Lesson 7: Data Sources

## INTRODUCTION

This lesson provides information to help you learn about data sources. The lesson discusses creating geospatial data and the various types of data available. The lesson also explains both automatic and manual digitizing with information, images and a comparison of the two methods. The lesson ends with an explanation of Global Navigation Satellite Systems (GNSS).

## LESSON OBJECTIVES

By the end of this lesson, you will be able to:

1. Define key terms relating to digitizing, data sources, and geocoding.
2. Build and collect primary spatial data by way of scanning, digitizing, geocoding, or similar methods.
3. Identify and use data from public sources.
4. Create a map.

## LEARNING SEQUENCE

|  |  |
| --- | --- |
|  | |
| Required Reading | Read the following:  Data Sources   * Creating Geospatial Data * Digitizing * Manual Digitizing * Automatic Digitizing * Global Navigation Satellite Systems |
| Assignments | Complete the following assignments:   * Lab 7: Data Sources * Quiz 7: Data Sources |

## INSTRUCTION

## Creating Geospatial Data

### Why Create Geospatial Data?

## There are many reasons why you might want to create your own geospatial data. The main reason is that the data you need has not been created by anyone else. Even though GIS and digital data have been around for a long time, there is still a large need for the creation of new data covering many different facets of our world at many different scales.

## A couple of additional reasons that you may want to create geospatial data are that the data you need is on a paper map and needs to be converted to digital format so it can be used in your GIS software, and the data available may be a different scale than you need. For example, let’s say you want to show a map of where all the state parks are in Virginia. However, all of the park data sets are extremely detailed, and look very poor at a statewide scale because of the intricate detail of the park boundaries.

## In this case, you would need to create a new park dataset suitable for display at a smaller scale, or, simplify the existing dataset so that it can display better at the smaller scale. Another reason to create new geospatial data is the possibility that the data available is too old for your purposes and you need an updated version of the data. There are many other reasons why you would want to create geospatial data, but these are the most common reasons that you should be aware of.

### Data Sources

## There are two types of data sources that you can derive digital geospatial data from: hardcopy, and digital. Each of these data sources has their own positive and negative aspects when it comes to storing data.

### Hardcopy Data

## Hardcopy data, such as data stored on paper as maps, was the most common medium of storage for GIS until the 1980’s. Here are some positive and negative aspects of hardcopy storage.

#### Positives

## The positive aspects of hardcopy storage is that it is a stable medium which means that it cannot get a virus, run out of battery power, or crash, like a computer. Hardcopy data is also cheap to produce, is easily portable, and does not require any specialized hardware or software to use. This provides a very low barrier for use, and makes a paper map easy to use and understand by a very wide range of users. Hardcopy data is semi-permanent so long it must be properly stored.

#### Negatives

## The negative aspects of hardcopy data are that it is not reusable because it cannot be easily updated. A hardcopy map, for instance, is a static representation, and can only be updated by printing over the existing map, which is often not desirable.

### Digital Data

Digital data has become a common storage medium for GIS data since the 1980s as computers became more widely used and storage capacities grew larger and prices dropped significantly. Here are some positive and negative aspects of storage data digitally.

#### Positives

The positive aspects of digital data are that it is stable so long as it is properly maintained, backed up, and occasionally transferred to a modern storage medium. Digital storage is now extremely cheap and can store massive amounts of information in a very small space. Digital data is semi-permanent. What is meant by that is that the data structures must be set up to allow for archiving of historical values information, however, this is often not the case, therefore previous data is often overridden when values are updated and there is no way to see the previous version or value. Digital data is extremely flexible. Digital data can easily be updated and can be used for many different purposes.

#### Negatives

Negative aspects of digital data are that it is not portable without specialized hardware and software. This significantly raises the barrier of entry for users, and the hardware and software may not be easily understood by the users either. Additionally, the data may be tied to specific systems that may deprecate over time, so there is a need to continuously update the data into newer hardware and software systems.

