steeve Ebener, World Health Organization
- Nate Heard, Humanitarian Information Unit, United States Department of State
- Mary Kratz, United States Department of Defense
- Libby Skolnik, Johns Hopkins University Center for Communication Programs
- John Spencer, MEASURE Evaluation
- James Stewart, MEASURE Evaluation
- Becky Wilkes, MEASURE Evaluation

We would like to acknowledge and thank the following experts for reviewing the course:

- Maria Au, USAID
- Shabani CISHAHAYO, TRAC Plus, Rwanda
- Mike Emch, Department of Geography, UNC-Chapel Hill
- Sabri GMACH, Regional Office of Public Health, Bizerte, Tunisia
- Eric Green, NYU/Population Council
- Janet Heitgard, CDC
- Carl Kinkade, CDC
- Jorge Luna, Mailman School of Public Health, Columbia University
- Michelle Monroe, US Department of State, Malawi
- Patrick NAPHINI, Ministry of Health, Malawi
- Adama NDIR, Ministry of Health and Prevention, Senegal
- Greg Russell, OGAC
- Virginia SIMUSHI, Zambia
- Hilary Spindler, UCSF Global Health Institute

Credits and Captions for Photos and Graphic Images

Special thanks are extended to those who provided the photos and graphic images that illustrate this course.

For general inquiries about the Geographic Approaches to Global Health please email MEASURE Evaluation by clicking here.

TABLE OF CONTENTS

SUMMARY

Purpose:

It is important to harness geography to manage, analyze, and leverage spatial data effectively when planning, monitoring, and evaluating health sector programs. Through this course, learners will gain an understanding of how to use spatial data to enhance the decision-making process for health program implementation in limited resource settings.

Time:

Approximately 3 hours

Objectives:

By the end of this course, the learner will—within a public health context—be able to

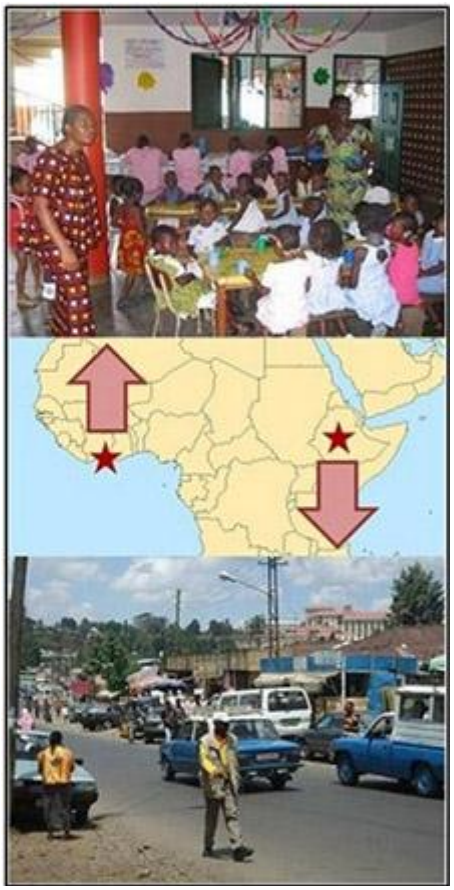
- Understand the basics of spatial data and its role in decision making
- Explain why geographic data and tools are important for decision makers
- Understand geography as a unifying framework and GIS as a tool to collect, link, analyze, visualize, manage, and share data and information
- Address the challenges and explore the opportunities of working with spatial data to produce geographic information and geographically based knowledge
- Distinguish between the different tools used to collect, analyze, and manipulate geographic data
- Communicate with technical specialists to develop program-specific geographic products

Audience:

This course is appropriate for public health program planners, managers, and professional staff who are interested in learning how geography and spatial data and tools can benefit their programs. The course is designed for individuals without a background in geographic information system(s) (GIS) or who are not specialists in medical geography

INTRODUCTION 1

EVERYTHING HAPPENS SOMEWHERE



Description: Everything happens somewhere, which provides a geographic component to human activity. Source: MEASURE Evaluation.

Human activities have a geographic component.

Except for the activities of a handful of individuals engaged in space exploration, everything we billions of people do occurs on Earth.

As a result, almost all human activity can be associated with physical locations on the surface of Earth, which can be mapped and analyzed to discern patterns.

The study of these patterns and their causes is known as geography.

Human health is intimately related to one's physical location on Earth. Where people live—or have lived in the past—can have a positive or negative influence on their health. The spatial pattern of disease in a population is vital evidence for public health analysts and decision makers when seeking to understand causes and plan interventions. In short, understanding the "where" helps understand the "why."

Note: In this course, the terms "geographic," "spatial," and "geospatial" will be used interchangeably.

Did you know?

The discipline of geography is concerned with the identification, analysis, and explanation of patterns—whether natural or man-made—on the surface of Earth, and provides a solid foundation for the study of human health. When applied to public or human health, the field is known as medical geography.

Highlights

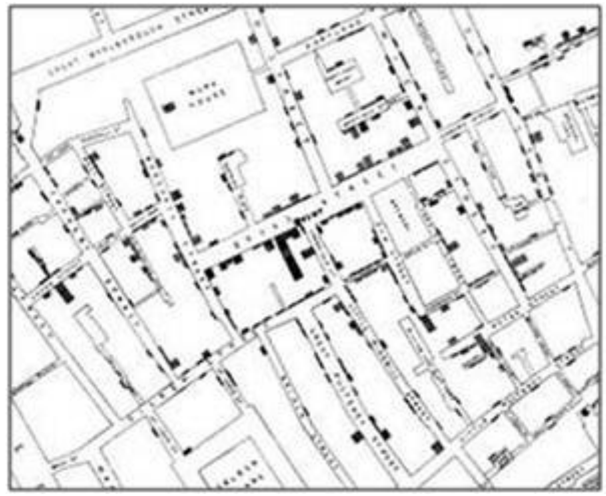
Geographic approaches to global health rely heavily on GIS. Definitions of GIS and other key terms used throughout the course can be found in the glossary.

DR. JOHN SNOW AND MEDICAL GEOGRAPHY

One of the earliest known examples of medical geography was Dr. John Snow's study of an 1854 cholera outbreak in London.

Dr. Snow observed that the cholera cases clustered around a particular public water pump on Broad Street (current-day Broadwick Street). Rejecting the prevailing theory that the outbreak was caused by miasma (bad air), Dr. Snow used a microscope to examine a sample of the pump's water, and found that it contained suspicious particles. Public officials reluctantly removed the handle from the pump, which effectively ended the outbreak.

Dr. Snow later published maps of the locations of cholera cases in relation to available water sources, and augmented his maps with statistics showing a positive relationship between water source and the disease.



Description: Excerpt from original 1854 cholera map by Dr. John Snow. Source: UCLA School of Public Health, Department of Epidemiology Web site, Jan 2010

> ***Did you know?***
>
> A useful introductory text on medical geography is [Medical Geography, Third Edition (Meade and Emch, 2010)]. The book is historical and international in scope, and surveys the perspectives, methodologies, and theories employed by geographers in the study of human health and disease.

Description: Map of HIV prevalence in Zambia by province, 2001. Source: MEASURE Evaluation

MAPPING AND EVIDENCE-BASED DECISION MAKING

Mapping quickly shows the "where" and helps develop better questions to understand the "why."

Modern medical geography uses computer-assisted mapping to identify patterns of human health.

The resulting maps, when based on high quality data that accurately represent real-world conditions, provide a ***quick and***

powerful means for decision making based on evidence rather than on opinion or pre-existing policy.

For an example of creating evidence by linking and mapping data from multiple sources, let's explore a situational analysis of orphans and vulnerable children (OVC) conducted by Catholic Relief Services (CRS) in eight states of Nigeria in early 2007.

Highlights

A map creates a picture worth a thousand words.

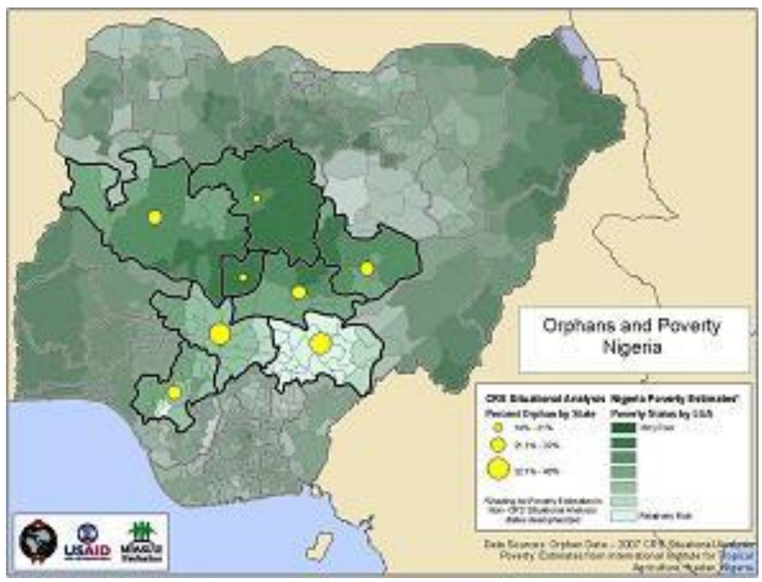
Did you know?

Mapping...

- Helps recognize spatial patterns
- Deepens insight into data
- Highlights data quality issues
- Provides a useful tool for decision support, analysis, and data display

MAPPING AND EVIDENCE-BASED DECISION MAKING IN ACTION

Map of orphans versus poverty in Nigeria that could be used for evidence-based decision making. To see a larger map, go to: https://www.cpc.unc.edu/measure/training/online-courses/my-courses/gis/files-and-links/S1_P4_Poverty_and_CRS_Orphans.pdf



Description: Nigeria: Percent orphan by state vs. poverty status by local government area. Source: MEASURE Evaluation.

According to the CRS situational analysis, issues of poverty make children more at risk for illness, malnutrition, and other undesirable conditions, and can contribute to the inability of guardians to care for OVC.

In order to explore the relationship between poverty and orphanhood, MEASURE Evaluation used common geographic identifiers to link CRS's state-level orphan data to International Institute for Tropical Agriculture (IITA) poverty data for local government areas (LGAs) and to create a map showing the relationship between these two theme.

Combining the previously separate CRS and IITA datasets in a map (see above) produced some surprising results. One might have expected a higher percentage of orphans (larger yellow circles) in areas of higher poverty (darker green shading). The map, however, revealed a different pattern: the percentage of orphans appeared to be lowest where poverty was highest!

Note: *Before linking datasets from different sources, it is important to ensure that there is no conflict between them, such as between the geographic extent or scale of the datasets or the time period for which the datasets are valid.*

> ***Ideas in Action***
>
> This example illustrates what a quick and powerful tool mapping can be for visualizing data and making decisions based on evidence rather than non-validated assumptions.

GEOGRAPHY STRENGTHENS THE NATIONAL DATA INFRASTRUCTURE

Geography allows datasets to be linked, which strengthens the national data infrastructure.

Data are often available from many sources, such as research institutes, health and education ministries, non-governmental organizations (NGOs), universities, and national mapping agencies (NMAs).

Many times these data streams are not connected and shared, whether through lack of awareness or administrative barriers. This is referred to as "stovepiping" of data. Using geography, datasets from different sources can be linked.

Linking datasets strengthens the national data infrastructure in a variety of ways.One of the most important ways linking datasets strengthens the national data infrastructure is by increasing organizational collaboration.

A strong national data infrastructure *strengthens data at multiple levels*, from local to national to international, and makes effective public health decisions more likely.

> ***Did you know?***
>
> Geography strengthens the national data infrastructure in *two key ways*:
>
> - Allows linking of datasets from different sources; and
> - Promotes standardization of data for sharing.

Table 1: District and Population

District	Population
North	3253
South	5621
East	8732
West	7715

Table 2: District and OVC Count

District	OVC
North	812
South	1011
East	2709
West	1411

Table 3: Data Joined by District

District	Population	OVC	Percent OVC
North	3253	812	24.96
South	5621	1011	17.99
East	8732	2709	31.02
West	7715	1411	18.29

Description: Geography is the key link between datasets. Source: MEASURE Evaluation.

> ***Ideas in Action***
>
> Linking datasets from different sources
>
> - Increases organizational collaboration for data creation, sharing, and maintenance
> - Extends inventory of data available to an organization
> - Potentially lowers cost of data creation and maintenance

KNOWLEDGE RECAP

1. In what key way(s) does geography strengthen the national data infrastructure?
    a. Creates more visually pleasing maps
    b. Allows linking of datasets from different sources
    c. Promotes standardization of data for sharing
    d. A and C
    e. B and C

2. Almost all human activity occurs on the surface of the Earth. What essential information does this provide for public health analysts and decision makers to help ask better questions and plan interventions to improve human health?
    a. Land cover
    b. Elevation
    c. Income, education, and occupation status
    d. Geographic locations
    e. None of the above

3. Mapping can accomplish which of the following?
    a. Recognition of spatial patterns
    b. Deeper insight into data and highlighting of data quality issues
    c. Provision of a powerful tool for decision support, analysis, and data
    d. All of the above
    e. None of the above

4. A strong national data infrastructure strengthens data at multiple scales, from local to national to international, and makes effective public health decision making more likely.
    True
    False

5. The field of geography is concerned with the identification, analysis, and explanation of what?
    a. Patterns in the physical environment
    b. Patterns in the man-made environment
    c. Patterns of human activity
    d. All patterns in relation to the surface of Earth
    e. Place names, capitals, cultures, and languages

KNOWLEDGE RECAP: ANSWERS

1. In what key way(s) does geography strengthen the national data infrastructure?

The correct answer is "e". Geography strengthens the national data infrastructure in two key ways: (1) allows linking of datasets from different sources and (2) promotes standardization of data for sharing. The successful integration of geography into the national data infrastructure may also create a greater demand for maps, although that is a secondary benefit. The artistic quality of maps may have no direct impact on the national data infrastructure.

2. Almost all human activity occurs on the surface of the Earth. What essential information does this provide for public health analysts and decision makers to help ask better questions and plan interventions to improve human health?

The correct answer is "d". The geographic locations of populations affected by health risks provide vital information for public health analysts and decision makers, as understanding the "where" helps ask better questions and plan interventions to understand the "why." Land cover, elevation, and socioeconomic status are examples of the many factors that can potentially influence a population's health based on its geographic location.

3. Mapping can accomplish which of the following?

The correct answer is "d". Some of the key strengths of mapping are that it helps recognize spatial patterns; deepens insight into data and highlights data quality issues; and provides a useful tool for decision support, analysis, and data display. Mapping is intended for graphic visualization of geographic data, not for publishing of text-based documents.

4. A strong national data infrastructure strengthens data at multiple scales, from local to national to international, and makes effective public health decision making more likely.

The correct answer is "true".

5. The field of geography is concerned with the identification, analysis, and explanation of what?

The correct answer is "d". The field of geography is concerned with understanding patterns on the surface of Earth, whether natural or man-made, including non-physical patterns generated by human activities (e.g., economic prosperity, airline or shipping routes, etc.).

STRATEGIC INFORMATION

Strategic information is the foundation for evidence-based program planning and decision making.

"Strategic information [SI] is knowledge that guides health policy, planning, programme management and service delivery. It is essential for evidence-based action at all levels of the health system" (WHO).

Description: Strategic information and strategic information systems. Source: MEASURE

SI is generated from raw data using a strategic information system (SIS).

Within the context of the US President's Emergency Plan for AIDS Relief (PEPFAR), an SIS would create SI through a synthesis of three key programs:

- ***Surveys and Surveillance,***
- ***Monitoring and Evaluation (M&E), and***
- ***Health Management Information Systems (HMIS).***

A complementary system used to collect, store, manage, analyze, display, and disseminate geographic data is known as a geographic information system (GIS).

PEPFAR considers GIS technology to be fundamental to the development of SI, as geographic mapping and analysis play an integral role in all three key SI programs.

Ideas in Action

Examples of Using GIS to Create SI:

Surveys and Surveillance

- Attaching GPS locations to survey data
- Mapping most-at-risk populations

M&E

- Improving routine health data via geographic linking
- Mapping health indicators

HMIS

- integrating map-based data viewers into health systems

NATIONAL DATA INFRASTRUCTURE

The data demand and use cycle integrates public health data into the national data infrastructure. Public health sector SI and SIS strengthen the national data infrastructure. The national data infrastructure is nourished by interlocking cycles of data demand and use (DDU) from different sectors of society.

Description: Data demand and information use across multiple sectors. Source: MEASURE Evaluation

The geographic elements of the national data infrastructure—including the geographic data, systems, people, and policies—constitute what is known as the national spatial data infrastructure (NSDI).

For the majority of countries, the NSDI effort is a formal program led by a single agency, although it can also be led by a committee made up of multiple agencies.

The development of a strong NSDI is a key step in the process of evidence-based decision making based on high-quality data.

Ideas in Action

U.S. Federal Geographic Data Committee provides helpful resources for understanding and implementing an NSDI.

Global Spatial Data Infrastructure Association is an excellent source of information on NSDI programs around the world.

Highlights

For more information on data demand and use, visit the PEPFAR course Data Use for Program Managers. (Coming Soon!)

GEOGRAPHY AS STRATEGIC INFORMATION

Geography improves health decisions and accountability.

Incorporating the geographic context of human activities into the SI framework helps decision makers ***understand the influence of place on health program objectives***.

This promotes the implementation of place-based accounting and accountability, as knowing where things occur allows decision makers to identify costs and responsible parties associated with specific locations.

Geographically informed systems are necessary to coordinate and plan large-scale, multi-agency programs. Whether a crisis is acute, such as an outbreak of avian influenza or the aftermath of a tsunami or earthquake, or is chronic and long-term, such as HIV/AIDS, effective responses require the active management of geospatial information (Heard 2007).

Other examples of how geography can be used for strategic decision making:

- Country teams can use maps as a common framework for coordinating with multiple donors.
- Thinking spatially helps define what information is required at each geographic level of the administrative hierarchy within a country.
- Decision makers can use geographic identifiers to link datasets from complementary sources and map over- and under-served populations. This helps avoid duplicating efforts and wasting resources.

Description: Geography improves health decisions and accountability as part of the data demand and use (DDU) cycle. Source: MEASURE

Ideas in Action

U.S. Federal Geographic Data Committee provides helpful resources for understanding and implementing an NSDI.

Global Spatial Data Infrastructure Association is an excellent source of information on NSDI programs around the world.

PRIORITIZING BASED ON GEOGRAPHIC EVIDENCE

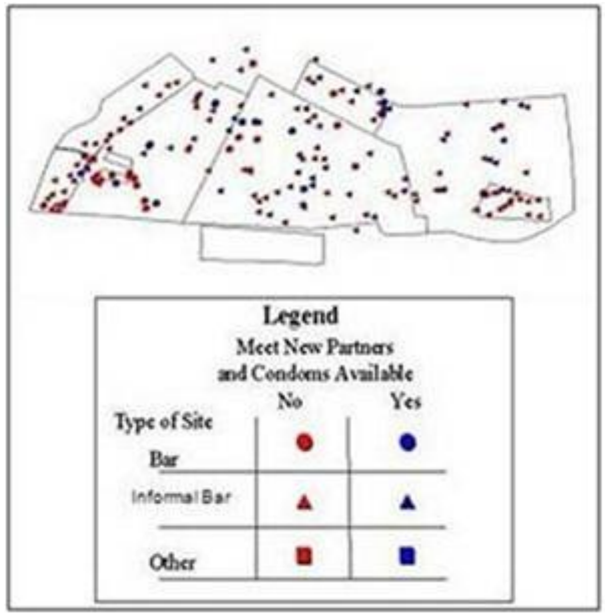
Geographic evidence can help define priorities.

Geographic evidence is intrinsic to decision making for public health, as ***populations affected by a disease cannot be identified, prioritized, and targeted for interventions without knowing where they are***.

In the context of HIV/AIDS, health officials often place the highest priority on preventing and treating disease in most at-risk populations (MARPs).

These include female sex workers, clients of female sex workers, injecting drug users, and men who have sex with men, as well as women and youths from the general population (UNAIDS 2008).

A proven method for identifying the geographic location of MARPs and HIV transmission networks, which generates strategic information for interventions, is the Priorities for Local AIDS Control Efforts (PLACE) protocol developed by MEASURE Evaluation.

