

Lecture 7: GIS Data Capture and Metadata

Section I: Overview of GIS Data Capture

As you read this lecture think about the type of data capture and how metadata can help with your final project.

Introduction

Back in the early days of the development of GIS getting the data to support your project was not a trivial task. Back in the 1990's, when GIS data was relative scarce (can be considered a data poor environment), a major component of most GIS projects were to generate your data; a task referred to as **primary data capture**. This work involves collecting data by surveying or in more recent days with GPS (to be discussed in the next lecture).

Today, with the proliferation of data on the world-wide-web (can be considered a data rich environment), many GIS projects rely on obtaining data over the web; referred to as **tertiary data capture**. Operationally, the main difference between different types of data capture is with primary data sources you have complete control over the accuracy and quality of the data. While with secondary and tertiary data sources you get what others have generated and the data may not be of sufficient accurate to support your project or application. Today, because many GIS projects rely on non-primary data sources, it has become extremely critical that the data collection process is described in detail to allow a potential user to evaluate whether the data can be utilized for a specific application. The role of **metadata** (discussed later) becomes particularly critical when using data that you did not create. Metadata will provide information about the accuracy of the data and its potentially suitability for your project.

Now let's consider how data is generated. Additionally, you need to realize that all non-primary data is initially primary created by a specific user or governmental agency. So you need to be mindful of the scale at which the data was developed as well as the degree of generalization.

Secondary Data Capture

Secondary data capture is first-hand data collection, which I suppose you can think of as old school GIS. There are a number of methods for secondary data collection. For entities defined based on a vector data model the most common method is to **digitize** data, which can convert features on a map into points, lines, and polygons.

Back in the early days of GIS, paper maps were converted into a digital format through digitizing

- today this is a less common activity. Digitizing is extremely tedious and labor intensive and involves hours of eye straining work just to convert a relatively simple map into a digital format. There are several methods of digitizing. If massive quantities of **data conversion** from a paper to digital format are required, a **digitizing tablet** (Fig. 1) is typically utilized; this is referred to as **hardcopy map digitizing**. Basically, this process involves placing a digitizing puck (a fancy multiple button mouse; Fig. 1) over a location on the tablet and presses one of the buttons on the puck to record the location of the feature of interest. An expensive tablet with a highly experienced operator can digitize the location of an object to within 0.006 mm! For a map at 1:24,000 this means that you can accurately locate an object on the map so that it is within 6 real world inches of its actual location. The map must be digitized under ideal conditions - for example, a small change in humidity can cause the map to shrink or expand altering results. If you want the highest possible accuracy associated with a data conversion project a digitizing tablet in a climate-controlled room is the way to go.



Figure 1. Digitizing tablet and digitizing puck. Material to be digitized is secured to the tablet and the puck is used like a mouse digitally capturing features of interest.

ArcGIS software also has a built in digitizing capability referred to as **heads up digitizing** (also referred to as on-screen digitizing) where the GIS uses a scan of a georeferenced raster image of a map and then the user generates vector data by clicking on features that define the entity. This

type of data capture will be explored in Assignment #7. In order to use this type of digitizing, registration marks or control points that define a known location must be present. For the USGS 7.5' quadrangle maps these marks are located at 2.5' minute intervals where parallels and meridians intersect and indicate the location of where registration marks will be present. The accuracy of digitized data will be governed by the resolution of your scan. If you have a high quality 1200 dpi (dots per inch) scan of a map it is possible to locate an object on the map so that it is within 20 real world inches of its actual location.

A map scanner (Fig. 2) can be used to capture paper media and convert it into a raster image. GIS systems have the capability of automatically converting raster to vector features - a process referred to as **vectorization** or raster to vector data conversion (Fig. 3). Digitizing, through this method also involves initially georeferencing the image. GIS tools can automatically convert a raster scan to vector lines. This process requires a very clean scan - any stray marks will lead to inaccuracies in the vectorization process, so scans should first be cleaned to remove spurious information. Even so the vector file produced by vectorization usually requires significant cleanup (or editing) after conversion because of incorrectly delineated features. If your image is not clean manual digitizing may be faster than to correct all the errors introduced by the vectorization process. The accuracy of the vectorized data is only as good as the input raster image, which is comparable to that associated with heads up digitizing. ArcGIS software includes numerous extensions that augment the functionality of the software. An extension that can be used to clean up raster images for vectorization is called the ArcScan extension (From ESRI). A critical component to consider when taking a scanned image and importing it into a GIS system is to as accurately as possible scan the image to avoid distortions.