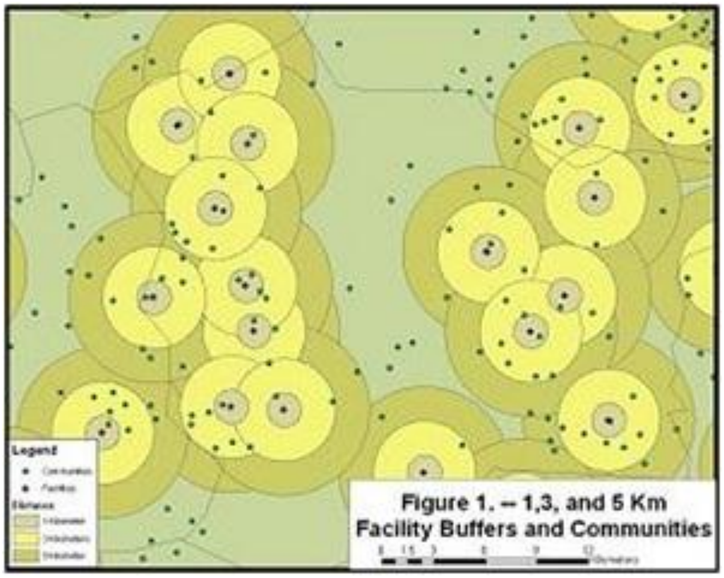


Description: PLACE map of most-at-risk populations (MARPs) in a township in South Africa. Source: MEASURE Evaluation

Did you know?

Evidence-based decision making for public health should adhere to some key practices. For more information, see Brownson et al., 2009

ASSESSING PHYSICAL ACCESS TO HEALTH SERVICES



Description: Facility buffers in Nicaragua, 2007. Source: MEASURE Evaluation

Access to health services can be analyzed using a GIS.
Physical access to health services differs based on one's geographic location within a country.

Using a GIS, these differences can be quantified and analyzed to inform planning efforts to improve services. *Note: Physical access to a health facility does not guarantee the availability of health workers, services, or medications to treat a patient.*

One of the GIS techniques for analyzing physical access is to (a) define distance thresholds from health facilities for covered versus non-covered populations and to (b) calculate straight-line (or Euclidean) distance between those populations and facilities (Noor et al. 2003; Noor et al. 2004).

One can also augment straight-line distance measurements with additional physical access indicators, such as facility size and proximity, or characteristics of the facility and population (Rosero-Bixby 2003). These studies based on straight-line distances have produced some valuable lessons:

- The likelihood that a person will choose a particular health facility decreases as the distance to it increases. This effect is more pronounced in rural settings in comparison to urban. Usage drops off significantly at a certain distance.
- Private sector providers and facilities are difficult for a ministry of health (MoH) to regulate and are generally under-reported on official lists. This creates blind spots in the knowledge base.

Ideas in Action

The kernel density estimation (KDE) approach allows researchers to estimate geographic access to health services more realistically than using straight-line distance calculations within artificial administrative boundaries.

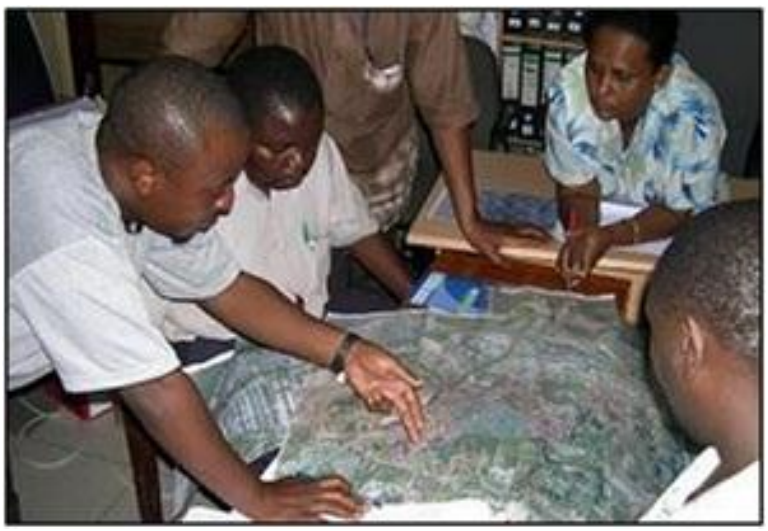
Did you know?

"The location of facilities in relation to population, other services and poverty or disease vulnerability underpins all efforts to achieve an equitable health service reform." (*Source*: Noor et al., 2004)

- Maps allow areas of “good” and “bad” physical access to be pinpointed, even when these areas are mixed.
- A GIS allows effective geographic targeting of interventions.

Caution: Estimating physical access using straight-line distances between populations and facilities is a simplistic approach. More advanced models take into account the transportation network, the varied nature of the landscape, and the modes of travel used by patients.

PRIORITIZING BASED ON GEOGRAPHIC EVIDENCE



Description: Participatory mapping to control malaria in Dar es Salaam, Tanzania. Source: Dongus et al. International Journal of Health Geographics 2007, 6:37.

Participatory mapping engages the community and deepens the understanding of target areas and populations.

Strategic information can be mapped at all geographic levels for which geographic identifiers are available, from national to sub-national to local.

In the context of public health, participatory mapping at the local level is a form of strategic mapping that engages the community in the identification of places and populations likely to be involved in the transmission of disease.

Participatory mapping is extraordinarily helpful to health planners, and is becoming an integral component of SI for understanding local areas and targeting interventions.

Did you know?

Participatory mapping is defined by:

- The community-based process of production
- A product that represents the agenda of the community
- The local knowledge and information depicted in the resulting maps

Participatory mapping is not defined by:

- Adherence to formal cartographic conventions (Source: IFAD, 2009)

Various terms used to refer to participatory mapping:

- Participatory mapping
- Participatory GIS
- Public participation GIS
- Counter mapping
- Community mapping

(Sources: IFAD 2009 and PPGIS 2010)

Dongus et al. 2007, for example, illustrated how participatory mapping was used to achieve a 100 percent inventory of mosquito larval habitats in Dar es Salaam, Tanzania. A key finding of the research was the cost-effectiveness of this method, as it required only a minimal level of technical skills and equipment.

The participatory process can enhance the planning process for an intervention:

- Participatory methods can significantly augment the knowledge base obtained from census data and GIS maps alone.
- Community-specific information allows researchers to see beyond the “neat, clear boundaries” between communities recorded in official documents.
- Resolving mismatches between local place names and official government census names reduces confusion and restores community confidence in the research.
- Community involvement allows researchers to identify a range of available health services that otherwise might go undetected.

KNOWLEDGE RECAP

1. A mature national data infrastructure is formed by interlocking cycles of what?
    a. Strategic planning
    b. Data demand and use
    c. Resource allocation
    d. Thematic mapping
    e. None of the above

2. What is the term used to refer to a strategic information system employed to collect, process, store, disseminate, and use *geographic data* for decision making?
    a. Health management information system (HMIS)
    b. Strategic information system (SIS)
    c. Routine health information system (RHIS)
    d. Geographic information system (GIS)
    e. None of the above

3. What is the term used to refer to *all* data available to national-level decision makers, as well as the people, policies, and systems required to collect, store, manage, analyze, and disseminate the data for decision making purposes?
    a. Health management information system (HMIS)
    b. Routine health information system (RHIS)
    c. National data infrastructure
    d. National spatial data infrastructure (NSDI)
    e. None of the above

4. Geographic data and tools allow users to perform which of the following?
    a. Link datasets from complementary sources based on geographic identifiers
    b. Generate maps to identify which populations are over- or under-served
    c. Identify information needed at each level of the administrative hierarchy to plan and execute interventions
    d. A and C only
    e. All of the above

5. What is the term used to refer to the geographic data available to national-level decision makers, as well as the people, policies, and systems required to collect, store, manage, analyze, and disseminate the data for decision making purposes?
    a. Health management information system (HMIS)
    b. Routine health information system (RHIS)
    c. National data infrastructure
    d. National spatial data infrastructure (NSDI)
    e. None of the above

KNOWLEDGE RECAP: ANSWERS

1. A mature national data infrastructure is formed by interlocking cycles of what?

    The correct answer is "b". The national data infrastructure, which can be defined as all of the data and systems available to decision makers within a country—as well as the people and policies needed to maintain them—is nourished by interlocking cycles of data demand and use (DDU) from different sectors of society. Source: MEASURE Evaluation, April 2010.

2. What is the term used to refer to a strategic information system employed to collect, process, store, disseminate, and use *geographic data* for decision making?

    The correct answer is "d". An SIS used to collect, store, manage, analyze, display, and disseminate geographic data (also referred to as spatial or geospatial data) is known as a geographic information system (GIS). A GIS is a powerful tool for modern public health research.

3. What is the term used to refer to *all* data available to national-level decision makers, as well as the people, policies, and systems required to collect, store, manage, analyze, and disseminate the data for decision making purposes?

    The correct answer is "c". In simple terms, the national data infrastructure consists of the data available to national-level decision makers and the people, policies, and systems required to collect, store, manage, analyze, and disseminate the data for decision making purposes. Source: MEASURE Evaluation, April 2010.

4. Geographic data and tools allow users to perform which of the following?

    The correct answer is "e". Geographic data and tools allow users to: Link datasets from complementary sources based on geographic identifiers; Generate maps to identify which populations are over- or under-served; Identify information needed at each level of the administrative hierarchy to plan and execute intervention.

5. What is the term used to refer to the geographic data available to national-level decision makers, as well as the people, policies, and systems required to collect, store, manage, analyze, and disseminate the data for decision making purposes?

    The correct answer is "d". The "technologies, policies, and people necessary to promote sharing of geospatial data throughout all levels of government, the private and non-profit sectors, and the academic community. The goal of this Infrastructure is to reduce duplication of effort among agencies, improve quality and reduce costs related to geographic information, to make geographic data more accessible to the public, to increase the benefits of using available data, and to establish key partnerships with [stakeholders] to increase data availability." Source: Federal Geographic Data Committee, www.fgdc.gov/nsdi/nsdi.html, April 2010.

Geographic data can be mapped.

Data become [geographic](#) by assigning them a location in relation to the surface of the Earth. In practical terms, this is accomplished using [geographic identifiers](#), which are pieces of information that specify the physical location of something.

Description: Geographic identifiers based on latitude and longitude. Source: MEASURE Evaluation (imagery from Google Earth)

By associating entities of interest (e.g., hospitals, clinics, households) and their attribute data (e.g., name, type, services provided) with a physical location, geographic identifiers allow data to be displayed on a map, and to be analyzed based on location-specific characteristics.

There are many types of geographic identifiers, including names or codes for administrative divisions (e.g., provinces, districts, communes); names or codes for human settlements (e.g., cities, villages, neighborhoods, informal settlements); or exact locations (e.g., street addresses or GPS coordinates).

Each distinct geographic entity (e.g., province, district, hospital) should be assigned a **[unique geographic identifier](#)**. This is vital for ***correctly linking geographic entities to their attribute data***, especially when the attribute data are derived from different sources.

WHAT MAKES DATA GEOGRAPHIC?

Ideas in Action

Geographic Identifiers: An Example from Kenya

Entity: Ministry of Public Health and Education

Street Address: Afya House, Cathedral Road

Postal code: 00100

City: Nairobi

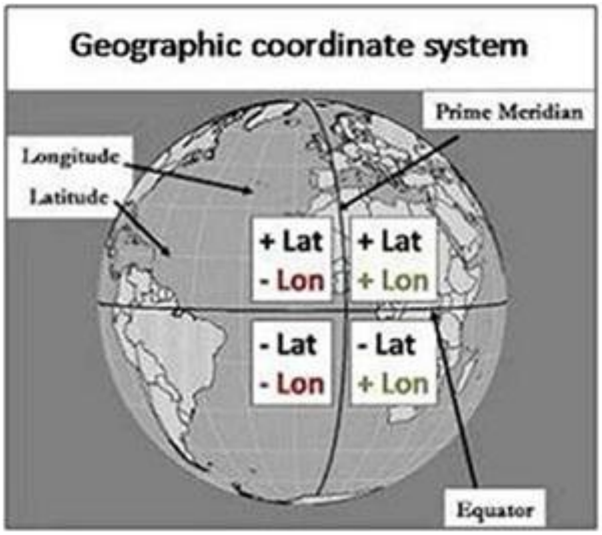
Province: Nairobi

Country: Kenya

Latitude/longitude:

1.291490 S, 36.815373 E

GEOGRAPHIC COORDINATE SYSTEMS



Description: Geographic coordinate systems represent locations on the spherical surface of the Earth. Source: MEASURE Evaluation

A geographic (unprojected) coordinate system.

A common geographic identifier is a combination of a latitude and a longitude.

Coordinates derived from latitude and longitude—as a result of being defined in relation to the more natural, three-dimensional surface of a globe—are considered to be "unprojected" or "geographic."

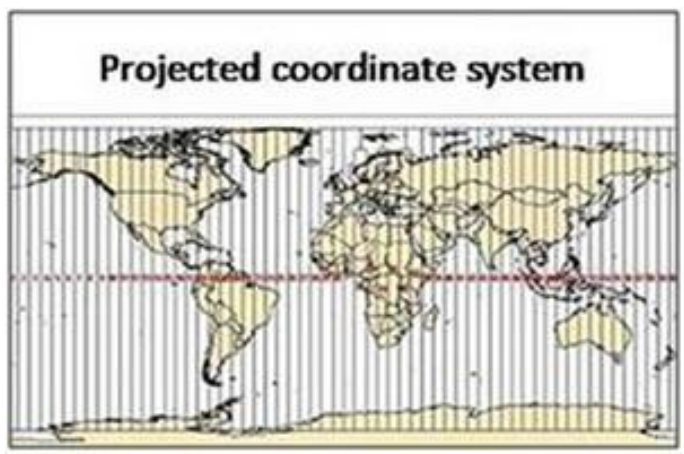
A coordinate system based on latitude and longitude is therefore known as a geographic coordinate system (GCS).

PROJECTED COORDINATE SYSTEMS

A projected coordinate system displays geographic data on a flat surface.

Although longitude and latitude can pinpoint exact locations on the surface of the Earth, they are ***not uniform units of measure*** around the globe (e.g., because lines of longitude converge at the poles, a degree of longitude ranges in length from about 111 kilometers at the equator to about half that at 60 degrees latitude).

As a result, a geographic coordinate system is excellent for original data capture and storage, but can be difficult to use to measure areas, perimeters, and distances accurately and consistently.



Description: Map of Universal Transverse Mercator (UTM) zones. Source: MEASURE Evaluation

This problem can be resolved by using a projected coordinate system or "projection," which transforms coordinates onto a flat, two-dimensional surface.

Before calculating geographic measurements, one should remember to "project first."

One of the most useful projections around the world is the Universal Transverse Mercator (UTM) coordinate system, as it uses the metric system and produces distance measurements that are accurate to within a meter (within one of its 60 zones).

Ideas in Action

Projections should be used based on which geographic properties they preserve:

- ***Conformal*** projections preserve local shapes.
- ***Equal-area*** projections preserve areas.
- ***Equidistant*** projections preserve distances between certain points.
- ***Azimuthal*** (or true-direction)

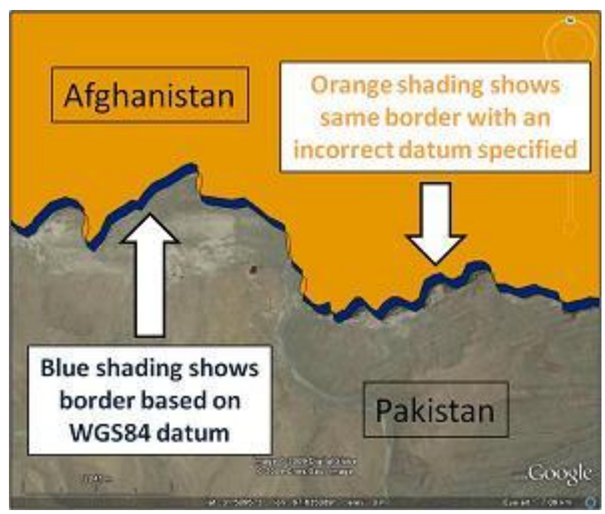
DATUMS

Specifying the wrong datum can cause a geographic dataset to be incorrectly represented.

Since the Earth is flattened at the poles, bulges at the equator, and has an irregular surface, it is not a perfect sphere.

As a result, specifying a geographic location with maximum accuracy also requires identification of a specific model of the Earth, known as a datum (e.g., the World Geodetic System 1984 or WGS 84).

This applies regardless of whether the coordinate system is geographic/unprojected or projected.



Description: "Datum shift" caused by specifying an incorrect datum for an international boundary. Source: MEASURE

Did you know?

The coordinate system and datum used to create a geographic dataset should be provided with the dataset as metadata.

For more information on metadata, see Session 3, page 7.

DATA FORMATS

Geographic reality can be represented in a GIS using two basic data formats, vector and raster.

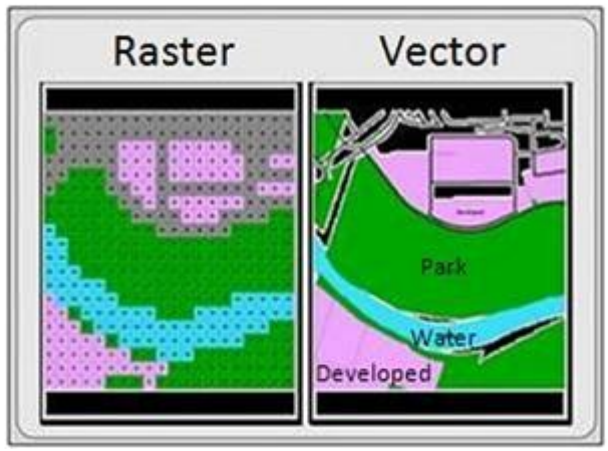
Geographic reality can be captured in a GIS as a collection of points, lines, or polygons, which is known as vector format. Or it can be stored in a grid in what is known as raster format.

> ***Did you know?***
>
> Raster datasets can be large, since a value needs to be stored for each grid cell. Vector format generally require much less storage

As examples of vector data, locations of health facilities could be stored in a GIS as **points**; roads which provide accessibility could be **lines**; and administrative divisions showing differences in population density could be **polygons**.

An example of raster data would be a satellite image. Such an image might be used to identify areas associated with mosquito breeding in order to help understand the geographic pattern of malaria prevalence.

Vector and raster data are analyzed using different sets of GIS tools. This will have an impact on the software package selected for analysis, as well as the technical expertise required for the analyst. For these reasons, it is important to understand the difference in the two formats.



Description: GIS data types: vector and raster. Graphic used under permissions of Creative Commons license. Source: GIS Commons

DATA SCHEMAS

A data schema is a description of how data in a computer database are organized into tables and fields, and identifies acceptable values for individual fields. A common way to capture a data schema is in a data dictionary

Using a proper data schema helps develop data that are standardized and complete, and that can be used to create accurate maps.

District Code	District	Province	Total Number Children Served
11	Getar	Alboma	388
12	Huma	Delmet	258
13	Jedanga	Delmet	267
22	Dumwick	Delmet	95
14	Flennet	Spudow	12
15	Pranan	Spudow	412
16	Spilitina	Spudow	426
17	Huma	Bronip	69
18	Star	Bronip	62
19	Trethel	Bronip	587

Description: Example of a geographic data table. Source: MEASURE Evaluation.

A good spatial data schema should require one record (one table row) per geographic unit with variables arranged in columns. To be mapped correctly, each geographic unit must have a well-formed geographic identifier, which means it should use a standardized spelling or code as defined by a national or international authority. The geographic identifier must also be unique.

The International Standards Organization (ISO) provides ISO 3166-1, which is a set of standardized identifier codes for countries, and ISO 3166-2, which is the complementary set of codes for country subdivisions. The United Nations publishes UN/LOCODE to provide standardized identifier codes for tens of thousands of locations used in trade and transport, such as ports, airports, and rail terminals. In addition, the United Nations publishes a free set of unique codes known as "SALB codes," which leverage the ISO codes and others, and which are available via the SALB project site. These sets of codes work together to identify unique geographic entities down to a detailed level.

During a crisis, the United Nations also develops unique geographic identifiers known as "P-codes" (place codes)—sometimes to a highly detailed level—to facilitate humanitarian efforts by the UN and others.

Following data schema best practices allows the accurate linking and mapping of datasets, which strengthens the national data infrastructure.

See the *Tip* below for more on ensuring unique geographic identifiers for data schema best practices for the table above.

TIP:

In the geographic data table shown on this page, each district appears once and only once in the table (there are two districts named Huma, but they are in different provinces). Each district also possesses a unique geographic identifier, which is provided in the District Code field. These two conditions adhere to data schema best practices.

Since the data in the table are arranged by district, the values in the column for Province (the larger geographic area) are not unique. To create a map by province, the district-level data would need to be summarized by province, and there would need to be a new table that contained a single record (row) for each province.

Ideas in Action

Creating unique geographic identifiers

In the geographic data table displayed on this page, district name "Huma" is not unique. This means district names alone cannot be used.

Two options:

- Use district codes
- Join district and province names (e.g., Huma:Delmet)

Did you Know?

Geographic coding systems that are standardized globally may not provide highly localized codes (e.g., for lower administrative divisions within a country). To obtain these codes, one might have to consult local central statistical agencies, ministries of health, or other authoritative offices.space and is considered more efficient.

METADATA

Dataset Title	Second Administrat the Congo
Theme Keywords	Democratic Republ
Dataset Topic Category	Boundaries
Geographic Location	Democratic Republ
Dataset Reference Date	20080703 (Dataset
Representativity	Start Date: March 2 End Date: April 200
Abstract	This ARC/INFO poly administrative bou
Supplemental Information In order to ensure a close m obtain compatibility with other standard medium resol with the international borders boundary database deve (International Boundaries Database, Version 1.0). Due t compiled, the SALB dataset is better adapted for thema spatial modeling. It is therefore recommended not to u scheme combines the ISO3 alpha code and a numerical	
Dataset Edition	First
Data Quality Comments Due to better adapted for thema	

Description: An example of metadata provided by the United Nations for second level administrative boundary files. Source: United Nations (www.unsalb.org)

Metadata provide vital information about geographic datasets.

Without knowing some basic information about a geographic dataset, the user will not be able to trust the validity of the data or even know if it is appropriate for the intended use. This basic information should be captured in the metadata.

At a minimum, the metadata should describe the following:

Source: Some sources are much more reliable than others. For example, the United Nations provides second-level administrative boundaries that have been vetted and approved by the national mapping agencies involved in their creation. To visit the site, click here.

Date(s) for which data are valid: The currency of spatial data is of critical importance, as data can become useless with the passage of time.

Coordinate system/projection and datum: This information enables one to align different spatial data layers accurately within a GIS. Also, different coordinate systems/projections can have different uses (See Session 3, page 3).

Scale: The scale at which a geographic dataset was created may impose limitations on its use.

File format: Geographic data come in a variety of file formats, such as shapefile format for vector data and MrSID format for a satellite image. One should verify that the file format is compatible with the GIS to be used.

Ideas in Action

Key Resources for Geographic Metadata Standards:

- ISO 19115
- U.S. Federal Geographic Data Committee

DATA SOURCES

Description: Primary and secondary spatial data sources. Source: MEASURE Evaluation

Spatial data can be obtained from a variety of sources, both primary and secondary.

Primary spatial data are *collected directly by the individual or organization conducting research.* Examples would be GPS coordinates or geographic identifiers (e.g., province names) collected during field surveys. New satellite imagery collected for a research project would be considered primary data.

Secondary spatial data are *collected from external sources* such as facility registries, MEASURE DHS, and other components of the NSDI such as the census, transportation data, and administrative unit boundaries. Satellite imagery collected by other organizations for other purposes would be considered secondary. Secondary data are available through such organizations as national mapping agencies (NMAs), other government agencies, NGOs, and universities. Before using secondary data, it is important to review the metadata (see session 3, page 7 on metadata).

Highlights

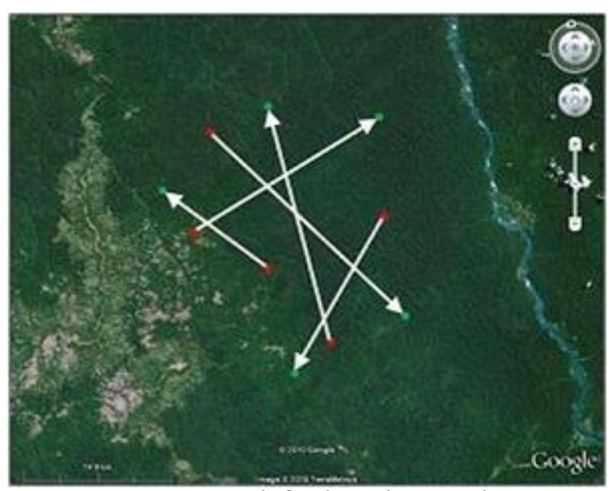
For a list of spatial data sources, visit https://www.cpc.unc.edu/measure/training/online-courses/my-courses/gis/files-and-links/gis%20data%20portals.pdf

MAINTAINING CONFIDENTIALITY, PART 1

To preserve confidentiality, respondent locations can be shifted randomly within a maximum allowable distance.

Geographic identifiers can allow an individual's location—and therefore identity—to be discovered. This can destroy data confidentiality and breach confidentiality agreements, a special concern when dealing with health data.

Some techniques for preserving the confidentiality of geographic data ***alter the spatial characteristics of the data***.



Description: Points shifted to obscure their true locations. Source: MEASURE Evaluation and Google Earth.

Examples:

1. Shifting point locations randomly within a maximum allowable distance to hide true locations.
2. Distorting the size, shape, or orientation of identifiable map features such as roads, rivers, and populated places.
3. Generalizing locations, such as putting survey respondents' latitudes into five-degree increments or rounding them to two decimal places.

Depending on the degree of geographic displacement, these methods have the potential to render meaningless the results of any analysis. Even though slight displacements may not significantly compromise results, these techniques should generally be employed only for publishing the results of analysis conducted using true locations.

Did you Know?

"Individual researchers must be concerned with the protection of confidentiality at all stages and in all types of research, including when they collect, disseminate, use, and read about data."

(*Source*: VanWey et al., 2005)

Ideas in Action

For more information on how to deal with issues concerning the confidentiality of geographic data, see this publication from MEASURE Evaluation: https://www.cpc.unc.edu/measure/training/ online-courses/my-courses/gis/files-and-links/wp-08-106_MEASURE _ confidentiality.pdf

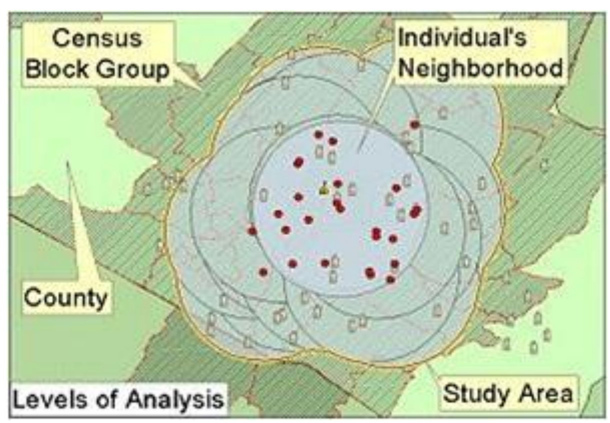
MAINTAINING CONFIDENTIALITY, PART 2

Omitting geographic identifiers from maps can obscure true locations without altering them.

There are many techniques available for maintaining the confidentiality of geographic data without having to modify the underlying spatial locations.

Examples include:

- Displaying spatial data only after falsifying or omitting names or other identifying information for map features such as roads, rivers, and populated places.
- Aggregating data from the individual level to more generalized spatial clusters or to



Description: Multiple levels of geographic analysis mapped without geographic identifiers in order to obscure true locations. Source:

administrative divisions, such as a districts or counties. ***Caution: Do not assume that aggregate data accurately represent particular individuals, which is known as the ecological fallacy.***

- Permitting access to spatial locations only for a select group of data developers and stripping personal and geographic identifiers from results provided to data users.
- Generalizing or rounding attribute data values from published data sources rather than using unaltered values that can be linked to geographic identifiers. This is especially useful for preventing deductive disclosure.

These are a few of the options for adhering to ethical standards and maintaining confidentiality without compromising the ability to conduct meaningful spatial analysis.

Ideas in Action

Before using a dataset that includes geographic identifiers, it is important to understand the quality of the geographic source data as well as how they were used to construct the dataset.

The first place to look for this information is in the metadata (See Session 3, Page 7).

KNOWLEDGE RECAP

1. How do data become geographic?

   a. Assign them a latitude and longitude
   b. Assign them a geographic identifier
   c. Use them to create a map
   d. A or B
   e. None of the above

2. Geographic identifiers make it possible to do what?

   a. Specify physical locations for entities such as hospitals, clinics, and households
   b. Associate physical locations with administrative divisions (e.g., provinces, districts)
   c. Link geographic entities with attribute data from multiple sources
   d. All of the above
   e. None of the above

3. A good spatial data schema does what?

   a. Describes how data in a computer database are organized into tables and fields
   b. Identifies acceptable values for individual fields
   c. Allows each geographic unit to have more than one record/observation
   d. A and B
   e. A and C

4. Metadata, which provide descriptive information about a geographic dataset, should include which of the following critical pieces of information?

   a. Name of person(s) who created the data
   b. Date(s) for which data are valid
   c. Coordinate system and datum
   d. B and C
   e. All of the above

5. What is the term used to refer to the protection of data and information that might identify individuals in a way that could cause harm or otherwise violate agreements made with them?

   a. Privacy
   b. Confidentiality
   c. Sensitivity
   d. Privilege
   e. None of the above

KNOWLEDGE RECAP: ANSWERS

1. How do data become geographic?

The correct answer is "d". Data become geographic by assigning them a location in relation to the surface of the Earth. In practical terms, this is accomplished using geographic identifiers. A set of latitude/longitude coordinates is a type of geographic identifier. A map cannot be created without first assigning geographic identifiers to the data to be mapped.

2. Geographic identifiers make it possible to do what?

The correct answer is "d". Geographic identifiers specify physical locations, which allow one to (a) identify the corresponding administrative division (e.g., province, state, or district) and to (b) link physical locations and their corresponding geographic entities with attribute data from different data sources.

3. A good spatial data schema does what?

The correct answer is "d". A data schema is a description of how data in a computer database are organized into tables and fields, and identifies acceptable values for individual fields. A proper data schema ensures that data are standardized and complete, and that they can be used to create accurate maps. A good spatial data schema should also require one record (one table row) per geographic unit with variables arranged in columns.

4. Metadata, which provide descriptive information about a geographic dataset, should include which of the following critical pieces of information?

The correct answer is "d". At a minimum, the metadata should describe the following: source, which is the organization that created the data; date(s) for which data are valid; coordinate system and datum; scale at which the dataset was created; and file format (e.g., shapefile format for vector data). Although potentially helpful, it is not necessary to know the names of the specific persons who created the data.

5. What is the term used to refer to the protection of data and information that might identify individuals in a way that could cause harm or otherwise violate agreements made with them?

The correct answer is "b". The term confidentiality refers to measures taken to prevent “the transmittal of personal information by someone other than the identified individual” (President’s Commission, 1971) in a way that could cause harm or otherwise violate agreements made with the identified individual (MEASURE Evaluation, 2008). “Individual researchers must be concerned with the protection of confidentiality at all stages and in all types of research, including when they collect, disseminate, use, and read about data” (VanWey et al., 2005).

TOOL COMPARISONS

There are many mapping software options available for purchase and many others which are free of charge. Each of these options requires a minimum amount of hardware on which to run.

When comparing mapping tools it is important to consider not only the initial cost of investment in the software, but also the necessary investment in equipment to run the software, and the time and manpower needed to learn and become proficient in their use.

Commercial packages can be expensive but provide a great deal of functionality and documentation. Free software often does not come with extensive documentation or training resources, and may be limited in features and capabilities.

As long as a mapping software package is easy to learn and use and has adequate documentation, simple features may be all the user needs.

If one is primarily interested in counting households within a certain distance of health facilities, for example, one would need a tool that searched within a radius but probably not a complex suite of tools for advanced processing of satellite imagery.

Description and Source: Logos from Google Earth, Quantum GIS, ESRI, and World Health Organization, the creator of HealthMapper.

Highlights

Online links to mapping software:

- DevInfo
- DIVA-GIS
- E2G
- EpiMap
- HealthMapper
- QGIS
- Google Earth
- ArcGIS

Highlights

For a comparison chart for mapping software options, visit: https://www.cpc.unc.edu/measure/training/online-courses/my-courses/gis/files-and-links/MappingSoftwareChart_v3.pdf

VIRTUAL GLOBES

Description: Eastern hemisphere as viewed in NASA's World Wind application. Source: NASA

Virtual Globes (also known as digital globes) are 3D representations of the Earth that provide the ability to zoom in and out through a wide variety of scales and to change viewing angle. They often combine satellite imagery collected at varying levels of detail with actual aerial or even street-level photography.

Features are often set to automatically label in a variety of ways according to the level of zoom or viewing angle. Quite a bit of this type of imagery is now available online, with several applications available.

> ***Did you know?***
>
> Sending sensitive locations to an online virtual globe may violate confidentiality agreements. To avoid this, one can download imagery from a virtual globe onto a local computer (cache the imagery), disconnect from the Internet, and then use the cached imagery for mapping.

The virtual globe with the highest user base and imagery library is currently Google Earth. It is constantly updating its imagery. It currently has the best detailed street data for Africa, even in rural areas. It also provides the richest supporting shared content like 3D buildings and photo links. Google Earth uses KML, which has been adopted as a standard by the Open Geospatial Consortium. KML files can be launched in Google Earth, sent as an attachment, posted on the Web, or embedded in a Web page.

Another globe that is freely available is World Wind by NASA. It is open source software and has elevation data as well as an extensive library of other types of satellite data, which can provide information on land use and vegetation as well as a variety of oceanic and atmospheric measurements over time cycles ranging from hours to years.

The newest available viewer/imagery base is Bing Maps, from Microsoft (formerly known as Microsoft Virtual Earth). It can operate in an Internet browser window when Microsoft Silverlight is installed. Bing Maps is also integrated into ArcGIS Explorer, which is available for no extra charge to organizations with an ArcGIS license. It also incorporates into some locations “birds’ eye views” (close-up photography displayed at a low viewing angle), and is also constantly updating its imagery.

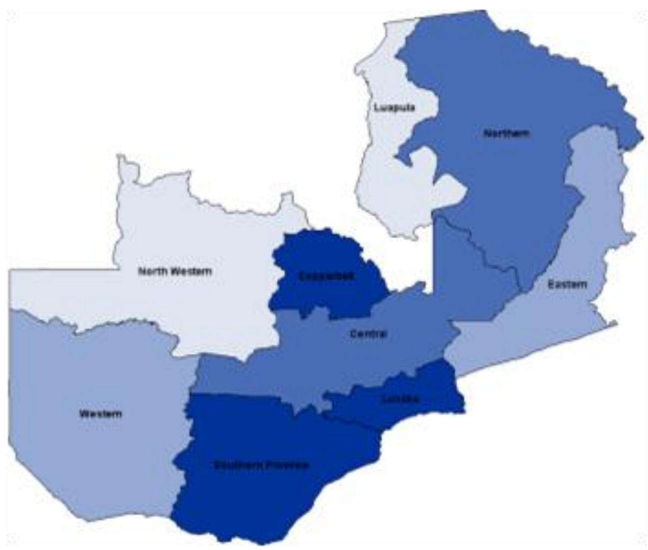
Limitations: For any of these virtual globes, a good Internet connection is necessary, at least for initial imagery download (see the “cache the imagery” link in the side bar). And a strong computer is required for the graphics-intensive displays. Keep in mind that imagery can be out-of-date or poor resolution for some areas. They are intended for data visualization only, not for analysis, and if administrative area boundaries are included they can be inaccurate or poorly detailed.

You decide

Virtual globes can serve as base information for any map. Depending on the location you are interested in, you may want to go online and compare imagery from each of the three main sources listed here. Each service has its own inventory of image information, which is constantly being updated for places all around the globe and results may vary at any given time.

THEMATIC MAPPING

Thematic mapping software allows the user to input data (often in spreadsheet form—click here for an example), link it to a map, and class the data for visual display, such as on a choropleth map, as shown here.



Description: Zambia: Percent population under age 18 living as orphans by province. Source: MEASURE Evaluation OVC project.

Examples of thematic mapping software available as free downloads from the web or for use on the web itself include:

HealthMapper—this tool is available from the World Health Organization and includes sample data for Mali. Good documentation is provided.

EpiMap—available from the Centers for Disease Control (CDC) and created specifically to work with EpiInfo, the CDC’s disease surveillance tool. EpiMap provides country administrative boundaries in shapefile format, although they are now somewhat dated.

Excel to Google Earth (E2G)—this tool is an Excel macro available from MEASURE Evaluation. It works with Excel spreadsheets and can display choropleth maps in Google Earth. Boundary files are included for the 15 original PEPFAR focus countries

Highlights

In limited resource settings, it is important to consider the **minimum hardware requirements** before selecting mapping or GIS software.

> ***Did you know?***
>
> Data Requirements for Thematic Mapping
>
> In order to create a thematic map you will need (1) geographic base map data for the points or areas to be mapped and (2) attribute data to display for these points or areas. The attribute data must contain geographic identifiers that allow the data to be joined to the base map.

plus the Democratic Republic of the Congo. There are accompanying tutorials in video and PDF formats.

GeoCommons—an online mapping service that offers an easy, step-by-step set of mapping tools, a rich set of base maps, and a professional interface. Output can be saved in KML, spreadsheet, or shapefile format. The site requires your data and maps be shared with the global community. If privacy issues are not a concern, this might be a viable option for creating and sharing maps over the web.

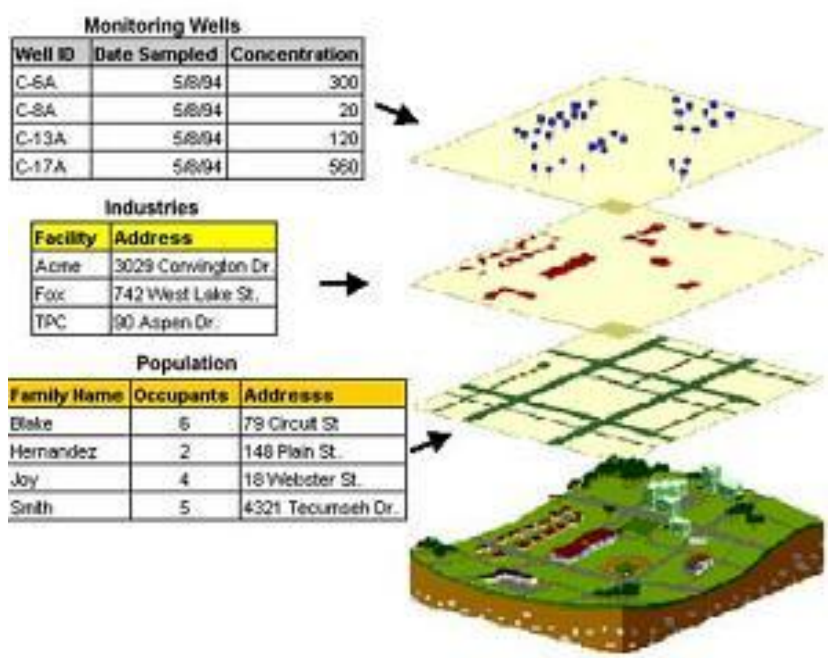
DevInfo—a data display software developed by UNICEF that contains a mapping module. It was originally created to work with the UN's Millenium Development Goals, but other data can be modified, with the help of the DevInfo support team, to work with the program.

These tools can all help you properly format your data for mapping. Geocodes in the data files must match geocodes in the base map files. If data is already properly formatted and geocoded, these programs are generally easy for the novice to use to produce simple thematic maps. These maps can then be shared with decision makers to shape and inform opinion.

GIS SOFTWARE

Spatial analysis, feature geometry editing, and higher level data manipulation functions can best be performed using GIS software.

There are many types of GIS software packages available. Some of these are commercial packages which are proprietary and quite expensive. Others are open source, which means they can be customized by the user. Although open source software may often be free for downloading, it can have hidden costs, associated with lack of documentation and training resources.



Description: A GIS can incorporate data from several sources using the common element of geographic location.