Additionally, it is possible to convert vector data into a raster. Some accuracy is always lost during the data conversion process when going either way. Because this process is almost never perfect it is possible to distort the scanned image to rectify it to eliminate distortions and inaccuracies. This procedure is referred to as **rubbersheeting**, which is the basis of correcting aerial photographs that we will discuss in more detail later in the course (Fig. 4). This process involves the selection of **control points** whose location is well-defined and can include USGS survey markers or road intersections (Fig. 5).

Tertiary Data Capture

Today there are large amounts of data now available, most of which is free or nearly free. Always check the internet for existing data before creating it. Also, just because data exists, it does not necessarily mean that it is suited for your application. Always check the metadata and

know exactly what you are getting. Additionally, not all GIS data is necessarily on the internet. Some private and governmental organizations that have invested vast human and capital resources in creating a GIS dataset may not be willing to openly share their data. However, if you know individuals within an organization, which you can do by going to regional GIS conferences or participating in a local GIS users group, and can assure them that you will not use their data for commercial purposes, *i.e.* turn around and sell it, you can generally get a hold of their data if you are persistent.

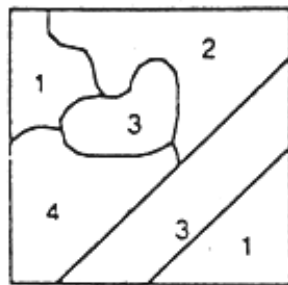


Figure 2. A map scanner where the map is placed on a roller and is scanned in a way analogous to a fax machine.

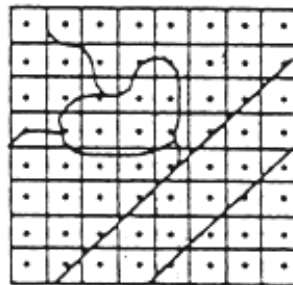
Tertiary data providers can be grouped into two categories: private and governmental. Governmental providers are at the local, state, and national level. The level of GIS data and expertise varies greatly at the local level generally depending on the size of the jurisdictions. Major cities can have a dedicated GIS department, which is constantly creating and maintaining GIS data. For example, check out the data available from the City of Austin.

<ftp://ftp.ci.austin.tx.us/GIS-Data/>

VECTOR TO GRID CONVERSION



Vector data set containing polygons with associated attributes.



A grid with the desired cell size is superimposed. The polygons which contain the centers of each of these grid cells are then determined.



1	2	2	2	2	2	2	2
1	1	2	2	2	2	2	2
1	1	2	3	2	2	2	3
1	1	3	3	2	2	3	3
4	4	4	4	4	3	3	1
4	4	4	4	3	3	1	1
4	4	4	3	3	1	1	1
4	4	3	3	1	1	1	1

The values of the grid cells become the values of the attributes of the polygons which contain them.

GRID TO VECTOR CONVERSION

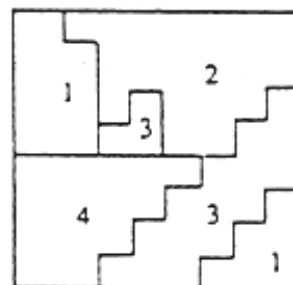
1	2	2	2	2	2	2	2
1	1	2	2	2	2	2	2
1	1	2	3	2	2	2	3
1	1	3	3	2	2	3	3
4	4	4	4	4	3	3	1
4	4	4	4	3	3	1	1
4	4	4	3	3	1	1	1
4	4	3	3	1	1	1	1

Gridded satellite classification with class numbers contained each grid cell.



1	2	2	2	2	2	2	2
1	1	2	2	2	2	2	2
1	1	2	3	2	2	2	3
1	1	3	3	2	2	3	3
4	4	4	4	4	3	3	1
4	4	4	4	3	3	1	1
4	4	4	3	3	1	1	1
4	4	3	3	1	1	1	1

Borders between differing class numbers are located.



Polygons are then generated for each contiguous area by storing the x, y coordinates of the points along these borders.