Examples of some of the many GIS software packages available include:

ArcGIS by ESRI—this is the oldest, most established, most comprehensive, and most widely used GIS software currently available. It has by far the best documentation and the most extensive help and support. Depending on how it is used, ArcGIS can also have a steep learning curve. Users wanting a free method of simply displaying their ArcGIS data can use ESRI's ArcExplorer.

MapInfo—is one of ESRI's biggest commercial competitors, sharing many of its most popular features. It is also proprietary (currently owned by Pitney Bowes), and is being marketed primarily for business use.

DIVA-GIS—has many of the most important and useful features of a basic GIS and is available for free download. However, it has limited output capabilities, and has been through several versions since the original basic manual was updated. It was originally developed for raster-based agricultural applications, but contains vector tools as well.

> ***You decide***
>
> A true GIS generally has a high learning curve and may provide more capability for geographic analysis than is required. Before deciding what software is needed it is important to determine the function that maps and spatial analysis will play in your organization's routine decision making.

> ***Highlights***
>
> GIS is a tool that uses geographic data to produce maps and to conduct spatial analysis. In limited resource settings, it is important to review the **minimum hardware requirements** before selecting GIS software.

Quantum GIS—this software is a good basic GIS package and is open source and free. It comes with an introductory GIS course which was developed for download or online use. It also comes with basic tutorials. It might be useful for beginners who are willing to invest time and energy in learning the software, or for experienced users who want a free but customizable GIS.

WORKING WITH THE SOFTWARE—DATA AND FILE FORMATS

The software discussed in this session generally involves the use of the following common types of file formats.

Common input formats:

- Excel data tables (.xls or .xlsx)
- Access database files (.mdb or .accdb)
- Shapefiles (includes .dbf and .shx and .shp files)
- KML (stands for Keyhole Markup Language) coordinate files which can be overlayed on satellite imagery in Google Earth

> ***Highlights***
>
> To import data into a GIS, it is important to organize the data into rows and columns. For more information, see Session 3, page 6.

- Comma-delimited or tab-delimited text files
- SPSS data tables
- SAS data tables

Common output formats:

- All of the above
- images such as JPEG or TIFF or PNG files (all types of images which can be used in other documents such as Word files and reports)
- PDF (portable document format) files
- Bitmapped (.bmp) image files

It is important that data to be used in a GIS be stored in rows and columns in a format that can be readily imported into the chosen software, such as in an Excel spreadsheet, Access database, or comma-delimited text file.

It is difficult to import and map data stored in a word processing or PDF document, even if the data are stored in rows and columns, as these formats cannot be easily read by a GIS.

INCORPORATING THE GEOGRAPHIC CONTEXT—GPS AS A DATA COLLECTION TOOL

A GPS receiver is a key device for collection of geographic locations.

The incorporation of the geographic aspect of any data needs to be considered from the start, when planning the data collection.
Geographic data collection will involve recording both the **location** and the **attributes** of the phenomena being studied.

As mentioned in Session 3, page 8 concerning primary data collection, a GPS can be used to initially record coordinate data when conducting a field survey.

Description: Garmin 72 GPS receiver with view of satellite position page showing location acquired in three dimensions (latitude, longitude, and elevation). Source: MEASURE Evaluation Global Positioning Toolkit.

GPS receivers record locations on the Earth with a high level of accuracy by receiving signals from satellites in space. Most GPS receivers are simple to operate; however, some care is needed to make sure that the points are as accurate as possible.

The exact process of data collection will vary based on the type of receiver and the data collection needs, however, there are important considerations in all uses of GPS when it comes to maintaining positional accuracy and reducing data errors. Click on the *Tip* at the bottom of the page for more information on obtaining positional accuracy during GPS data collection.

Once the GPS receiver has locked onto signals from the satellites, it will display the current location as a coordinate. There are many coordinate systems the receiver can use, but latitude and longitude is the most common.

Did you know?

There are 24 to 32 GPS satellites in orbit, and their position relative to any particular point on the Earth varies.

Highlights

For detailed instructions on using a Garmin 72 receiver to capture GPS data, see the MEASURE Evaluation Global Positioning System Toolkit.

TIP:

Factors to consider in obtaining the best positional accuracy during GPS data collection:

Arrangement of satellites—there needs to be a good spread ("constellation") in the arrangement of locational satellites; this arrangement will vary according to time of day. If only a few are available, or the available ones are all clustered in one area of the sky, the positional accuracy will be diminished.

Existence of a clear line of sight for the satellite signal path—buildings, trees, or even the body of the data collector himself can all serve to block the signal from one or more satellites, impeding the ability of the unit to read an accurate position.

KNOWLEDGE RECAP

1. Free GIS and mapping software packages lack sufficient functionality, documentation, and technical support to justify their use.Assign them a latitude and longitude

   a. True
   b. False

2. A fully-featured GIS can be distinguished from a simpler mapping tool by the inclusion of what capabilities:Specify physical locations for entities such as hospitals, clinics, and households

   a. The ability to output a map.
   b. The ability to input data.
   c. The ability to create customized modules for specific types of analysis.
   d. The ability to create a scale bar and legend.
   e. All of the above.

3. Geographic identifiers are only important when inputting GPS data.Describes how data in a computer database are organized into tables and fields

   a. True
   b. False

4. Open source GIS software is readily available, free of charge, and always comes with extensive documentation.

   a. True
   b. False

5. Which of the following is a strength of virtual globes?

   a. Imagery is always up-to-date.
   b. A good Internet connection is not required.
   c. They are generally capable of sophisticated analysis of data.
   d. Administrative area boundaries provided are usually highly detailed and accurate.
   e. None of the above.

KNOWLEDGE RECAP: ANSWERS

1. Free GIS and mapping software packages lack sufficient functionality, documentation, and technical support to justify their use.Assign them a latitude and longitude

The correct answer is "b". There are several free GIS and mapping software options that may be able to produce the maps and other spatial products needed by a public health program. More details are provided in Session 3, Geographic Tools.

2. A fully-featured GIS can be distinguished from a simpler mapping tool by the inclusion of what capabilities:

The correct answer is "c". Other types of mapping software can often input data and output simple maps, including those with scale bars and legends, but generally speaking only a full-featured GIS can perform more complicated data analysis.

3. Geographic identifiers are only important when inputting GPS data.

The correct answer is "b". Geographic identifiers are the best way to keep track of any geographic data, including polygon (area) names. Data to which geographic identifiers have been added is called "geocoded". To facilitate mapping and analysis, a good data schema will allow for one geocoded record per geographic unit. If a data file for a country contained both a population figure and a geographic identifier for each province, and the geographic identifiers in the data file matched those in a map file, it would allow for the population data to be mapped. A choropleth map showing population by province could then be produced from the geocoded data.

4. Metadata, which provide descriptive information about a geographic dataset, should include which of the following critical pieces of information?

The correct answer is "b". It is true that open-source software is becoming more readily available and is often free, but documentation is often scarce or not subject to updates along with the software. Effective use may require willingness by the user to experiment and even to seek out online forums for help.

5. Which of the following is a strength of virtual globes?

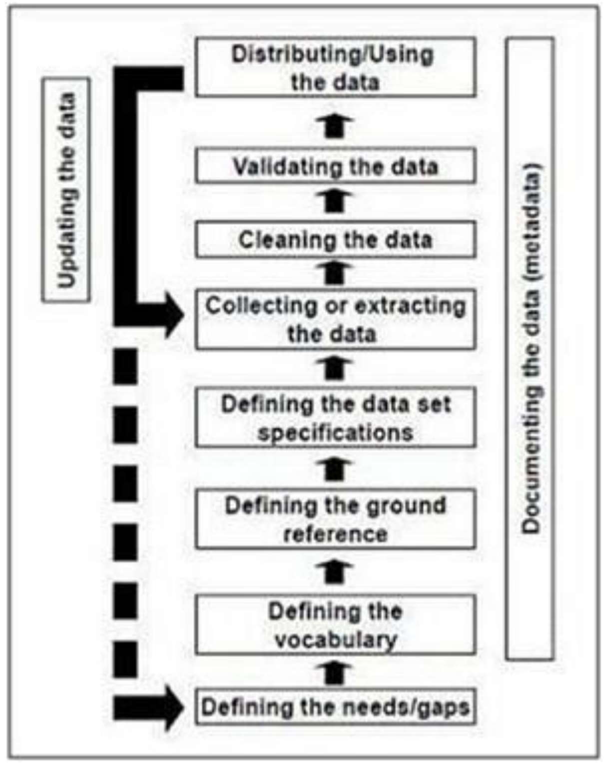
The correct answer is "e". Imagery may be out-of-date because of the time and expense required to obtain it. A good Internet connection is required to access most virtual globes, at least for the initial download and caching of imagery on a local computer. Virtual globes are intended more for display of geographic data than for analysis. The national organizations that create and maintain administrative boundaries generally do not provide detailed, accurate versions of these boundaries to external organizations such as virtual globe developers.

INCORPORATING GEOGRAPHIC DATA INTO THE DECISION-MAKING PROCESS

The first step in the decision-making process is to identify which decisions need to be made. These decisions may involve how best to identify target populations, strengthen health systems, or monitor and evaluate disease interventions. Identifying these decisions will help define the need for geographic data, which is the first step in the GIS data production chain.

Defining the need for geographic data would include the following questions:

- What specific questions need to be answered in order to make a decision?
- What geographic data is needed to answer those questions, and what geographic data is already available versus will need to be collected?
- What is the process for cleaning and validating the data, and who will be responsible for this process?
- What geographic products, such as maps and graphs, will be needed to help guide the decision-making process?
- What type of analysis is required, and what software is available for this purpose?
- How will the geographic products be presented and/or distributed?



Description: Garmin 72 GPS receiver with view of satellite position page showing location acquired in three dimensions (latitude, longitude, and elevation). Source:

Description: Entering malaria survey data in Angola. Source: 2007, Alfredo L. Fort, courtesy of Photoshare. (latitude, longitude, and elevation). Source: MEASURE Evaluation Global Positioning Toolkit.

IDENTIFYING THE RESOURCES

Resources can include data, equipment, and expertise.

An important step in the process of using geographic data and tools is to identify the resources available to you.

Did you know?

Note: Unless specified, some contractors deliver graphics files but withhold data. To avoid this, contractor deliverables should include maps in the preferred format (JPG, PDF, etc.), the spatial data used to create those maps, and the appropriate metadata.

Data are available either internally or externally. External data sources can include other partners, external consultants, academic institutions, or local/regional data centers. Such data centers and academic institutions are often major components in the NSDI. Internal data can be available in-house or collected in the field (for example, new GPS data).

The most common GIS data formats and software programs are detailed in Session 4.

Equipment refers to the actual hardware needed. This will include computers to run the mapping software and store the data, GPS units to collect new point-level data, and possibly printers and backup storage devices.

Expertise refers to the skills and knowledge necessary to use spatial tools. It also may be found internally or externally. If external technical specialists are to be hired, it will be helpful to have at least some internal staff that is able to communicate with them on a technical level. Training may be necessary to elevate the level of internal expertise.

Highlights

When to consider hiring external contractors:

- Project requires skills not available internally
- Project goals are well-defined
- Face-to-face meetings do not need to be frequent
- Privacy/security not an issue or can be worked around
- Project has a short timeline with a clear deadline

DATA COLLECTION AND REVIEW

Depending on whether you have decided to collect your geographic data as part of a field survey or to use existing data, you will need a variety of personnel and equipment.

If you are **collecting new data**, you will need a team that is trained in the use of GPS and you will need GPS-enabled equipment. This may mean that (1) you choose to use laptops with a GPS card or that (2) you choose to use hand-held GPS machines and later download the data to a computer.

Source/Description: 1996 CCP, Courtesy of Photoshare. A field worker interviews a mother about child health needs in Peru.

Important considerations for newly collected GPS data:

- **Review** and "**clean**" the data. If you are using a higher-end GPS receiver, you may be able to use post-processing software, such as Trimble Navigation's Pathfinder Office, to obtain better positional accuracy. Even if using a less expensive recreational receiver, you should check points for obvious errors such as whether they occur outside the collection area or have been recorded or saved improperly.
- Make note of collection or recording **errors**.
- Remember that **backups** are important!

If you are using **existing data** you may not have the ability to clean the data but you may need to:

- **Obtain agreements** from the original data collectors in order to use it for your purposes.
- **Reformat** the data.
- **Convert** file types to work with your software.
- Check the existence of any **metadata** associated with the secondary data files. Historical metadata needs to be incorporated into any new metadata once you have used the data for new purposes so that future users will have a history of changes.

Did you know?

Most GIS programs require GPS data to be imported in decimal degrees.

Two key benefits of setting the GPS receiver to collect data in this format:

- Saves a great deal of time when preparing data for use in a GIS
- Facilitates merging of data from different data collection efforts

Did you know?

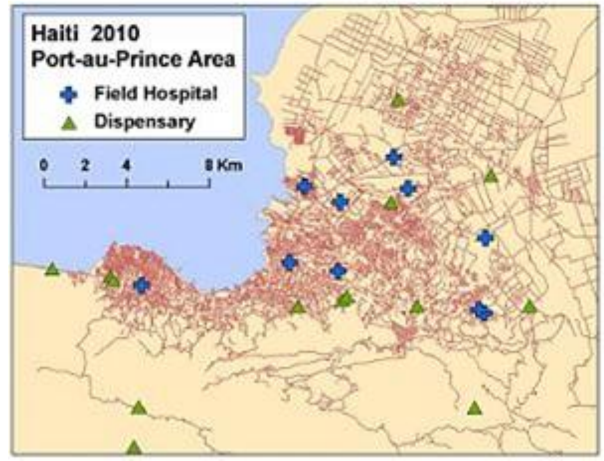
Reversals in the x,y coordinates are a common error when the coordinates are hand-typed.

GIS AS A DECISION SUPPORT TOOL

Decision support tools will need to be identified, produced, and evaluated to make evidence-based decisions.

Examples of the ways in which GIS can be used as a decision-support tool include:

Mapping of points or "spot mapping": Simple point locations can be shown along with relevant background information such as political boundaries and roads, and possibly land cover or surface features such as mountains or rivers. This can often be done using freely available tools such as Google Earth or ESRI's Map Explorer tool. Such a map can show patterns in the data at a glance, such as location of HIV treatment centers clustering along major transportation routes.



Description: Haiti: Location of field hospitals and dispensaries displayed on road map of Port-au-Prince area. Source: USAID in January 2010 (data based on the 2005 List of Health Institutions from the Ministère de la Santé Publique et de la Population).

Thematic mapping: Classifying data according to standardized analytical methods or determining whether two complementing variables can be shown together on a map by use of more than one type of symbol. A choropleth map is one of the most common types of thematic map.

Spatial analysis: A more in depth type of data analysis involving spatial statistics and allowing predictions and hypotheses about the data, generally performed in a more robust GIS program.

GIS AS A DECISION SUPPORT TOOL—THEMATIC MAPPING

There are **two main considerations** in planning how to display data on a thematic map.

1. **How to classify the data**—data are generally classified according to standardized analytical methods such as natural breaks, quantiles, or equal intervals, and then displayed using color or symbol gradations. Usually, large values are represented by a dark shade or a large symbol and smaller values are de-emphasized using a proportionally lighter shade or a smaller symbol.

Highlights

These are the **most common types of thematic maps:**

- Choropleth
- Proportional symbol
- Isarithmic
- Dot density

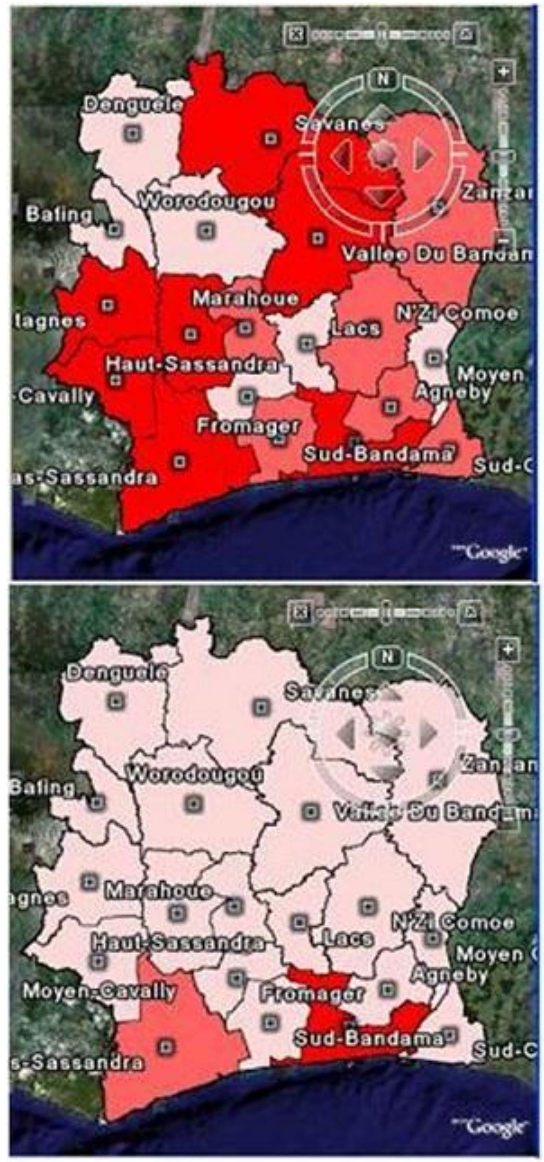
For a guide to choosing map colors, please see the Color Brewer.

2. **How to represent the data**—two complementary variables may be shown together on a map by combining the use of several types of symbols. For instance, shaded areas and pie charts or graphs. This can be a good way of combining data from more than one source.

When preparing data to be displayed on a thematic map, it is also important to consider the use of ratios rather than raw counts, especially when the areas to be shaded are of varying sizes.

> ***You decide***
>
> These two maps of Côte d'Ivoire on the left both show the same population data, but use different methods of data classification. The first uses quantiles and the second uses equal intervals.
>
> Which map emphasizes the dense coastal population? Which map better shows balanced groups of high, medium, and low population?

For example, a choropleth map representing population density (a ratio of population count to land area) will allow geographic areas of different sizes to be compared more readily than one representing simple population counts. Likewise, one showing percent orphans by province should facilitate province-level comparisons more easily than one showing total numbers of orphans for the same areas.



Population by province, Côte d'Ivoire, using two different data classification methods: quantiles and equal intervals. Source: MEASURE Evaluation.

For more information about normalizing data for choropleth mapping, visit:
http://www.gsd.harvard.edu/gis/manual/normalize/

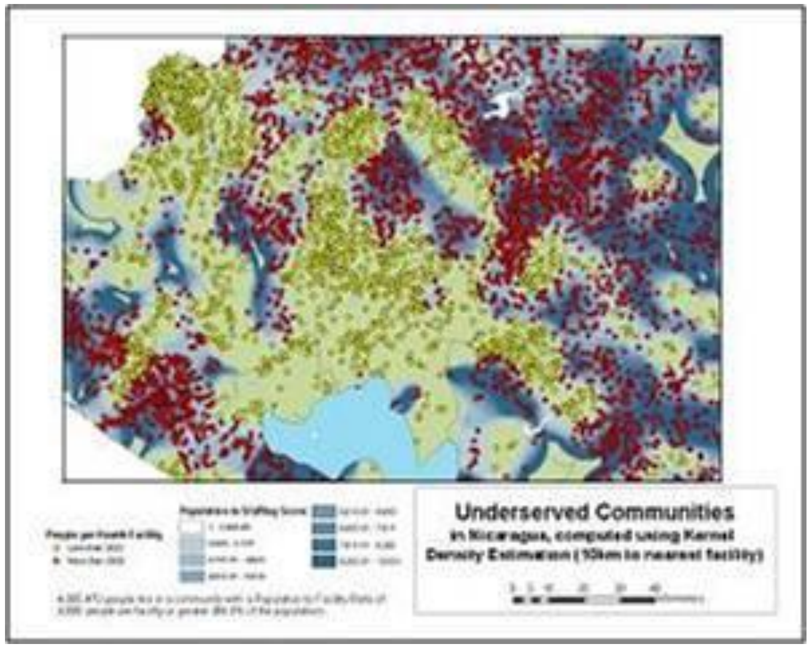
For an excellent article on data classification for choropleth mapping, visit:
http://www.migration.directionsmag.com/article.php?article_id=718

GIS AS A DECISION SUPPORT TOOL—SPATIAL ANALYSIS

One type of spatial data analysis that can be performed using a GIS is kernel density estimation.