Figure 3. Vector to raster and raster to vector data conversion.

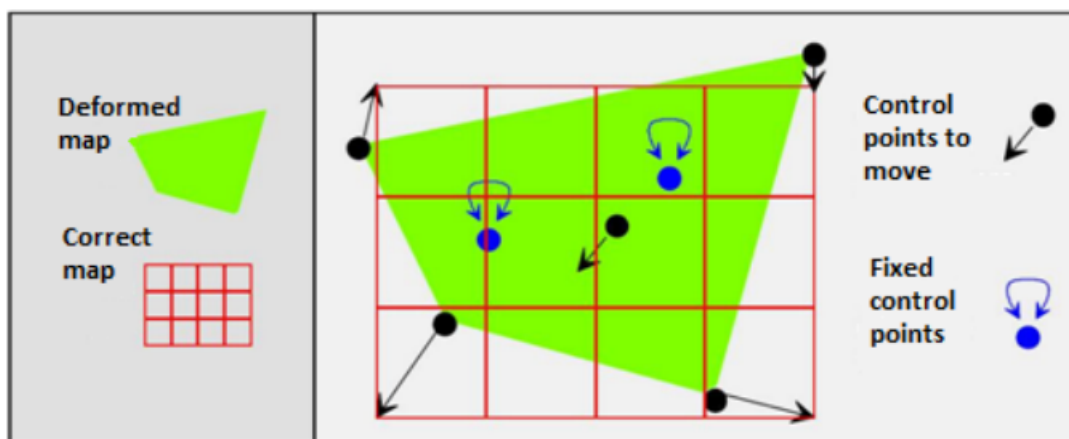


Figure 4. Rubbersheeting involves the stretching of control points to a new corrected position.



Figure 5. USGS survey marker that can be used as a control point. From Penn State Geography program.

You can find many webpages with free access to GIS data. Nowadays, GIS has become an integral part of governments to maintain a database of everything from boundaries, infrastructure, census, etc. Most governments have their own official GIS clearinghouses. Some clearinghouses collect as much data as possible in one single location, and others have very specific purposes. The United States has several webpages dedicated to the dissemination of GIS data. The official census designated areas are always available for download. This is a database of all the official boundaries in the country <https://www.census.gov/geo/maps-data/data/tiger-line.html>. The National Map server is a source to find datasets that can serve as basemaps such as topographic maps, elevation products, hydrography, etc. <https://viewer.nationalmap.gov/basic/>. The Geospatial Data Gateway maintained by the USDA is a database of all datasets related to natural resources and the environment such as land cover, soil types, precipitation, <https://gdg.sc.egov.usda.gov/>.

Some states have their own GIS databases in which some datasets are replicated from the national databases, and others are results of state-run projects. Texas, for example, has the Texas Natural Resources Information System (TNRIS, <https://tnris.org/>). Aside from the usual statewide political datasets, this is a good source for orthoimagery and lidar data specific to the state. In the same manner, there is the Mississippi Clearinghouse (<http://www.gis.ms.gov/portal/download.aspx?dom=&x=1536&y=720&browser=Netscape>), the New York Clearinghouse (<http://gis.ny.gov/>), etc.

Within each state there are also departments with their own GIS clearinghouses . For example, the Texas Commission on Environmental Quality (<http://www.tceq.state.tx.us/gis/download-tceq-gis-data>), the Texas Parks and Wildlife Department (<http://tpwd.texas.gov/gis/data>).

There are also other websites with specific purposes such as natural conservation, endangered species protection, etc. The North American Atlas is an organization for the protection of the environment of the U.S., Canada, and Mexico (<http://www.cec.org/tools-and-resources/north-american-environmental-atlas/map-files>). The U.S. Fish and Wildlife Service maintains the Environmental Conservation Online System (<https://ecos.fws.gov/ecp/report/table/critical-habitat.html>) as a way to keep track of the geographic distribution of protected species.

You can always find sources of GIS data online using Google to search for what you want. Most of this data is free to get and use as long as you give credit to the sources. There are, of course, private companies that sell GIS data as a means of saving the time and effort to search for it, or also because they carry out their own data collection and maintenance of databases.

II. Overview of Metadata

Metadata can be described as “data about data” because it describes the contents and usefulness of electronic files so that they can be easily shared across the Internet. In GIS metadata is an essentially requirement if a user will engage in secondary data collection. Simply put, it is critical to know exactly where data comes from and how it was developed for it to be used in a purposeful manner. Geospatial data is more valuable when it comes accompanied of good metadata. There are a number of characteristics about geospatial data that need to be specified for users that have not used to data before. Characteristics such as the coordinate system, datum, geographic extent, and attribute fields related to the data are part of the metadata of geospatial files. In addition, knowing the main purpose of the data and who developed it is extremely useful to seek help if problems are encountered. Overall, metadata ensures that sharing

information about how geospatial data is created, organized, and updated can transfer geospatial knowledge.

Metadata standards have to exist because it would be close to impossible to understand data that flows across the globe and is modified by a countless number of people. Currently, there are two main standards that are used to create metadata about geospatial data.

The Federal Geographic Data Committee (FGDC) developed the **Content Standard for Digital Geospatial Metadata** (CSDGM) in 1998. This standard was created to facilitate the creation and sharing of metadata by ensuring that a common template is utilized and that the most critical information is included. The CSDGM has become the most widely used standard in the United States for the creation of geospatial metadata. The FGDC defined metadata as “a file of information, usually presented as an Extensible Markup Language (XML) document, which captures the basic characteristics of a data or information resource. It represents the who, what, when, where, why and how? of the resource. Geospatial metadata are used to document geographic digital resources such as Geographic Information System (GIS) files, geospatial databases, and earth imagery. A geospatial metadata record includes core library catalog elements such as Title, Abstract, and Publication Data; geographic elements such as Geographic Extent and Projection Information; and database elements such as Attribute Label Definitions and Attribute Domain Values.

The **International Organization for Standardization** (ISO) developed the ISO19115 standard for geospatial data in 2004 to create a more universal standard for metadata creation. It also created the ISO19139 standard specifically for the formatting and exchange of metadata. This standard utilizes XML as the preferred formatting style. XML is the most widely used format to encode documents to be transferred over the Internet.

The main aspects of data included in a geospatial metadata file are the following:

- Identification: Title, area covered, theme, time period, restrictions.
- Data Quality: Accuracy, completeness, consistency, lineage.
- Spatial Data Organization: Vector or Raster? Type and number of elements.
- Spatial Reference: Projection, datum, grid system, coordinate system.
- Entity and Attribute Information: Features, Attributes, Attribute Values.
- Distribution: Author, formats, sources.

Depending on the nature of the spatial data and the author, some areas may be omitted.

Even though **XML** is the preferred format for metadata, metadata can be obtained in other formats such as **HTML**, **plain text**, and even in **frequently asked questions (FAQ)** format. There are software packages aimed at producing metadata that conform to the previously mentioned standards and they are recommended by the FGDC for individuals or companies that create geospatial data. Some packages are free and others such as ESRI's ArcCatalog are part of commercial software suites that specifically work with geospatial data.

Examples of how metadata look depends on format are shown on the following pages. Realize that even though metadata can come in different formats that the same basic information is conveyed no matter the format. I would like you to answer questions based on an exploration of the metadata below.