One of the simplest types of **spatial data analysis** and one that is often available in mapping software is the buffer.

For example, a one mile buffer on either side of a river can be shown to contain a certain number of villages that may be subject to water-borne illnesses such as cholera.



Description: Kernel density estimation as used to determine underserved communities in Nicaragua. Source: MEASURE Evaluation.

Ideas in action

Spatial Analysis Examples

- Buffer, spatial autocorrelation,and kriging: See the case studies in the next session.
- Kernel density estimation and availability of health services: Spencer and Angeles 2007
- Network analysis and contraceptive choice: Entwisle et al. 1997

Another similar but slightly more complex way of measuring these types of phenomena (such as accessibility to a point location) is kernel density estimation. This is a way of weighting certain points statistically to discover areas that are more or less heavily influenced by them.

Network analysis is a way of creating travel routes and times or determining accessibility, using road networks (line-based vector data).

Spatial statistics such as spatial autocorrelation (a measure of the degree to which nearby points are similar to each other, or how much spatial clustering exists in the data) and kriging, which is a form of spatial interpolation, can give added insights into the influences and patterns operating in geographic data.

ONGOING CONSIDERATIONS

Geographic data is a useful input for decision making and can be used to guide policy and strategically prioritize program funding.

Maps can be an effective means for showing and sharing data. They can bring to light issues previously unnoticed. They can also clarify trends in the data. In order to ensure that the maps are being used and to track how they are being used, a system should be put in place, if possible, for tracking who gets these maps.

Description: Kernel density estimation as used to determine underserved communities in Nicaragua. Source: MEASURE Evaluation.

The data demand and use cycle (see session 2, page 2) illustrates how use of the data will lead to more questions and more data collection demand. Subsequent data collection processes will both update and expand the data infrastructure.

Geographic data can be used to generate strategic information and need to be readily shared. Other agencies may be able to use the data collected for a specific project in other unforeseen ways. Using a good data schema and a common data format can help ensure this happens.
And finally, if data is to be collected on a routine basis, a system needs to be put in place to enable regular updates with minimum effort. There also needs to be maintenance of the data reporting systems themselves. Software and hardware require routine updating and maintenance, and personnel must be continually informed of and trained in how to perform updates, so that the data demand and use cycle can continue.

KNOWLEDGE RECAP

1. Factors to consider in using existing data:
    a. Existence of metadata
    b. Permissions and use agreements
    c. Formatting and data schema
    d. Timeliness of data
    e. All of the above

2. A simple point map that shows relevant background information can be an effective decision-support tool.
    a. True
    b. False

3. Population totals by province are a good type of data to use for a choropleth map.
    a. True
    b. False

4. The method used to classify data for a thematic map can have a big impact on the way the resulting map looks.
    a. True
    b. False

5. Data that has been collected by GPS is always free of errorsImagery is always up-to-date.
    f. True
    g. False

KNOWLEDGE RECAP: ANSWERS

1. Factors to consider in using existing data:

   The correct answer is "e".

2. A simple point map that shows relevant background information can be an effective decision-support tool.

   The correct answer is "a". Base map information should be relevant, however, and can include administrative boundaries, roads, rivers and other information such as land use or rainfall records.

3. Population totals by province are a good type of data to use for a choropleth map.

   The correct answer is "b". Population density would generally be considered a better candidate for choropleth mapping, especially for a map that involves a variation in the size of the areas represented. For example, take two provinces in a given country which each have the same total population. Both of these provinces, when classed according to total population, will fall into the same category. But a larger, more sparsely populated area will show up the same color as a small, densely populated area if raw numbers are used. If population per square kilometer is used, the sparsely populated area will show up a lighter color than the densely populated area, which will be a better visual representation of reality.

4. The method used to classify data for a thematic map can have a big impact on the way the resulting map looks.

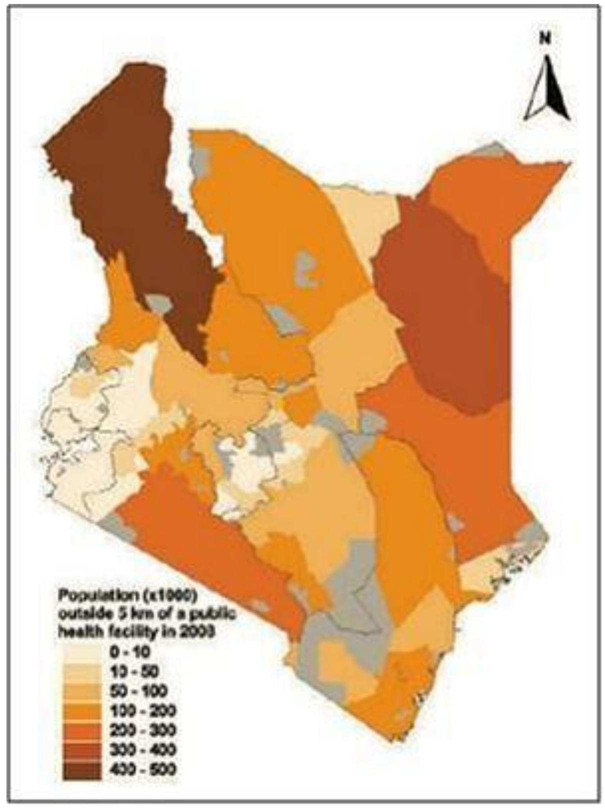
   The correct answer is "a". For example, equal interval classification can result in empty classes (and thus missing colors on the map) if the data are highly skewed. For more on this issue see "Choropleth Mapping with Exploratory Data Analysis".

5. Data that has been collected by GPS is always free of errorsImagery is always up-to-date.

   The correct answer is "b". Positional accuracy can vary depending on the number of satellites overhead at a particular time, the position of those satellites in the sky, or nearby obstructions that might block or redirect the signal. Post-processing of the data (also known as "data cleaning", and often requiring additional software) is frequently required. Writing down GPS coordinates and entering them into a computer by hand can result in transcription errors.

MAPPING ACCESS TO HEALTH CARE SERVICES IN KENYA

Mapping access to health care services can greatly benefit national planning efforts.
Source: Noor et al. 2009.



Description: Areas in 2008 within 5 km benchmark for access recommended by the Kenyan Ministry of Health.

OVERVIEW

Noor et al. (2009) documents an effort in 2008 to develop a comprehensive database of public health service providers in Kenya to facilitate comparison with a national inventory from 2003. The end goal was to provide a foundation for evaluating changes in geographic access to health care services.

GEOGRAPHIC METHODS

1. *Georeferenced facilities:* Assigned a unique geographic identifier to every verifiable public health service provider using locations derived from a combination of GPS coordinates, 1:50,000 topographic maps, hand-drawn maps, and Google Earth. Organized facilities by type of service(s) and by sector.

2. *Mapped population:* Using a population density map for 1999 from the Malaria Atlas Project (MAP), which is provided in raster format using 100 m × 100 m pixels, re-projected population density to 2003 and 2008 population counts using provincial inter-censal growth rates.

3. *Calculated distances:* Computed Euclidean (straight-line) distances from each health facility to each population pixel for both 2003 and 2008 using ArcGIS 9.2 from ESRI.

4. *Classified population map:* Identified population pixels and summed corresponding population counts for 2003 and 2008 based on whether they were (a) within 5 km or (b) greater than 5 km from a public health facility.

RESULTS

- Demonstrated that it is possible to develop a relatively complete national inventory of spatially referenced health facilities to identify areas of the country where access to health facilities has improved versus where it needs to be improved.
- Developed a spatially referenced inventory of health facilities that can be linked with other spatially referenced datasets to facilitate national disease risk modeling and health commodity distribution planning.

Did you know?

For health information systems to function well, ***a reliable inventory of health service providers*** is critical.

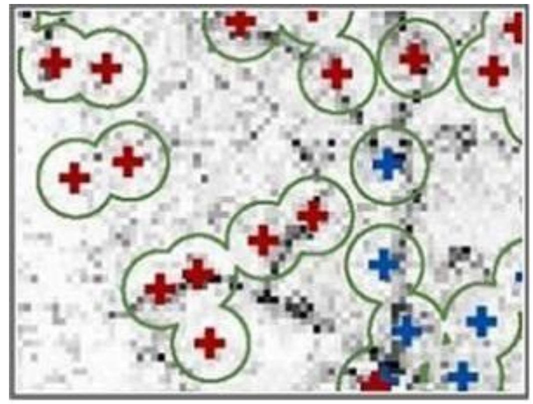
Highlights

The ***spatial referencing of service providers*** to allow their representation in a GIS is vital if the full planning potential of such data is to be realized.

CAUTION

- *Physical access to a health facility does not guarantee the availability of health workers, services, or medications to treat a patient.*
- *Estimating physical access using straight-line distances between populations and facilities is a simplistic approach. More advanced models take into account the transportation network, the varied nature of the landscape, and the modes of travel used by patients.*

ESTIMATING GEOGRAPHIC COVERAGE OF ART SERVICES IN RWANDA



Description: 5 km radius drawn around facilities in Rwanda that offered antiretroviral therapy (ART) services supported by PEPFAR or the Global Fund to Fights AIDS, Tuberculosis and Malaria.

The buffer tool can be used to estimate geographic coverage of health services. Source: U.S. Department of State, Humanitarian Information Unit 2009.

OVERVIEW

In the literature on geographic accessibility to health services, researchers often draw circular buffers around facilities using the estimated catchment area radius to divide population according to whether it is inside or outside the catchment. While a crude method for identifying catchment areas, the buffer tool is common to virtually all GIS software and provides a rough estimate of which parts of a country lack service coverage, as it is reasonable to assume that opportunity to access a service is greater within buffer boundaries than without. This information can be used to estimate how much coverage would increase according to different service expansion scenarios.

GEOGRAPHIC METHODS

The buffer layer can be combined with population data in raster format to quantify the percent of the total population that is within the buffer. Such an analysis requires us to combine the buffers with population data using a zonal statistics tool. The zonal statistics tool analyzes the raster cells that fall within a specified area (in this case, the buffers serve as zones) and calculates a variety of statistics on those cells. Calculating zonal statistics requires an advanced GIS software package.

In the figure above, a buffer with a 5 kilometer radius was drawn around facilities in Rwanda that offered antiretroviral treatment (ART) supported by PEPFAR or the Global Fund to Fights AIDS, Tuberculosis and Malaria. The buffer was then displayed over population density, with dark cells signifying high population density and white cells signifying low density.

Did you know?

Limitations of the circular buffer tool:

- Assumes catchment areas are simple and uniform in size and shape
- Population data does not account for spatial variation of disease prevalence
- Should always be used with the best available data regarding distribution of a disease

Did you know?

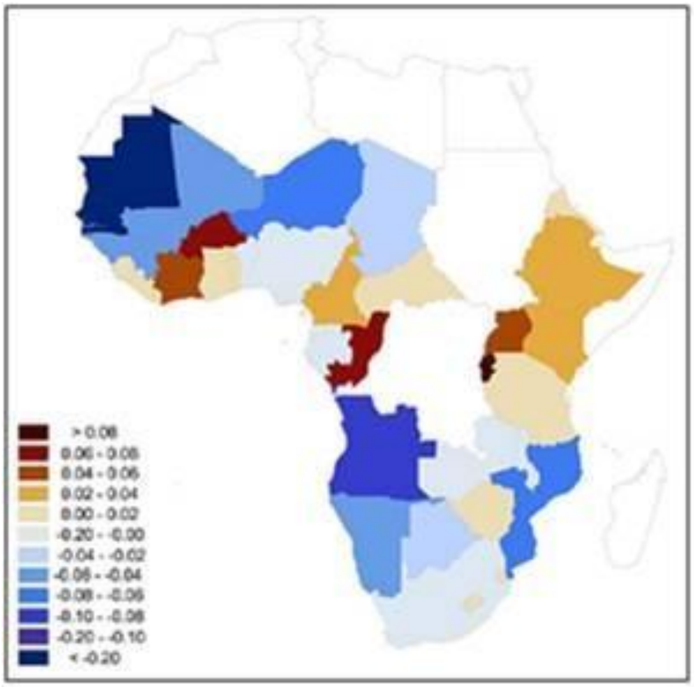
Minimum data required to estimate geographic coverage with the circular buffer tool:

- Latitude and longitude of service location(s)
- Type of service(s) delivered at the facility
- Population dataset in raster format

RESULTS

A simple visual inspection suggests where population may lack access to ART services.

MONITORING TB/HIV CO-EPIDEMICS IN SUB-SAHARAN AFRICA



Description: Change in TB incidence per 100,000 for 2000–2006 relative to change in HIV prevalence for 1997–2003.

SOURCE: Sánchez et al. 2010.

OVERVIEW

Tuberculosis (TB) ranks among the most deadly and prevalent re-emerging infections of people living with HIV/AIDS (PLWHA). In the last 20 years, the number of new TB cases has tripled in high HIV prevalence countries, and at least 33% of the world's 33.2 million PLWHA are infected with TB. Furthermore, drug-resistant TB can be more prevalent and virulent in PLWHA. Approximately 80% of persons with TB/HIV coinfections live in sub-Saharan Africa, where TB is the leading cause of death among PLWHA.

GEOGRAPHIC METHODS

1. *Calculate indicator:* Calculate $R_{TB/HIV}$, which is an indicator that encompasses the ratio of two measures: $\mathbf{R_{TB}}$ quantifies the mean change in TB incidence over a defined time period and $\mathbf{R_{HIV}}$ quantifies the mean change in HIV prevalence over an earlier time period. This can be done at any geographic level for which the data are available.
2. *Map indicator:* Map the indicator and conduct a visual assessment of the results:
    - $\mathbf{R_{TB/HIV} = 0}$ if TB and HIV rates are same
    - $\mathbf{R_{TB/HIV}}$ **is positive** if (a) TB incidence increasing faster than HIV prevalence or (b) TB increases and HIV decreases or (c) both rates are decreasing but TB incidence is decreasing at lower rate

Ideas in Action

Geographic considerations for monitoring co-epidemics:

- Indicators should be comparable between geographical areas within a country and between countries
- Data should be broken down and reported by the smallest administrative unit possible

(Source: WHO, 2009)

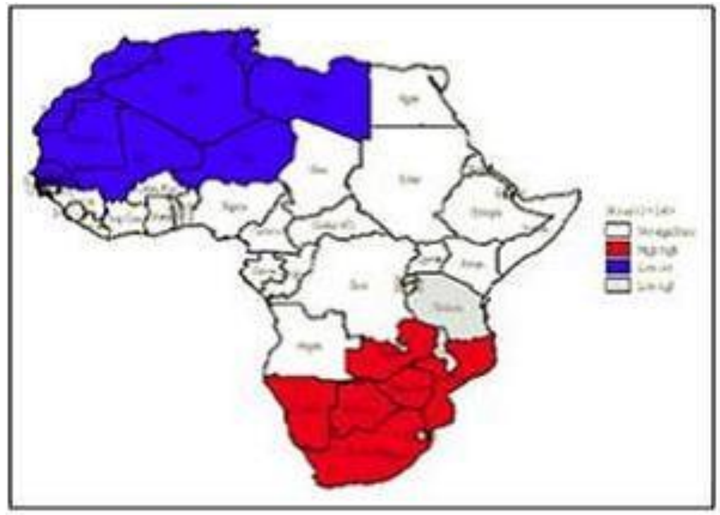
- $R_{TB/HIV}$ **is negative** if inverse conditions exist in relation to the preceding case (where $R_{TB/HIV}$ is positive).

RESULTS

Indices such as $R_{TB/HIV}$ can be calculated using joint analysis of data gathered regularly and independently by monitoring agencies. Simple indicators specifically designed to integrate information from closely linked diseases can allow monitoring of co-epidemics over time, and are powerful tools for spatial comparisons. **If used as a routine monitoring tool, the $R_{TB/HIV}$ indicator could allow public health officials to maximize the use of existing data by evaluating a single number.**

MONITORING TB/HIV CO-EPIDEMICS IN SUB-SAHARAN AFRICA

High TB-HIV deaths with similar neighbors (shown in red) versus low TB-HIV deaths with similar neighbors (shown in blue).
SOURCE: Uthman et al. 2009.



Description: Local Indicator of Spatial Association (LISA) cluster map for annualized average TB-HIV deaths in Africa, 1991 to 2006

OVERVIEW

The challenge facing TB and HIV control programs in Africa is that the disease burden is not homogenous but varies geographically. Minimizing the risk of TB-HIV deaths can be assisted by recognizing its geographical and temporal distribution and identifying areas of unusually high death rates. Testing for spatial autocorrelation is a valuable geographic technique for identifying clusters or "hot spots" of disease.

GEOGRAPHIC METHODS

- Calculated average TB-HIV deaths per 100,000 for each administrative region for each year of the 16-year period 1991 to 2006.
- Conducted spatial rate smoothing and grouped countries into categories based on quartiles.
- Applied a likelihood function to test for elevated risk within each country versus outside of each country.
- Calculated global and local Moran's I values of spatial autocorrelation to test for statistically significant clustering of TB-HIV deaths.
- Performed exploratory spatial data analysis (ESDA) using GeoDa software to identify spatial patterns and to generate hypotheses based on those patterns.

RESULTS

- Spatial distribution of TB-HIV deaths in Africa is non-random, with statistically significant test results found for clustering.
- ESDA allowed pinpointing of geographic areas with higher risk, as well as assessment of temporal variability of risk areas. This provided a working hypothesis on risk of TB-HIV deaths versus environmental factors.

Highlights

The Kulldorff spatial scan statistic can be used to identify geographic clusters of TB incidence that are higher or lower than expected at a local level. (Source: Houlihan et al., 2010)

Ideas in Action

Spatial autocorrelation is one of four methods of geographic analysis used by the USAID-funded Deliver Project to identify and resolve supply chain management issues related to stocks of health commodities (e.g., contraceptives). (Source: Deliver Project, 2009)

COMBATING MALARIA IN BURKINA FASO WITH REMOTE SENSING

High-resolution satellite imagery allows the identification of populated areas at high and low risk for malaria.

Source: Dambach et al. 2009.

OVERVIEW

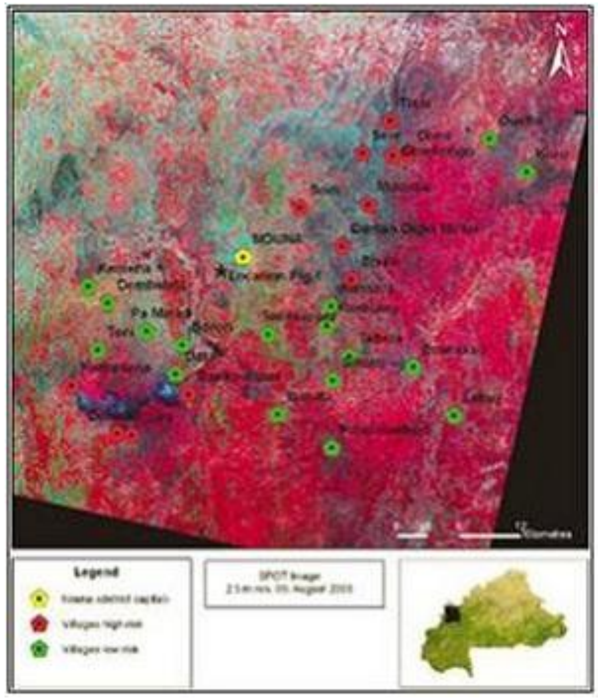
In 2008, the authors conducted a remote sensing study in the malaria endemic lowlands of northwestern Burkina Faso to see whether high spatial resolution satellite imagery could be used to distinguish between mosquito larvae habitats that posed different levels of risk for local malaria transmission.

GEOGRAPHIC METHODS

1. *Acquired high-resolution satellite imagery:* Obtained orthorectified, multispectral SPOT 5 images to coincide with late rainy season (1 Sep 2008).
2. *Conducted field work:* Executed six-week field phase using handheld GPS devices to collect 45 ground truth points. Developed classification scheme for land cover based on field observations.
3. *Assigned risk categories:* Based on a literature review, assigned land cover classes to risk categories (low, medium, high, and very high) in relation to the expected presence of mosquito larvae.
4. *Classified imagery:* Implemented a supervised classification of the satellite image based on the land cover classification scheme.
5. *Identified high-risk areas:* Calculated percent of land cover of high and very high risk within 500 meters of 30 villages.

RESULTS

Villages with the greatest percentages of high-risk land cover types could readily be identified using SPOT 5 satellite imagery. This allowed for spatial targeting of populations at most risk of malaria.



Description: Villages in Burkina Faso with similar land cover risk in their 500 m buffer zones based on the potential to contain breeding habitats for Anopheles mosquitoes.

> ***Ideas in Action***
>
> **Imagery Options**
>
> A study in Kenya in 2006 compared three types of satellite imagery for correctly identifying habitats associated with malaria-causing mosquito larvae. IKONOS imagery was best, followed by aerial photography and Landsat TM7.
> (*Source*: Mushinzimana et al., 2006)

> ***Ideas in Action***
>
> **Mapping Malaria with Existing Data**
>
> Kriging offers an alternative to remote sensing for creating a national map of malaria prevalence based on point-based data one may already have.
> (*Source*: Noor et al. 2008, Noor et al. 2009)

REDUCING DIARRHEA IN BENIN

Geographic deviation from the national average of DHS cluster-level data on diarrhea prevalence.
Source: Pande et al. 2008.

OVERVIEW

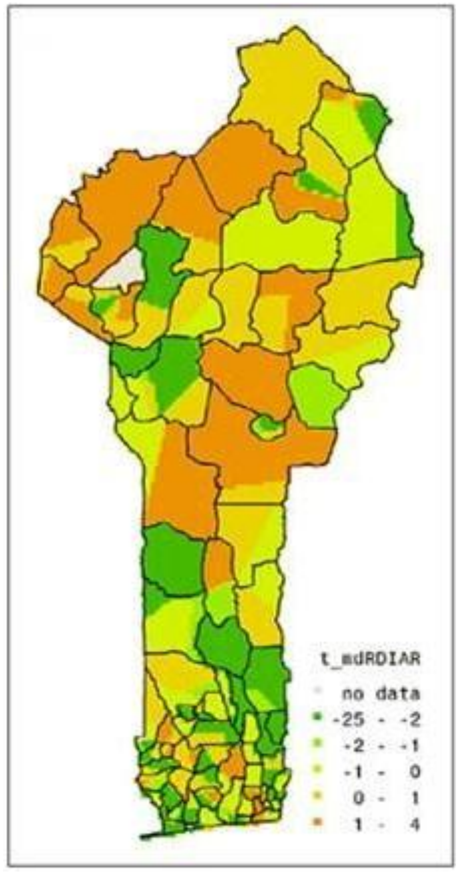
Reducing diarrhea prevalence is linked to the Millennium Development Goals (MDGs) on child mortality and environmental sustainability. Diarrhea prevalence is only partly due to lack of safe drinking water, as social factors also play a role. Pande et al. (2008) analyzed diarrhea prevalence based on conditions within households as well as external, geographical factors in the Oueme River Basin of Benin.

GEOGRAPHIC METHODS

1. *Created household variables:* Used Demographic and Health Surveys (DHS) cluster-level data to create household-specific variables for such things as diarrhea prevalence, mother's education, hygiene conditions, and distance to hospital.
2. *Created environmental variables:* Constructed variables related to water availability and quality, such as mean annual rainfall and a groundwater quality stress indicator.
3. *Mapped:* Mapped diarrhea prevalence and covariates at different levels of geographic summarization (clusters and municipalities) and developed hypotheses based on visual inspection.
4. *Analyzed socioeconomic and environmental influences separately:* Assessed variance of diarrhea prevalence and its covariates both within and between cluster locations to see the influences of socioeconomic and environmental factors, respectively.
5. *Developed combined statistical model:* Estimated a mixed effect logit model of diarrhea prevalence to evaluate the significance of household-level socio-economic variables and cluster-level environmental variables.

RESULTS

The logit model highlighted the significant adverse impact on diarrhea prevalence caused by poor access to water, but indicated that the intensity of the effect depended on household-level socioeconomic conditions. The conclusion was that households suffering from worse socioeconomic conditions will experience higher rates of diarrhea prevalence, even at the same geographic location.



Description:Spatial deviation of cluster-level diarrhea prevalence from the national average.

Ideas in action

Data recorded at DHS cluster locations can be used to estimate values for an entire country using Thiessen polygons.)

Highlights

In this case, improvement in access to safe water on its own would not significantly reduce diarrhea prevalence, as ***social factors also played an important role***

MAPPING DENGUE IN NICARAGUA

Neighborhood data collection maps incorporated free satellite imagery from Google Earth.

Source: Chang et al. 2009.

Description: During routine epidemiologic surveillance, public health workers mark 8.5 x 11 inch black and white maps like this to locate potential larval development sites.

OVERVIEW

The authors evaluated Google Earth as a source of high-resolution imagery for GIS base maps for dengue surveillance and control efforts in the limited resource setting of Bluefields, Nicaragua

METHODS

1. *Created Google Earth base map:*
   - Developed high-resolution mosaic of satellite imagery from Google Earth.
   - Georeferenced base map using GPS control points.
2. *Overlaid GIS layers:*
   - Neighborhood boundaries.
   - Indices of larval infestations by neighborhood.
   - GPS points for dengue fever case households.
   - Point locations for potential larval sites based on combination of GPS points and hand-drawn locations on satellite image maps.
3. *Performed visual inspection:* Examined spatial patterns of known dengue case locations in relation to potential larval development sites and neighborhood-based larval indices.

RESULTS

- Pinpointed areas of greatest risk to prioritize where to send limited supplies of larvicide, insecticide, and human resources.
- Used maps very effectively to communicate with the central Ministry of Health regarding daily progress of interventions.

Did you know?

Google Earth has been used in the Democratic Republic of the Congo to track the polio virus down the Congo River. This allowed mobile populations, who were missed by previous routine immunization services, to be identified and vaccinated.
(*Source*: Kamadjeu, 2009)

KNOWLEDGE RECAP

1. The development of a comprehensive, spatially referenced national inventory of health facilities will provide the foundation for which of the following?
    a. Assignment of unique geographic identifiers to help pinpoint locations of health service facilities
    b. Evaluation of geographic access to health facilities
    c. Linking of the health facility inventory to other spatially referenced datasets
    d. All of the above
    e. None of the above

2. It is possible to construct a simple indicator of TB incidence versus HIV prevalence and map it to monitor changes in TB/HIV co-epidemics over time.
    a. True
    b. False

3. High spatial resolution satellite imagery, such as IKONOS or SPOT 5, can be used to identify local areas that are likely to provide breeding habitats for mosquito larvae, which is helpful in combating the transmission of malaria.
    a. True
    b. False

4. In the absence of spatial analysis and planning at the national level, local and international organizations may unintentionally cluster in urban or select rural areas, leaving other parts of the country with inadequate coverage for health services.
    a. True
    b. False

5. What are some benefits of using the buffer tool to estimate geographic coverage of health services?
    a. Takes into account the often irregular size and shape of catchment areas
    b. Population data in raster format provide a highly accurate representation of spatial variation of disease prevalence within a geographic area
    c. Can be used reliably with any dataset containing disease distribution information
    d. All of the above
    e. None of the above

KNOWLEDGE RECAP: ANSWERS

1. The development of a comprehensive, spatially referenced national inventory of health facilities will provide the foundation for which of the following?

   The correct answer is "d"

2. It is possible to construct a simple indicator of TB incidence versus HIV prevalence and map it to monitor changes in TB/HIV co-epidemics over time.

   The correct answer is "a". Sánchez et al., 2010 demonstrated the feasibility of constructing a simple indicator from data that are regularly gathered by monitoring organizations. It uses the ratio of TB incidence for a time period to HIV prevalence for an earlier time period. This indicator can be mapped at any geographic level for which the data are available, and can be assessed visually in relation to three possible results: RTB/HIV = 0 if TB and HIV rates are same; RTB/HIV is positive if (a) TB incidence increasing faster than HIV prevalence or (b) TB increases and HIV decreases or (c) both rates are decreasing but TB incidence is decreasing at lower rate; and RTB/HIV is negative if inverse conditions exist in relation to the preceding case (where RTB/HIV is positive)

3. High spatial resolution satellite imagery, such as IKONOS or SPOT 5, can be used to identify local areas that are likely to provide breeding habitats for mosquito larvae, which is helpful in combating the transmission of malaria.

   The correct answer is "a". Several researchers have proven the feasibility of using high-resolution satellite imagery to identify geographic areas that are likely to provide suitable habitats for mosquito larvae. Dambach et al. (2009), for example, found that villages in northwestern Burkina Faso at greatest risk of malaria could readily be identified using SPOT 5 high spatial resolution satellite imagery. Mushinzimana et al. (2006) conducted a study in Kenya in 2006 to compare the ability of three types of imagery to identify habitats associated with anopheline mosquito larvae. They found that high-resolution IKONOS imagery allowed 89% of habitats to be classified correctly.

4. In the absence of spatial analysis and planning at the national level, local and international organizations may unintentionally cluster in urban or select rural areas, leaving other parts of the country with inadequate coverage for health services.

   The correct answer is "a".

5. What are some benefits of using the buffer tool to estimate geographic coverage of health services?

   The correct answer is "e". When using the buffer tool, one should consider the following limitations: Assumes catchment areas are simple and uniform in size and shape, when they are often irregular.; Population data do not account for spatial variation of disease prevalence. In addition, population data in raster format are aggregated into pixels, which generalizes the data even more.; To provide the most accurate picture possible for a given geographic area, the buffer tool should always be used with the best available disease distribution data for that area.

1. Thematic mapping software can help you class and color shade your data.
   a. True
   b. False

2. A mature national data infrastructure is formed by interlocking cycles of what?
   a. Strategic planning
   b. Data demand and use
   c. Resource allocation
   d. Thematic mapping
   e. None of the above

3. A primary national-level step in the process of making health decisions based on geographic evidence is development of a strong what?
   a. Health management information system (HMIS)
   b. Routine health information system (RHIS)
   c. Geographic information system (GIS)
   d. National spatial data infrastructure (NSDI)
   e. None of the above

4. GIS studies of access to health services have produced all but which of the following lessons?
   a. The likelihood that a person will choose a particular health facility decreases as the distance to it increases.
   b. Usage of services drops off significantly at a certain distance from a facility.
   c. Proximity to a highway affects access.
   d. Access measures differ based on whether analysis is conducted using straight-line (Euclidean) or transportation network distances.
   e. A GIS does not permit effective geographic targeting of interventions to improve access.

5. Participatory mapping is expensive, but is warranted based on the high quality of results.
   a. True
   b. False

6. In what key way(s) does geography strengthen the national data infrastructure?
   a. Creates more visually pleasing maps
   b. Allows linking of datasets from different sources
   c. Promotes standardization of data for sharing
   d. A and C
   e. B and C

7. A strong national data infrastructure strengthens data at multiple scales, from local to national to international, and makes effective public health decision making more likely.
   a. True
   b. False

8. Almost all human activity occurs on the surface of the Earth. What essential information does this provide for public health analysts and decision makers to help ask better questions and plan interventions to improve human health?
   a. Land cover
   b. Elevation
   c. Income, education, and occupation status
   d. Geographic locations
   e. None of the above

9. What is the term used to refer to the ***geographic*** data available to national-level decision makers, as well as the people, policies, and systems required to collect, store, manage, analyze, and disseminate the data for decision making purposes?
   a. Health management information system (HMIS)
   b. Routine health information system (RHIS)
   c. National data infrastructure
   d. National spatial data infrastructure (NSDI)
   e. None of the above

10. Geographic identifiers make it possible to do what?
   a. Specify physical locations for entities such as hospitals, clinics, and households
   b. Associate physical locations with administrative divisions (e.g., provinces, districts)
   c. Link geographic entities with attribute data from multiple sources
   d. All of the above
   e. None of the above

11. How are geographic and projected coordinate systems different?
    a. Geographic coordinate systems are three-dimensional
    b. Projected coordinate systems are two-dimensional
    c. Geographic coordinate systems make it easier to measure areas, shapes, and distances accurately
    d. A and B only
    e. All of the above

12. What is a good source for standardized unique geographic identifiers?
    a. ISO 3166-1
    b. ISO 3166-2
    c. UN/LOCODE
    d. A and B only
    e. A, B, and C

13. What is a good way to maintain the confidentiality of geographic data?
    a. Shifting point locations using a small, uniform distance and uniform direction
    b. Publishing and distributing maps made from the data subject to no restrictions
    c. Avoiding generalization of spatial information (e.g., rounding of latitudes) so that true locations will be available to users of the data
    d. Including all personal and geographic identifiers in data provided to users
    e. None of the above

14. Which of the following is a strength of virtual globes?
    a. Imagery is always up-to-date.
    b. A good Internet connection is not required.
    c. They are generally capable of sophisticated analysis of data.
    d. Administrative area boundaries provided are usually highly detailed and accurate.
    e. None of the above.

15. A fully-featured GIS can be distinguished from a simpler mapping tool by the inclusion of what capabilities:
    a. The ability to output a map.
    b. The ability to input data.
    c. The ability to create customized modules for specific types of analysis.
    d. The ability to create a scale bar and legend.
    e. All of the above

16. The most common coordinate system used by most GPS units is UTM.
   a. True
   b. False

17. Identify the example(s) of external data:
   a. Data available in-house.
   b. Data from academic institutions.
   c. Data from local or regional data centers.
   d. Field data.
   e. B and C

18. The term spatial analysis generally refers to more complicated GIS operations such as the creation of predicted data points or data surfaces, and often involves the use of spatial statistics.
   a. True
   b. False

19. Which of the following is NOT a type of thematic map:
   a. A highway map
   b. A choropleth map
   c. A dot-density map
   d. An isarithmic map
   e. None of the above (all are thematic)

20. Spatial autocorrelation refers to:
   a. Tendency of data to predict event occurrence in time.
   b. Automatic sorting of data according to location.
   c. Tendency of data to cluster together spatially.
   d. Ability to predict data at a point based on knowledge of nearby points.
   e. C & D

21. The development of a comprehensive, spatially referenced national inventory of health facilities will provide the foundation for which of the following?
   a. Assignment of unique geographic identifiers to help pinpoint locations of health service facilities
   b. Evaluation of geographic access to health facilities
   c. Linking of the health facility inventory to other spatially referenced datasets
   d. All of the above
   e. None of the above

22. What is the benefit of calculating global or local Moran's I values of spatial autocorrelation?

a. Eliminates the need to conduct exploratory spatial data analysis (ESDA).
b. Creates box plots and histograms to generate a graphical summary for evaluating spatial dependence of observations.
c. Provides a test for whether values are clustered by location, which can help identify "hot spots" for a disease.
d. All of the above
e. None of the above

23. What method of geographic analysis can be employed to construct a highly accurate, national map of malaria prevalence using data attached to georeferenced survey locations, such as health care facilities, that are distributed throughout a country?

a. Supervised classification
b. Kriging
c. Ground truthing
d. All of the above
e. None of the above

24. What geographic technique can be used with data attached to point-based survey locations to create a polygon-based map for an entire country?

a. Buffer polygons
b. Ground truthing
c. Thiessen polygons
d. All of the above
e. None of the above

25. In addition to a coordinate system, what else must be specified to identify a location with maximum accuracy?

a. Projection
b. Datum
c. Zone
d. Geographic identifier
e. None of the above

26. Open source GIS software is readily available, free of charge, and always comes with extensive documentation.

a. True
b. False

1. Thematic mapping software can help you class and color shade your data.

   The correct answer is "a". A choropleth map is one type of thematic map.

2. A mature national data infrastructure is formed by interlocking cycles of what?

   The correct answer is "b". The national data infrastructure, which can be defined as all of the data and systems available to decision makers within a country—as well as the people and policies needed to maintain them—is nourished by interlocking cycles of data demand and use (DDU) from different sectors of society. *Source*: MEASURE Evaluation, April 2010.

3. A primary national-level step in the process of making health decisions based on geographic evidence is development of a strong what?

   The correct answer is "d". While a strong HMIS, RHIS, or GIS can all facilitate decision making based on geographic evidence, they depend on a strong national spatial data infrastructure (NSDI) as a prerequisite.

4. GIS studies of access to health services have produced all but which of the following lessons?

   The correct answer is "e". GIS studies have produced some valuable lessons learned, including the following:

- The likelihood that a person will choose a particular health facility decreases as the distance to it increases. This effect is more pronounced in rural settings in comparison to urban.
- Distance to services is extremely important for determining access, as usage drops off significantly at a certain distance.
- Proximity to a road influences access.
- Access measures differ based on whether analysis is conducted using straight-line (Euclidean) or actual transportation network distances.
- Private sector providers and facilities, which are difficult for a ministry of health (MoH) to regulate, are generally under-reported on official lists. This creates blind spots in the knowledge base required to determine equitable access to health care.
- Maps allow areas of “good” and “bad” access to be pinpointed, even when they are mixed.
- A GIS allows effective targeting of interventions.

   Sources: Noor et al., 2003; Noor et al., 2004; Rosero-Bixby, 2003.

5. Participatory mapping is expensive, but is warranted based on the high quality of results.

   The correct answer is "b". Dongus et al. (2007) demonstrated the cost-effectiveness of participatory mapping, as it required only a minimal level of technical skills and equipment.

6. In what key way(s) does geography strengthen the national data infrastructure?

   The correct answer is "e". Geography strengthens the national data infrastructure in two key ways: (1) allows linking of datasets from different sources and (2) promotes standardization of data for sharing. The successful integration of geography into the national data infrastructure may also create a greater demand for maps, although that is a secondary benefit. The artistic quality of maps may have no direct impact on the national data infrastructure.

7. A strong national data infrastructure strengthens data at multiple scales, from local to national to international, and makes effective public health decision making more likely.

   The correct answer is "a."

8. Almost all human activity occurs on the surface of the Earth. What essential information does this provide for public health analysts and decision makers to help ask better questions and plan interventions to improve human health?

   The correct answer is "d". The geographic locations of populations affected by health risks provide vital information for public health analysts and decision makers, as understanding the "where" helps ask better questions and plan interventions to understand the "why." Land cover, elevation, and socioeconomic status are examples of the many factors that can potentially influence a population's health based on its geographic location.

9. What is the term used to refer to the ***geographic*** data available to national-level decision makers, as well as the people, policies, and systems required to collect, store, manage, analyze, and disseminate the data for decision making purposes?

   The correct answer is "d." The "technologies, policies, and people necessary to promote sharing of geospatial data throughout all levels of government, the private and non-profit sectors, and the academic community. The goal of this Infrastructure is to reduce duplication of effort among agencies, improve quality and reduce costs related to geographic information, to make geographic data more accessible to the public, to increase the benefits of using available data, and to establish key partnerships with [stakeholders] to increase data availability." *Source*: Federal Geographic Data Committee, www.fgdc.gov/nsdi/nsdi.html, April 2010.

10. Geographic identifiers make it possible to do what?

    The correct answer is "d". Geographic identifiers specify physical locations, which allow one to (a) identify the corresponding administrative division (e.g., province, state, or district) and to (b) link physical locations and their corresponding geographic entities with attribute data from different data sources.

11. How are geographic and projected coordinate systems different?

The correct answer is "d". Geographic coordinate systems are based on a sphere, and are therefore three-dimensional or spherical. It is difficult to obtain accurate, consistent measurements of area, shape, and distance using the spherical coordinates of a geographic coordinate system. Projected coordinates transform or "project" three-dimensional coordinates onto a flat, two-dimensional surface. Projected coordinate systems can preserve areas, shapes, or distances, but not all three simultaneously. Depending on the geographic properties preserved by a projection, the projection makes it easier to measure them accurately.

12. What is a good source for standardized unique geographic identifiers?

The correct answer is "e". The International Standards Organization (ISO) provides standardized codes for countries (ISO 3166-1) and their subdivisions (ISO 3166-2). The United Nations publishes UN/LOCODE to provide standardized codes for tens of thousands of locations used in trade and transport, such as ports, airports, and rail terminals. The three sets of codes work together to identify geographic entities consistently down to a detailed level.

13. What is a good way to maintain the confidentiality of geographic data?

The correct answer is "e". Some good ways to maintain the confidentiality of geographic data are as follows:

- Shifting point locations by a randomly selected distance and direction.
- Restricting the publication or distribution of maps made from the data.
- Generalizing spatial information (e.g., rounding of latitudes) to obscure true locations.
- Removing all personal and geographic identifiers from data provided to users.