XML format is more appropriate for electronic devices because it is formatted in machine-readable language and it looks somewhat confusing for an individual. The same information can be presented in more user-friendly formats so that the information can be easily read. See next page.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE metadata (View Source for full doctype...)>
- <metadata>
- <idinfo>
- <citation>
- <citeinfo>
  <origin>U.S. Geological Survey</origin>
  <pubdate>2007</pubdate>
  <title>binat_hospitals_bw</title>
  <geoform>vector digital data</geoform>
  <onlink>http://borderhealth.cr.usgs.gov</onlink>
</citeinfo>
</citation>
- <descript>
  <abstract>This binational hospitals dataset was developed by the U.S. Geological Survey's (USGS) Border Environmental Health
  barriers for environmental issues. The major issues surrounding the U.S.-Mexico border involve economics and population growth
  environment, binationally integrated baseline datasets that portray the status of the landscape are needed. The primary object
  datasets for display and further analysis within a geographic information system (GIS) framework. Datasets created for the B
  Committee Issues Team, Woodward and Durall used surface-water drainage basins as the primary basis for defining and delin
  region, local and regional BEHI datasets have been clipped to this boundary. This feature class contains the locations of hospit
  along the entire Mexican side of the border, but on the U.S. side data are only available in Texas. Data for Mexico, excluding Co
  from the Instituto Municipal de Investigación y Planeación (IMIP), Cd. Juárez in 2005. Data for Texas were obtained from the
  Nacional de Salud Pública and IMIP), and 2003 for DSHS. The final dataset was created by the USGS's BEHI (http://borderhe
  <purpose>This dataset was created for the USGS's BEHI, and displays the location of hospitals along the U.S.-Mexico Border. Fu
  accurate, impartial, scientific information to strike a balance among human needs for resource utilization, environmental qualit
  <supplinf>Additional information about the BEHI can be found at http://borderhealth.cr.usgs.gov and datasets are available for v
</descript>
- <timeperd>
- <timeinfo>
  - <mdattim>
    - <sngdate>
      <caldate>Data in Mexico - 2005</caldate>
    </sngdate>
    - <sngdate>
      <caldate>Data in Texas, United States - 2003</caldate>
    </sngdate>
  </mdattim>
  </timeinfo>
  <current>ground condition</current>
</timeperd>
- <status>
  <progress>In work</progress>
  <update>Unknown</update>
</status>
- <spdom>
  - <bounding>
    <westbc>-117.021022</westbc>
    <eastbc>-96.837772</eastbc>
    <northbc>32.925976</northbc>
    <southbc>25.650011</southbc>
  </bounding>
</spdom>
- <keywords>
- <theme>
  <themekt>Library of Congress</themekt>
  <themekey>Environmental Health</themekey>
  <themekey>Water</themekey>
  <themekey>Hospital and Community</themekey>
  <themekey>Hospital Care</themekey>
</theme>
- <theme>
```

The main format for web pages is HTML, which stands for HyperText Markup Language. Metadata in HTML format is illustrated on the following page.

binat_hospitals_bw

Metadata also available as

Metadata:

- [Identification Information](#)
- [Data Quality Information](#)
- [Spatial Data Organization Information](#)
- [Spatial Reference Information](#)
- [Entity and Attribute Information](#)
- [Distribution Information](#)
- [Metadata Reference Information](#)

Identification Information:

Citation:

Citation Information:

Originator: U.S. Geological Survey
Publication Date: 2007
Title: binat_hospitals_bw
Geospatial Data Presentation Form: vector digital data
Online Linkage: <http://borderhealth.cr.usgs.gov>

Description:

Abstract:

This binational hospitals dataset was developed by the U.S. Geological Survey's (USGS) Border Environmental Health Initiative (BEHI) (<http://borderhealth.cr.usgs.gov/>). Mexico border involve economics and population growth that present challenges to environmental management and natural resource planning. To monitor trends and analyze U.S. Geological Survey's Border Environmental Health Initiative is to develop integrated, environmental resource and human health datasets for display and further analysis. Datasets created for the BEHI follow watershed boundaries as defined by Woodward and Durrall (1996). As part of the U.S.-Mexico Border Field Coordinating Committee I: shared-water resources perspective. In order to provide integrated datasets in the border region, local and regional BEHI datasets have been clipped to this boundary. This feature class contains the locations of hospitals along the U.S.-Mexico Border, and is being completed in phases by subareas defined by the USGS's BEHI. Currently, Data for Mexico, excluding Cd. Juárez in Chihuahua, were obtained from the Instituto Nacional de Salud Pública in 2006. However, hospitals in Cd. Juárez were obtained from Services (DSHS) in 2004. The dates for the ground condition of the datasets are 2005 (for the Instituto Nacional de Salud Pública and INIP), and 2003 for DSHS. The final c