14. Which of the following is a strength of virtual globes?

The correct answer is "e". Imagery may be out-of-date because of the time and expense required to obtain it. A good Internet connection is required to access most virtual globes, at least for the initial download and caching of imagery on a local computer. Virtual globes are intended more for display of geographic data than for analysis. The national organizations that create and maintain administrative boundaries generally do not provide detailed, accurate versions of these boundaries to external organizations such as virtual globe developers.

15. A fully-featured GIS can be distinguished from a simpler mapping tool by the inclusion of what capabilities:

The correct answer is "c". Other types of mapping software can often input data and output simple maps, including those with scale bars and legends, but generally speaking only a full-featured GIS can perform more complicated data analysis

16. The most common coordinate system used by most GPS units is UTM.

The correct answer is "b". Although most GPS receivers can be configured to record geographic locations using a variety of coordinate systems and datums, the most common coordinate system used is a geographic coordinate system based on latitude/longitude and the WGS 84 datum. UTM stands for Universal Transverse Mercator, which is one of the most useful projections around the world based on the fact that it uses the metric system and produces distance measurements that are accurate to within a meter when taken within one of its 60 zones.

17. Identify the example(s) of external data:

The correct answer is "e". B and C are good sources of external data that are often available for third-party use. In-house data or data collected in the field would be considered internal data.

18. The term spatial analysis generally refers to more complicated GIS operations such as the creation of predicted data points or data surfaces, and often involves the use of spatial statistics.

The correct answer is "a".

19. Which of the following is NOT a type of thematic map:

The correct answer is "a". Thematic maps all involve the classification of data, and most road maps do not.

20. Spatial autocorrelation refers to:

The correct answer is "c". Spatially autocorrelated data clusters together. Negative autocorrelation refers to regularly spaced data. Data with no autocorrelation is simply random in location.

21. The development of a comprehensive, spatially referenced national inventory of health facilities will provide the foundation for which of the following?

The correct answer is "d".

22. What is the benefit of calculating global or local Moran's I values of spatial autocorrelation?

The correct answer is "c". Moran's I is a statistical measure that incorporates the element of spatial location. Given a set of features with known locations for a study area, Global Moran's I evaluates whether the overall pattern of attribute values attached to those locations is clustered, dispersed, or random. Source: ESRI, 2010, accessed June 2010. The local Moran's I is used as an indicator of local spatial association. Source: Uthman et al., 2009. Global and local Moran's I values are expressed numerically using Z scores and p-values, which can be mapped, but not as graphical summaries such as box plots and histograms generated through exploratory spatial data analysis (ESDA).

23. What method of geographic analysis can be employed to construct a highly accurate, national map of malaria prevalence using data attached to georeferenced survey locations, such as health care facilities, that are distributed throughout a country?

The correct answer is "b". Kriging refers to a group of geostatistical techniques used to interpolate the value of a variable (e.g., number of malaria cases) for locations situated between known locations using observations of its value at nearby locations. Source: http://en.wikipedia.org/wiki/Kriging, accessed June 2010.

This would allow a researcher to generate a spatially continuous set of data for an entire geographic area (e.g., a national map of malaria prevalence), using data corresponding to a set of geographically distributed points with known locations, such as health care facilities. For examples of kriging to create highly accurate malaria prevalence maps for Kenya and Somalia, see Noor et al., 2009 (Kenya) and Noor et al., 2008 (Somalia).

For more information on the incorrect answers, see the glossary entries for supervised classification and ground truthing.

24. What geographic technique can be used with data attached to point-based survey locations to create a polygon-based map for an entire country?

The correct answer is "c". Thiessen polygons use point locations to create a polygon-based surface for a geographic area. Thiessen polygons have the unique property that each polygon contains only one input point, and any location within a polygon is closer to its associated point than to the point of any other polygon. Source: http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=1349&pid=1347&topicname=Create_Thiessen_Polygons_%28Analysis%29, accessed June 2010.

Buffer polygons, which would be created using a buffer tool, could be of any radius and might not result in polygons that covered an entire country. For more information on the buffer tool, see session 7.2. Ground truthing is not associated with the creation of polygons, but with the execution of fieldwork to verify features on the surface of Earth.

25. In addition to a coordinate system, what else must be specified to identify a location with maximum accuracy?

The correct answer is "b". Since the Earth is flattened at the poles, bulges at the equator, and has an irregular surface, it is not a perfect sphere. As a result, specifying a geographic location with maximum accuracy also requires identification of a specific model of the earth, known as a datum (e.g., the World Geodetic System 1984 or WGS 84). This applies regardless of whether the coordinate system is geographic/unprojected or projected.

26. Open source GIS software is readily available, free of charge, and always comes with extensive documentation.

The correct answer is "b". It is true that open-source software is becoming more readily available and is often free, but documentation is often scarce or not subject to updates along with the software. Effective use may require willingness by the user to experiment and even to seek out online forums for help.

GLOSSARY OF TERMS

A	Top
Aerial photography	Photography from airborne platforms. (Canada Centre for Remote Sensing)
Antenatal care (ANC)	Antenatal care constitutes screening for health and socioeconomic conditions likely to increase the possibility of specific adverse pregnancy outcomes; providing therapeutic interventions known to be effective; and educating pregnant women about planning for safe birth, emergencies during pregnancy, and how to deal with them. Antenatal care coverage is an indicator of access and use of health care during pregnancy. (WHO)
Antiretroviral therapy (ART)	Standard antiretroviral therapy (ART) consists of the use of at least three antiretroviral (ARV) drugs to maximally suppress the HIV virus and stop the progression of HIV disease. Huge reductions have been seen in rates of death and suffering when use is made of a potent ARV regimen. (WHO)
B	Top
Box plot	A box plot is a graphical summary of the following statistical measures: median, upper and lower quartiles, and minimum and maximum data values. (NETMBA)
Buffer	An area created by specifying a particular distance from a point, line, or polygon on a map. Can be used to identify geographic features that occur inside or outside a certain distance from another feature.
Buffer tool	Creates buffer polygons to a specified distance around the Input Features. An optional dissolve can be performed to remove overlapping buffers. (ESRI)
C	Top
Catchment area	Catchment area is the area from which a health facility attracts patients. A simple means of estimating a catchment area is to define a radius beyond which individuals are unlikely to access the services offered at that facility.
Choropleth map	A map that uses colors or shading to display attribute data for geographic areas rather than for points. To display values that take into account the size differences of the geographic areas, data should first be normalized (e.g., calculating population density, such as people per square kilometer, rather than using simple population counts). Choropleth maps are most visually informative when they display between 2 and 7 classes of data using colors or shades that gradually darken as values increase.
Classification scheme	Hierarchical system of user-defined classes for the classification of remote sensing imagery. (FWIE)
Cluster	For Demographic and Health Surveys (DHS) data collection, the geographic location is collected based on what is known as the "cluster." DHS clusters are usually census enumeration areas, sometimes villages in rural areas or city blocks in urban areas, containing the households selected for survey. A single GPS location is recorded at the center of the settlement area of the cluster. Collecting only one point for the cluster greatly reduces the chance of

	compromising confidentiality of the respondents, but is enough to allow the integration of multiple datasets for further analysis. (MEASURE DHS)
Confidentiality	The result achieved by protecting data and information that might identify individuals in a way that could cause harm or otherwise violate agreements made with them. For more information, see the MEASURE Evaluation publication "Overview of Issues Concerning Confidentiality and Spatial Data" which is listed on the Reference & Links page.
Confidentiality agreement	Agreement entered into between a public health organization and an individual regarding the protection and non-disclosure of personally identifiable information.
Coordinate system	A coordinate system is a reference system used to represent the locations of geographic features, imagery, and observations such as GPS locations within a common geographic framework. (ESRI)
D	Top
Data demand and use	Demand for and use of data to enhance evidence-based decision making for public health. Activities that foster data demand and use involve a systematic approach that applies proven, effective best practices and appropriate tools to help increase demand for health system data and ensure that the information is used in an evidence-based decision making process. (MEASURE Evaluation)
Data dictionary	A data dictionary is a text-based description of tables and fields in a database. It provides a solid foundation for writing data cleaning programs, and offers a common language for facilitating communications between managers and analysts. (CDC)
Data schema	A data schema is a description of how data in a computer database are organized into tables and fields, and identifies acceptable values for individual fields. A common way to capture a data schema is in a data dictionary. A proper data schema ensures that data are standardized and complete, and that they can be used to create accurate maps.
Datum	A set of control points, which are points on the surface of Earth with known locations, and a corresponding mathematical model used to approximate the shape of Earth and to calculate the location of any given point on that shape.
Decimal degrees	A numeric format for storing latitude and longitude that makes it easier to import coordinates into a GIS and to use for location-based calculations. For example, a comparison of latitude and longitude formats for the location of the Library of Alexandria in Egypt is as follows: Degrees, minutes, seconds: 31°12'31.93"N, 29°54'33.62"E Decimal degrees: 31.208870, 29.909339
Deductive disclosure	The act of piecing together information from various sources or from distinct elements within a single dataset to gain insight or knowledge that should be kept secret. (MEASURE Evaluation 2008)
Demographic and Health Surveys (DHS)	MEASURE Demographic and Health Surveys (DHS) are nationally-representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition. Standard DHS surveys have large sample sizes (usually between 5,000 and 30,000 households) and are

	typically conducted every five years to allow comparisons over time. The DHS Project captures survey locations based on "clusters" rather than individual households. MEASURE DHS provides an inventory of available survey datasets at http://www.measuredhs.com/accesssurveys/search/start.cfm. (MEASURE DHS)
Dot density map	A map that uses dots to display data on a map. Each dot generally represents a given quantity of a certain occurrence (not necessarily one), such as 10 people per dot.
E	Top
Environmental Systems Research Institute (ESRI)	The leader in commercial GIS software and services. Based in Redlands, CA (www.esri.com).
Equal intervals	A method of data classification which often works better for data that are continuous, that is, not highly skewed. Each resulting range (interval) of values will be approximately equal, but there may be a very different number of observations per class. Attention will be more focused on outliers. The map resulting from this method of data classification will tend to highlight any data with particularly high or low values, and may show an uneven distribution of colors.
Euclidean distance	A distance calculated using a straight line to connect the beginning and ending points.
Exploratory data analysis (EDA)	Exploratory data analysis (EDA) applies a variety of statistical tools, such as box plots, to quickly summarizing and gaining insight into a set of data. (Encyclopedia Britannica)
Exploratory spatial data analysis (ESDA)	Exploratory spatial data analysis (ESDA) applies EDA statistical tools to the evaluation of spatial data. A basic technique for ESDA is to link observations in a histogram, box plot, and map to detect spatial patterns, such as outliers. (Anselin 2005)
G	Top
Geocodes	Coded variables in a table of geographic data which indicate position, either at a point or within an area. They can be codes indicating latitude/longitude coordinates or administrative areas. This is a coded version of a geographic identifier.
GeoDa	GeoDa is a collection of software tools available from http://geodacenter.asu.edu/ designed to implement techniques for exploratory spatial data analysis (ESDA) on lattice data (see note below). It is intended to provide a user friendly and graphical interface to methods of descriptive spatial data analysis, such as autocorrelation statistics and indicators of spatial outliers. The design of GeoDa consists of an interactive environment that combines maps with statistical graphics, using the technology of dynamically linked windows. Note: Lattice data are discrete spatial units that are not a sample from an underlying continuous surface (geostatistical data) or locations of events (point patterns). GeoDa currently does not yet contain specific techniques to analyze geostatistical or point pattern data. (Anselin 2003)
Geographic coordinate system (GCS)	A coordinate system based on a three-dimensional, spherical surface. As a result of being defined in relation to the more natural, three-dimensional surface of a globe, a GCS is considered to be "unprojected" rather than "projected." For an explanation of geographic versus projected coordinate systems, see Session 3, pages 2 and 3.

Geographic data	Information describing the location and attributes of things, including their shapes and representation. (ESRI [ArcGIS documentation])
Geographic identifier	A geographic identifier is any piece of information that indicates the geographic or spatial location of features on the landscape, such as latitude and longitude, street address, P-code (place code) from the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA), or administrative division name (i.e., province, district, county, etc.). Common geographic identifiers play a critical role in joining data from different sources.
Geographic information system(s) (GIS)	A computer-based system used to collect, store, manage, analyze, display, and distribute geographic data (points, lines, and polygons referenced to the surface of the Earth) and their attributes (e.g., unique identifier, name, type, date collected, etc.). (MEASURE Evaluation)
Geography	Geography is the study of patterns on the surface of Earth, and the causes of those patterns. The patterns can be the result of natural forces or human activity. This glossary entry is a synthesis of definitions from several sources, as many definitions of geography emphasize sub-fields of geography, and can be too narrow in scope. (MEASURE Evaluation)
Georeferenced	Assigned to a geographic location.
Global Fund to Fights AIDS, Tuberculosis and Malaria (Global Fund)	An international financing institution that invests the world's money to save lives. To date, it has committed US$ 19.3 billion in 144 countries to support large-scale prevention, treatment and care programs against the three diseases. (The Global Fund)
Global positioning system (GPS)	Satellite-based system originally created by the United States Department of Defense to provide accurate data on position, velocity, and time to both military and civilian users. Coordinates are generally given in digital degrees relative to the equator and prime meridian.
Google Earth	A virtual globe from Google (www.google.com). It has the highest user base and satellite imagery library currently available, and can be downloaded from http://earth.google.com. For more information, see Session 4, page 2.
Ground truth	Information about a feature on the Earth's surface that is collected in the field. In remote sensing, the process of acquiring ground truth data is referred to as "ground truthing."
H	Top
Health management information system(s) (HMIS)	A planned system of collecting, processing, storing, disseminating, and using health-related information to carry out functions of management. It consists of people, tools (paper-based and electronic) and procedures to gather, sort, and distribute timely, accurate information to decision-makers. (Kotler and Keller 2006)
Histogram	A histogram is a graphical summary showing the count of data points falling in various ranges. It provides a rough approximation of the frequency distribution of the data. (NETMBA)
I	Top
IKONOS	The IKONOS Satellite is a high-resolution satellite operated by GeoEye (www.geoeye.com).
Imagery	Pictures or graphical representations. The term is used in remote sensing and GIS to describe digital representations of the surface of Earth. (FWIE)

Incidence	The number of new events, such as new cases of a disease, occurring over a specific period of time. It is often expressed as a rate, for example the number of cases per 100,000 population. (U.S. Global Health Policy Web site)
Isarithmic map	A map that uses contours lines to show change in a continuous variable over the land surface, such as temperature, precipitation, or elevation.
K	Top
Kernel density estimation (KDE)	A geographic technique that disperses discrete phenomena across continuous space without the constraints of administrative boundaries. It provides a more realistic representation of the spread of people and services across a landscape. (Spencer and Angeles 2007)
KML	KML, which originally stood for Keyhole Markup Language, is an XML-based file format that can incorporate descriptive text, image links, and geographic information associated with points, lines, and polygons. It is an open standard officially named the OpenGIS® KML Encoding Standard (OGC KML). KML files can be read by Google Earth and several mapping software packages. (Open Geospatial)
Kriging	Kriging refers to a group of geostatistical techniques used to estimate the values of variables at locations with no values based on the known values at known locations nearby. As such, it can be used to create a map covering an entire geographic area using values attached to a collection of points, such as health facility locations, distributed throughout that area. For more details on kriging, see A Practical Primer on Geostatistics. (USGS 2009)
Kulldorff spatial scan statistic	A measure of spatial autocorrelation created by Martin Kulldorff and incorporated into a software package known as SaTScan (www.satscan.org). (Kulldorff 2009)
L	Top
Land cover	What can be seen remotely, from satellite data or aerial photographs. Current mapping techniques of land cover would not be possible today without milestones such as James Anderson’s 1976 publication, A Land Cover Classification System for Use with Remote Sensor Data. (CPC and USGS)
Landsat Thematic Mapper (TM) 7	The Landsat Project is a joint initiative of the U.S. Geological Survey (USGS) and the National Aeronautics and Space Administration (NASA) designed to gather Earth resource data from space. Landsat represents the world's longest continuously acquired collection of space-based moderate-resolution land remote sensing data. Landsat Thematic Mapper (TM) 7 was launched April 15, 1999, and provides imagery captured using eight spectral bands. (USGS)
Latitude	Angle between a line connecting the center of the Earth to the equator and a line connecting the center of the Earth to a point on the Earth’s surface on, north, or south of the equator along a line of longitude. Latitude ranges from 0 degrees at the equator to 90 degrees at the poles. Latitude is positive north of the equator (0 to 90 degrees) and negative below it (0 to -90 degrees). Lines of constant latitude can be visualized as circles drawn around the Earth horizontally in parallel with the equator.
Likelihood function	Likelihood function is a fundamental concept in statistical inference. It indicates how likely a particular population is to produce an

	observed sample. (Statistics.com)
Logit model (also call logistic regression or logistic model)	Logistic regression (sometimes called the logistic model or logit model) is used to predict the probability of an event's occurrence. For example, the probability that a person has a heart attack within a specified time period might be predicted from knowledge of the person's age, sex, and body mass index. Logistic regression is used extensively in the medical and social sciences as well as marketing applications such as prediction of a customer's propensity to purchase a product or cease a subscription.
Longitude	Angle between (a) a line connecting the center of the Earth to the equator at a prime meridian, such as the meridian that passes from pole to pole through Greenwich, England (also known as the Prime Meridian or Greenwich Meridian), and (b) a line connecting the center of the Earth to the equator at its intersection with a meridian that passes through the point of interest. Longitude ranges from 0 degrees at the Prime Meridian to 180 degrees along the meridian on the opposite side of the Earth. The 180th meridian roughly parallels the International Date Line, where the date changes as travelers cross going east or west. Lines of constant longitude can be visualized as half circles drawn on the Earth's surface vertically from pole to pole.
M	Top
Medical geography	Medical geography applies the discipline of geography to the study of patterns of public or human health. (Meade and Emch 2010)
Metadata	Data about data, such as the source, date of data creation, date(s) for which data are relevant, etc. As a best practice, metadata should be provided with any geographic dataset. International standards for geographic metadata are available as ISO 19115. (ISO)
Millennium Development Goals (MDGs)	A set of eight broadly supported, comprehensive and specific goals developed by the United Nations in 2000 that provide concrete, numerical benchmarks for tackling extreme poverty in its many dimensions. Targeted for completion by 2015, the MDGs form a blueprint agreed to by all of the world's countries and all of the world's leading development institutions to meet the needs of the world's poorest. (UN)
Monitoring and evaluation (M&E)	Monitoring of a program or intervention involves the collection of routine data that measure progress toward achieving program objectives. It is used to track changes in program performance over time and to allow stakeholders to make informed decisions regarding the effectiveness of programs and the efficient use of resources. Evaluation measures how well the program activities have met expected objectives and/or the extent to which changes in outcomes can be attributed to the program or intervention. (CPC)
Moran's I	Moran's I is a statistical measure that incorporates the element of spatial location. Given a set of features with known locations for a study area, Global Moran's I evaluates whether the overall pattern of attribute values attached to those locations is clustered, dispersed, or random. The local Moran's I is used as an indicator of local spatial association. (ESRI [ArcGIS documentation] and Uthman 2009)
Most at-risk populations (MARPs)	Primarily used in association with HIV prevalence, the term MARPs focuses on female sex workers, clients of female sex workers, injecting drug users, and men who have sex with men. Also of concern are the sexual partners of MARPs, who may not be aware that they are at risk. These other priority populations, which are

	primarily composed of women of reproductive age and adolescents, would include sexual partners of injecting drug users, female partners of men who have sex with men, and partners of sex worker clients. (UNAIDS 2008)
MrSID (pronounced Mister Sid)	MrSID is a proprietary data format that allows raster image files to be substantially compressed with little to no loss of image quality. This format, developed and marketed by LizardTech, Inc. (http://www.lizardtech.com/), allows you to view any portion of a MrSID image at any resolution very quickly. The name stands for "Multi Resolution Seamless Image Database." (University of Virginia Library)
Multispectral (imagery)	Two or more images taken simultaneously, but each image taken in a different part of the electromagnetic spectrum. The electromagnetic spectrum is the total range of wavelengths or frequencies of electromagnetic radiation, extending from the longest radio waves to the shortest known cosmic rays. A portion of the electromagnetic spectrum corresponds to light that is visible to the naked eye. (CCRS)
N	Top
National data infrastructure	The national data infrastructure consists of all the data available to national-level decision makers as well as the people, policies, and systems required to collect, store, manage, analyze, and disseminate the data for decision-making purposes. (MEASURE Evaluation)
National mapping agency (NMA)	National organization responsible for the creation and maintenance of map series and related data for a country. A good source of NMA contacts is the United Nations Second Administrative Level Boundaries (UN SALB) Web site (www.unsalb.org).
National spatial data infrastructure (NSDI)	The “technologies, policies, and people necessary to promote sharing of geospatial data throughout all levels of government, the private and non-profit sectors, and the academic community. The goal of this Infrastructure is to reduce duplication of effort among agencies, improve quality and reduce costs related to geographic information, to make geographic data more accessible to the public, to increase the benefits of using available data, and to establish key partnerships with [stakeholders] to increase data availability.” (FGDC)
Natural breaks	A method of data classification that assigns data to classes such that the variance within classes is minimized while the variance between classes is maximized. The primary advantage of this method is that it takes the natural distribution of the data into account before assigning observations to classes. A disadvantage is that the breaks between the classes could be irregular and therefore not intuitive.
Network analysis	A method of geographic analysis that calculates measures along a network of such entities as roads, railroads, or rivers. It can be used to study accessibility of health care services.
O	Top
Orphans and Vulnerable Children (OVC)	In the PEPFAR context, the term OVC is defined as a child, 0-17 years old, who is either orphaned or made more vulnerable because of HIV/AIDS. Orphan: Has lost one or both parents to HIV/AIDS. Vulnerable: Is more vulnerable because of any or all of the following factors that result from HIV/AIDS: (a) Is HIV-positive; (b) Lives without adequate adult support (e.g., in a household with chronically ill parents, a household that has experienced a recent death from chronic illness, a household headed by a grandparent, and/or a

	household headed by a child); (c) Lives outside of family care (e.g., in residential care or on the streets); or (d) Is marginalized, stigmatized, or discriminated against. (PEPFAR OGAC 2006)
Orthorectification	A process by which geographic distortions caused by sensor tilt and terrain relief are removed. An image that has been submitted to orthorectification is referred to as orthorectified. (CCRS)
Outlier	An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. This definition leaves it up to the analyst (or a consensus process) to decide what will be considered abnormal. (ITL NIST)
P	Top
Participatory mapping	In the context of public health, participatory mapping at the local level is a form of strategic mapping that engages the community in the identification of places and populations likely to be involved in the transmission of disease. For more information, see Session 2, page 6.
PEPFAR	The U.S. President's Emergency Plan for AIDS Relief (www.pepfar.gov). Launched in 2003 by President George W. Bush, PEPFAR represents the largest, most focused effort in history to defeat a single disease. (PEPFAR)
Percentile	In a population or a sample, the Pth percentile is a value such that at least P percent of the values take on this value or less and at least (100-P) percent of the values take on this value or more. (Statistics.com)
Pixel	"Picture element" is the ground area corresponding to a single element of a digital image dataset. (CCRS)
Plasmodium falciparum parasite rate (PfPR)	The proportion of a random sample of population with malaria parasites in their peripheral blood. The PfPR has become the benchmark indicator by which malaria risk is modeled and mapped in Africa. (Noor et al. 2009)
Positional accuracy	The degree to which a user's location can be accurately determined. GPS data results can vary according to the arrangement of satellites overhead.
Prevalence	The number or proportion of events, such as cases of a disease, within a population at a specific point in time, e.g., the proportion of a population living with a disease. (U.S. Global Health Policy Web site)
Priorities for Local AIDS Control Efforts (PLACE)	Priorities for Local AIDS Control Efforts (PLACE) is a rapid assessment tool that applies epidemiologic and social science principles to identify specific geographic pockets within countries where HIV transmission is most likely to occur. The method systematically monitors the extent to which AIDS prevention programs are reaching key sexual and injecting drug use networks and recommends where prevention programs should be targeted. (MEASURE Evaluation)
Projected coordinate system	A coordinate system in which locations on the three-dimensional surface of the Earth are transformed or "projected" onto a flat, two-dimensional surface for display, measurement, or other analysis.
Proportional symbol map	A map that uses varying sizes of symbols (often circles) to display attribute data for geographic areas or points.
Q	Top

Quantiles	A method of data classification that attempts to place an equal number of values into each class. For example, if there were 50 observations, each with a different value so that no duplicates were encountered during data classification, grouping them into five classes (also known as quintiles) would result in 10 observations per class. Quantiles are useful for classifying ordinal (rank-order) data, and for comparing maps with the same number of classes. Outlier values will be less visible and attention will be focused on relative rankings. The map resulting from this method of data classification will tend to produce an even distribution of map colors. Caution: If the data is highly skewed, the quantiles classification method will still place the data into the number of classes specified. This forcing of unevenly distributed data into classes containing equal numbers of observations could lead to the false impression that the data is normally distributed.
Quartile	The 1st, 2nd, and 3d quartiles are the 25th, 50th, and 75th percentiles respectively. (Statistics.com)
R	Top
Raster data	Spatial data stored in a computer as a series of values in a grid pattern (pixels). This type of data generally requires much more computer storage space than vector data. Larger numbers of pixels over a smaller area provide greater spatial resolution but take up much more memory. This type of data can show continuous change over a surface, such as land cover. Satellites collect data in this format.
Remote sensing	The science, technology and art of obtaining information about objects or phenomena from a distance (i.e., without being in physical contact with them). (CCRS)
S	Top
Scale	The Earth is too large to draw on a map without reducing its size. This reduction is expressed as map scale, which is the ratio of the distance on a map to the actual distance on the surface of the Earth. As a result, a small-scale map displays a small amount of detail, but covers a large geographic area. A large-scale map shows a large amount of detail, but for a small area. Scale can be expressed graphically as a scale bar, or in writing using text or numeric forms: (i) Text: 1 inch = 24,000 inches OR 1 inch = 2,000 feet; or (ii) Numeric: 1:24,000.
Shapefile	A spatial data format originally developed by ESRI and in widespread use today. A shapefile is actually a collection of at least three files: (i) Main file (filename.shp), which contains geometric information for the features of interest on a record-by-record basis. (ii) Index file (filename.shx), which identifies the positional offset of each record in the main file from the beginning of the main file. (iii) dBASE file (filename.dbf), which contains a table of attribute data for each geometric feature described in the main file. A shapefile can also have a projection file (filename.prj) to specify the coordinate system and datum. Although a projection file is optional with respect to the shapefile technical specification, it is essential for accurate geographic analysis.
Signature domain (for health facilities)	The health facility signature domain from MEASURE Evaluation contains all the information necessary to identify a facility uniquely, and should be explicitly included in all health facility surveys. While each element in the Signature Domain may or may not uniquely

	identify a facility (and indeed some elements may not exist), the collection of as many elements as possible will more reliably result in correct matching of facilities across surveys. (MEASURE Evaluation 2007)
Spatial autocorrelation	Spatial autocorrelation refers to the dependence on spatial location of observed values of a phenomenon. If spatial autocorrelation exists, values are considered to be clustered or grouped geographically. If no autocorrelation exists, values are considered to be randomly distributed geographically. For more information, see glossary entry for Moran's I.
Spatial data	Data that describe the geographic shape and location of entities in relation to the physical space defined by the Earth's surface. In terms of shape, spatial data can take the form of points, lines, or polygons. With respect to location, spatial data are organized and displayed according to coordinate systems and datums.
Spatial rate smoothing	Rate smoothing is a procedure to address the variance instability related to estimating rates in areas with widely varying populations. Variance instability is particularly pertinent in areas with small population numbers. Raw rates and smoothed rates will differ less as underlying population numbers in areas increase. Smoothing increases the precision of risk estimates. (GeoDa Center)
SPOT 5	High spatial resolution satellite owned and operated by the French company SPOT Image (www.spot.com). For a technical specifications sheet and a gallery of SPOT imagery, see www.satimagingcorp.com
Stovepiping	Term referring to the practice of keeping data separate from, and not connected to, other data. Stovepiping of data can lead to isolated, one-time solutions and narrow interpretations of data.
Strategic information (SI)	The knowledge that guides health policy, planning, program management and service delivery. It is essential for evidence-based action at all levels of the health system. (WHO)
Strategic information system (SIS)	The "hardware, software, facilities, data, and personnel" needed "to achieve the most cost-effective system for satisfying [an] organization's [information] needs." (U.S. GAO 1992)
Supervised classification	A procedure for identifying spectrally similar areas on an image by identifying 'training' sites of known targets [known via ground truthing conducted during fieldwork] and then extrapolating those spectral signatures to other areas of unknown targets. In contrast, an unsupervised classification is the categorization of digital image data by computer processing based solely on the image statistics without availability of training samples or a-priori knowledge of the area. (CCRS)
Supply chain management	A supply chain is a network that includes vendors of raw materials, plants that transform those materials into useful products, and distribution centers to get those products to customers. Without management, each organization in the overall supply chain system has its own agenda and operates independently from the others. (QUICKMBA and SCMS)
Surveys and surveillance	Biological and behavioral surveys and surveillance are essential to determine the drivers and the spread of the HIV epidemic in a country. HIV surveys and surveillance may focus on the general population, most-at-risk populations, or both. Protocols and data collection tools for surveys should be based on international

	standards, such as the Demographic and Health Survey [www.measuredhs.com], the AIDS Indicator Survey, and the Multiple Cluster Indicator Survey. (UNAIDS 2008)
T	Top
Thiessen polygons	Thiessen polygons use point locations to create a polygon-based surface for a geographic area. Thiessen polygons have the unique property that each polygon contains only one input point, and any location within a polygon is closer to its associated point than to the point of any other polygon. (ESRI)
U	Top
Unique geographic identifier	A name or code that uniquely identifies a geographic entity. Examples include province or district name for an administrative region; waypoint ID and latitude/longitude combination for a GPS point; P-code (place code) from the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA). Unique geographic identifiers are essential for distinguishing between individual geographic entities and ensuring the accuracy of attribute data for those entities, as non-unique identifiers can create confusion and produce errors during all phases of data use. (MEASURE Evaluation)
V	Top
Vector data	Spatial data stored in a computer as points, lines, and polygons. In the case of a straight line, the coordinates of one point, the distance and direction to a second point, and the coordinates of the second point will all be stored. This is usually the most effective method of spatial data storage.
Virtual globe	A 3D representation of the Earth that provides the ability to zoom in and out through a wide variety of scales and to change viewing angle. Virtual globes often combine satellite imagery collected at varying levels of detail with actual aerial or even street-level photography. They also often allow additional overlays such as points, maps, or images.
X	Top
XML	XML, which stands for eXtensible Markup Language, is a simple, flexible text format that plays an increasingly important role in the exchange of data on the Web. (w3)
Z	Top
Z scores and p-values	The Z score is a test of statistical significance that helps you decide whether or not to reject the null hypothesis. The p-value is the probability that you have falsely rejected the null hypothesis. Z scores are measures of standard deviation. For example, if a tool returns a Z score of +2.5 it is interpreted as "+2.5 standard deviations away from the mean". P-values are probabilities. Both statistics are associated with the standard normal distribution. This distribution relates standard deviations with probabilities and allows significance and confidence to be attached to Z scores and p-values. (ESRI)
Zonal statistics tool	Calculates statistics on values of a raster within the zones of another dataset. (ESRI)

Brody, et al. 2000. "Map-making and myth-making in Broad Street: the London cholera epidemic, 1854." *The Lancet* Volume 356 Issue 9223, pages 64-68.

Brownson, RC et al. April 2009. Evidence-Based Public Health: A Fundamental Concept for Public Health Practice. *Annual Review of Public Health* Volume 30, 175-201.

Catholic Relief Services and Catholic Secretariat of Nigeria, 2008. A Situational Analysis of Orphans and Vulnerable in Eight States of Nigeria, accessed June 2010.

Chang et al. 2009. Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. *International Journal of Health Geographics* 8:49.

Chirowodza et al. 2009. Using Participatory Methods and Geographic Information Systems (GIS) To Prepare For an HIV Community-Based Trial in Vulindlela, South Africa. *Journal of Community Psychology*, Volume 37 Issue 1, pages 41-57.

Dramowicz, E and K. December 29, 2004. Choropleth Mapping with Exploratory Data Analysis. *Directions Magazine*, accessed June 2010.

Dambach et al. 2009. Using high spatial resolution remote sensing for risk mapping of malaria occurrence in the Nouna district, Burkina Faso. Global Health Action.

Department of Epidemiology. Accessed September 2010. Broad Street Pump Outbreak. Los Angeles, CA: University of California, Los Angeles (UCLA) School of Public Health.

Dongus et al. 2007. Participatory mapping of target areas to enable operational larval source management to suppress malaria vector mosquitoes in Dar es Salaam, Tanzania. *International Journal of Health Geographics* 6:37, accessed April 2010.

Duncan, G, Jabine, T, and V de Wolf. 1993. “Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics,” Washington, DC, National Academy Press.

Entwisle et al. 1997. "Geographic Information Systems, Spatial Network Analysis, and Contraceptive Choice" *Demography* Volume 34, Number 2, pages 171-187.

Federal Committee on Statistical Methodology. 2005.

Golden, ML, Downs, RR, and K Davis-Packard. 2005. “Confidentiality Issues and Policies Related to the Utilization and Dissemination of Geospatial Data for Public Health Applications, A Report to: The Public Health Applications of Earth Science Program, National Aeronautics and Space Administration, Science Mission Directorate, Applied Sciences Program,” Socioeconomic Data and Applications Center (SEDAC), Center for International Earth Science Information Network (CIESIN), Columbia University.

Green, E. April 2010. "Participatory Mapping and Geospatial Technologies to Inform the Development of a Setting-Level HIV Prevention Intervention in Muhuru Bay, Kenya," presentation at AAG: Geographic Contexts of Global Health.

Heard, NJ, U.S. Department of State, Humanitarian Information Unit. January 2007. "Spatial Data as Strategic Information within PEPFAR," Presentation at ESRI Federal User Conference, Washington, DC.

Houlihan et al. 2010. The tuberculosis challenge in a rural South African HIV programme. *BMC Infectious Diseases* 10:23.

International Fund for Agricultural Development (IFAD). 2009. Good Practices in Participatory Mapping, accessed April 2010.

International Institute for Tropical Agriculture (IITA). Accessed September 2010. Ibadan, Oyo State, Nigeria: IITA.

International Organization for Standardization (ISO). Accessed September 2010. ISO 3166-1. Geneva: ISO.

International Organization for Standardization (ISO). Accessed September 2010. ISO 3166-2. Geneva: ISO.

Kamadjeu, R. 2009. Tracking the polio virus down the Congo River: a case study on the use of Google Earth™ in public health planning and mapping. *International Journal of Health Geographics* 8:4.

Khan, O.A. et al. 2010. Geographical information systems and tropical medicine. *Annals of Tropical Medicine and Parasitology*. 104(4): 303-318.

Koch, T. 2005. Cartographies of Disease: Maps, Mapping, and Medicine. ESRI Press, Redlands, California.

Kotler, P and KL Keller. 2006. Marketing Management, 12th Ed. Pearson Education. Upper Saddle River, NJ.

Meade and Emch, 2010. Medical Geography, Third Edition. Guilford Press, New York, New York.

Mmari et al. 2009. The Impact of Regionalization of Treatment Partners on Roll-out of the Tanzania National HIV Treatment Plan. Poster. HIV Implementers Meeting, Windhoek, Namibia.

Mushinzimana et al. 2006. Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands. *Malaria Journal*.

Noor, AM, Zurovac, D, Hay, SI, Ochola, SA and RW Snow. 2003. Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. *Tropical Medicine and International Health* 8(10): 917-926.

Noor, AM, Gikandia, PW, Haya, SI, Muga, RO, and RW Snow. 2004. Creating spatially defined databases for equitable health service planning in low-income countries: the example of Kenya. *Acta Tropica* 91(3): 239-251.

Noor et al. 2008. Spatial prediction of Plasmodium falciparum prevalence in Somalia. *Malaria Journal* 7:159.

Noor et al. 2009. The risks of malaria infection in Kenya in 2009. *BMC Infectious Diseases* 9:180.

Noor et al. 2009. A spatial national health facility database for public health sector planning in Kenya in 2008. *International Journal of Health Geographics* 8:13.

Pande et al. 2008. Addressing diarrhea prevalence in the West African Middle Belt: social and geographic dimensions in a case study for Benin. *International Journal of Health Geographics* 7:17.

President's Commission on Federal Statistics., 1971. "Federal Statistics: Report of the President's Commission on Federal Statistics." Washington, DC.

PPGIS. 2010. Open Forum on Participatory Geographic Information Systems and Technologies.

Rambaldi, G, Kyem, PAK, McCall, M, and D Weiner. 2006. Participatory Spatial Information Management and Communication in Developing Countries. *Electronic Journal of Information Systems in Developing Countries* 25.

Renger, R, Cimetta, A, Pettygrove, S, and S Rogan. 2002. Geographic Information Systems (GIS) as an Evaluation Tool. *American Journal of Evaluation* 23(4): 469-479.

Rosero-Bixby, L. 2004. Spatial access to health care in Costa Rica and its equity: a GIS-based study. *Social Science & Medicine* 58: 1271–1284.

Sánchez et al. 2010. Monitoring Linked Epidemics: The Case of Tuberculosis and HIV. *PLoS ONE* 5(1): e8796.

Slocum, N. 2003. Participatory Methods Toolkit, A Practitioner's Manual, accessed April 2010.

Spencer, J and G Angeles. 2007. Kernel Density Estimation as a Technique for Assessing Availability of Health Services in Nicaragua. *Health Services and Outcomes Research Methodology* 7: 145-157.

Tanser, FC and D le Sueur. 2002. The application of geographical information systems to important public health problems in Africa. *International Journal of Health Geographics* 1:4.

Tanser, F. 2006. Geographical Information Systems (GIS) Innovations for Primary Health Care in Developing Countries. *Innovations: Technology, Governance, Globalization* 1(2): 68-79.

Tomlison, RF. 2003. Thinking about GIS: geographic information system planning for managers. ESRI Press.

United Nations. Handbook on geospatial Infrastructure in Support of Census activities. Series F. No. 103. UN Statistics Division, New York, New York.

United Nations Economic Commission for Europe (UNECE). Accessed September 2010. UN/LOCODE. Geneva: UNECE.

United Nations Programme on HIV/AIDS [UNAIDS]. 2008. A Framework for Monitoring and Evaluating HIV Prevention Programmes for Most-At-Risk Populations.

United States Agency for International Development (USAID). Accessed September 2010. Strategic Information for HIV/AIDS. Washington, D.C.: USAID.

United States Government Accounting Office. June 1992. Strategic Information Planning, Framework for Designing and Developing System Architectures.

United States President's Emergency Plan for AIDS Relief (PEPFAR). Accessed September 2010. Annex VI – Strategic information resources. Washington, D.C.: PEPFAR.

Uthman et al. 2009. A trend analysis and sub-regional distribution in number of people living with HIV and dying with TB in Africa, 1991 to 2006. *International Journal of Health Geographics* 8:65.

VanWey, L K, Rindfuss, RR, Gutmann, MP, Entwisle, B, and DL Balk. 2005. Confidentiality and spatially explicit data: Concerns and challenges. Proceedings of the National Academy of Sciences 102(43): 15337-15342.

Wilson, DP and S Blower. 2007. How far will we need to go to reach HIV-infected people in rural South Africa? *BMC Medicine* 5:16.

World Health Organization. 2009. A Guide to Monitoring and Evaluation for Collaborative TB/HIV Activities. Geneva: WHO.

World Health Organization. Accessed September 2010. Strategic Information. Geneva: WHO.

World Health Organization Health Metrics Network. January 2008. Framework and Standards for Country Health Information Systems, Second Edition

MEASURE Evaluation
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill NC 27517 USA
P: +1 919-445-9350
F: +1 919-445-9353
E: measure@unc.edu
www.measureevaluation.org

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) through Cooperative Agreement GHA-A-00-08-00003-00 and is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with ICF International, John Snow, Inc., Management Sciences for Health, Palladium, and Tulane University. The views expressed in this publication do not necessarily reflect the views of USAID or the United States government. MS-12-56

eCourse available online:
globalhealthlearning.org/course/geographic-approaches-global-health
eCourse available for download:
http://www.measureevaluation.org/resources/publications/ms-12-56

USAID FROM THE AMERICAN PEOPLE
PEPFAR
K4Health Knowledge for Health
MEASURE Evaluation