Purpose:

This dataset was created for the USGS's BEHI, and displays the location of hospitals along the U.S.-Mexico Border. Furthermore, this dataset was created to provide resource needs for resource utilization, environmental quality, and human health.

Supplemental Information:

Additional information about the BEHI can be found at <http://borderhealth.cr.usgs.gov> and datasets are available for viewing through an Internet Mapping Service (fol

Time Period of Content:

Time Period Information:

Multiple Dates/Times:

Single Date/Time:

Calendar Date: Data in Mexico - 2005

Single Date/Time:

Calendar Date: Data in Texas, United States - 2003

Currentness Reference: ground condition

Status:

Progress: In work

Maintenance and Update Frequency: Unknown

Plain text file is illustrated below. This file has no formatting and is typically indicated in a Windows system with the .txt suffix.

Identification_Information:
Citation:
Citation_Information:
originator: U.S. Geological Survey
Publication_Date: 2007
Title: binat_hospitals_bw
Geospatial_Data_Presentation_Form: vector digital data
Online_Linkage: <http://borderhealth.cr.usgs.gov>
Description:
Abstract:
This binational hospitals dataset was developed by the U.S. Geological Survey's (USGS) Border Environmental Health Initiative
Datasets created for the BEHI follow watershed boundaries as defined by Woodward and Durrall (1996). As part of the U.S.-Mexico
This feature class contains the locations of hospitals along the U.S.-Mexico Border, and is being completed in phases by subare
Data for Mexico, excluding Cd. Juárez in Chihuahua, were obtained from the Instituto Nacional de Salud Pública in 2006. Howev
Purpose: This dataset was created for the USGS's BEHI, and displays the location of hospitals along the U.S.-Mexico Border. Furt
Supplemental_Information: Additional information about the BEHI can be found at <http://borderhealth.cr.usgs.gov> and datasets are
Time_Period_of_Content:
Time_Period_Information:
Multiple_Dates/Times:
Single_Date/Time:
Calendar_Date: Data in Mexico - 2005
Single_Date/Time:
Calendar_Date: Data in Texas, United States - 2003
Currentness_Reference: ground condition
Status:
Progress: In work
Maintenance_and_Update_Frequency: Unknown
Spatial_Domain:
Bounding_Coordinates:
West_Bounding_Coordinate: -117.021022
East_Bounding_Coordinate: -96.837772
North_Bounding_Coordinate: 32.925976
South_Bounding_Coordinate: 25.650011
Keywords:
Theme:
Theme_Keyword_Thesaurus: Library of Congress
Theme_Keyword: Environmental Health
Theme_Keyword: Water
Theme_Keyword: Hospital and Community
Theme_Keyword: Hospital Care
Theme:
Theme_Keyword: Border Health
Theme_Keyword: Human Health
Theme_Keyword: US-Mexico
Theme_Keyword: Hospital
Theme:
Theme_Keyword_Thesaurus: ISO 19115 Topic Category
Theme_Keyword: ImageryBaseMapsEarthCover
Theme_Keyword: Imagery and Base Maps
Place:
Place_Keyword: Mexican-American Border Region
Place_Keyword: United States of America
Place_Keyword: Mexico
Place_Keyword: Rio Grande
Place_Keyword: US
Place_Keyword: U.S.
Place_Keyword: U.S.-Mexico
Place_Keyword: US-Mexico
Place_Keyword: United States-Mexico
Access_Constraints: None, the data are available for viewing at <http://borderhealth.cr.usgs.gov>.
Use_Constraints: There is no guarantee concerning the accuracy of the data. Users should be aware that temporal changes may have oc
Point_of_Contact:
Contact_Information:
Contact_Organization_Primary:
Contact_Organization: U.S. Geological Survey
Contact_Person: Public Information Officer
Contact_Address:
Address_Type: mailing and physical address
Address: 8027 Exchange Dr
City: Austin
State_or_Province: Texas
Postal_Code: 78754
Country: USA
Contact_Voice_Telephone: 512-927-3500
Contact_Facsimile_Telephone: 512-927-3590
Contact_Electronic_Mail_Address: gs-w-txpublic-info@usgs.gov
Data_Set_Credit: This dataset was created by the USGS BEHI with data obtained from the Texas Department of State Health Services, t
Native_Data_Set_Environment: Microsoft Windows XP Version 5.1 (Build 2600) Service Pack 3; ESRI ArcCatalog 9.3.1.3000

Finally, Frequently Asked Question Format (FAQ) provides information most common requested by data users.

What does this data set describe?

Title: binat_hospitals_bw

Abstract:

This binational hospitals dataset was developed by the U.S. Geological Survey's (USGS) Border Environmental Health Initiative (BEHI) (<http://borderhealth.cr.usgs.gov/>) issues surrounding the U.S.-Mexico border involve economics and population growth that present challenges to environmental management and natural resource planning. To address these issues, a detailed landscape are needed. The primary objective of the U.S. Geological Survey's Border Environmental Health Initiative is to develop integrated, environmental resource and human health datasets created for the BEHI follow watershed boundaries as defined by Woodward and Durall (1996). As part of the U.S.-Mexico Border Field Coordinating Committee, the extent of the border area from a shared-water resources perspective. In order to provide integrated datasets in the border region, local and regional BEHI datasets have been created. This feature class contains the locations of hospitals along the U.S.-Mexico Border, and is being completed in phases by subareas defined by the USGS's BEHI. Currently, data for Mexico, excluding Cd. Juárez in Chihuahua, were obtained from the Instituto Nacional de Salud Pública in 2006. However, hospitals in Cd. Juárez were obtained from the Texas Department of State Health Services (DSHS) in 2004. The dates for the ground condition of the datasets are 2005 (for the Instituto Nacional de Salud Pública and I

Supplemental Information:

Additional information about the BEHI can be found at <http://borderhealth.cr.usgs.gov/> and datasets are available for viewing through an Internet Mapping Service (follow

1. How should this data set be cited?

U.S. Geological Survey, 2007, binat_hospitals_bw.

Online Links:

- <http://borderhealth.cr.usgs.gov/>

2. What geographic area does the data set cover?

West_Bounding_Coordinate: -117.021022

East_Bounding_Coordinate: -96.837772

North_Bounding_Coordinate: 32.925976

South_Bounding_Coordinate: 25.650011

3. What does it look like?

4. Does the data set describe conditions during a particular time period?

Calendar_Date: Data in Mexico - 2005

Currentness_Reference: ground condition

5. What is the general form of this data set?

Geospatial_Data_Presentation_Form: vector digital data

6. How does the data set represent geographic features?

a. How are geographic features stored in the data set?

This is a Vector data set. It contains the following vector data types (SDTS terminology):

- Entity point (125)

b. What coordinate system is used to represent geographic features?

The map projection used is Albers Conical Equal Area.

Projection parameters:

Standard_Parallel: 27.000000

Standard_Parallel: 33.000000

Readings

GIS Commons webpage; Chapters 2.

Campbell, J and Shin, M., 2011, Essentials of Geographic Information Systems. Chp. 3.2, 3.3.

Terms

Primary Data Capture	Secondary Data Capture	Tertiary Data Capture	Metadata
Digitizing Tablet	Heads Up Digitizing	Data Conversion	Vectorization
Rasterization	Control Points	Rubbersheeting	Benchmarks
CSDGM	ISO		

Concepts

Clearly know the difference between primary, secondary, and tertiary data capture

Know about the different techniques for GIS secondary data capture

Make a list of active GIS websites that you can potentially obtain data from for your final project

What are the different metadata formats?

Make a list of the different types of information that you can find within a metadata file

DATA CAPTURE HOMEWORK

1. Compare the advantages and disadvantages of secondary and tertiary data capture. Discuss potential issues associated with accuracy and cost of data acquisition.

2. Comment on the importance of metadata especially associated with using GIS data from tertiary sources.

3. Describe the pitfalls with relying on a GIS layer created by vectorization.

METADATA HOMEWORK

1. When was the data published? _____

2. What organization originated/developed the data set? _____

3. Find the website that distributes the website and print out the web page and attach to the back of the assignment.

4. The bounding coordinates determine the geographic extent that the layer covers. This coordinates are given in decimal degrees. Realize east-west coordinates are longitude and north-south coordinates are latitude.

West Bounding Coordinates _____

East Bounding Coordinates _____

North Bounding Coordinates _____

South Bounding Coordinates _____

5. If you want to talk to someone about this data set what number would you call?

6. What type of map projection is used for this layer? _____

7. What is the latitude of the standard parallels associated with this projection?

8. What class of projection is this? Circle one

Conic

Cylindrical

Azimuthal

Pseudo-cylindrical

9. Is this projection tangent or secant ?

10. What is the reference that defines the watershed boundaries used to define the data layer?

11. What type of data model is used to define the data? _____

12. Indicate the use constraints that exist for this data.