## Digitizing

### What is Digitizing?

Digitizing is the process of converting coordinates of features from a data source, such as a paper map, into a digital format. Digitizing can be thought of as tracing an object. The general idea is that we will use an input data source, such as an aerial photograph rectified to earth-based coordinates, and then use an input device to trace and record the features from the map as a vector feature.

There are two methods of digitizing: manual digitizing and automatic digitizing. Manual digitizing is when a person digitizes the features from the source data manually. This process is often time-consuming, tedious, and prone to human error, however, it allows the human to intelligently determine where to digitize, and what to digitize. The second method of digitizing, automatic digitizing is where a computer program is trained to identify features on the input data source, and then automatically traces those features.

Automatic digitizing can save a significant amount time; however, the features need to be easily identifiable on the input data source. Training the software to identify the features may take quite a bit of time and a human may still need to clean up errors in the automatic digitizing results.

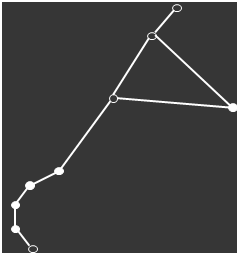
### Anatomy of a Digitized Object

When an object is digitized into the computer from the source data set, it is comprised of two parts: a node, and a vertex. A node is a start or end point of a line segment, or, a point. A vertex, is an intermediate point of a segment, and must exist between two nodes.

As an example, if we look at the example of a digitized object, we can see that there are four nodes, and five vertices. As we know that nodes determine the end and beginning of a segment, and vertices are continuations of the same line segment, we can identify four separate digitized line segments in this figure.

Starting at the bottom, we have a node, followed by four vertices, that terminates at the base of the triangle. That is one line segment. The second line segment is between the two nodes on the small leg of the triangle.

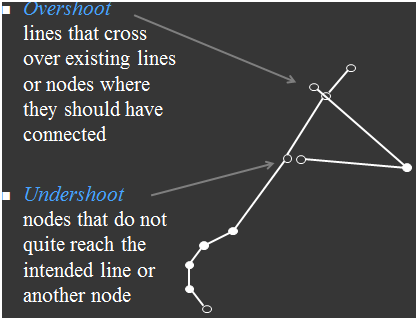
The third line segment, are the two nodes on the short leg of the triangle, that go through the vertex at the other end of the triangle. The fourth and final segment is the short line that offshoots from the top node of the triangle to the end of the short line.



### Common Digitizing Errors

When digitizing manually, there are two common digitizing errors. The first common digitizing error is known as an overshoot. An overshoot is where lines cross over existing lines and nodes where they should’ve actually connected. For instance, imagine if we are digitizing pipelines off of a map and we have two pipes that connect in a T intersection.

However, when digitizing, the pipe that is forming the stem of the T does not connect exactly with the other pipeline; instead, it passes over and through it in the digital data making a cross instead. This is an example of an overshoot. A undershoot is when nodes do not quite reach the intended line or node. Going back to the pipeline example, if the two pipelines do not intersect because the second pipeline was terminated just short of the pipeline was supposed connect to, this would be considered a undershoot.



### Digitizing Process Steps

When performing digitizing, there are five steps that you should follow.

Step 1: Create, or use an existing data set to store the digitized data. You should select the geometry type of the data set.

Step 2: Load the source data into the GIS program.

Step 3: Register the source to earth-based coordinates if it is not already.

Step 4: Manually, or automatically, digitize the features into the digital data set.

Step 5: Save the newly digitized features.

## Manual Digitizing

### What is Manual Digitizing?

This section defines manual digitizing in more detail, and shows an example of a manual digitizing task.

Manual digitizing is the human guided capture of features from a map image or source. There are two methods of manual digitizing: on-screen digitizing, and hardcopy digitizing. On-screen digitizing is probably the most common form of digitizing today and is where a source is scanned or downloaded into the computer’s memory and loaded into digitizing software.

The mouse and keyboard are used to digitize features while the sources are fed on the screen. Hardcopy digitizing is where the original source is taped to a digitizing table which is connected to a computer. The digitizing puck, which is similar to a mouse, is used to digitize features from the hardcopy source into the computer. The purpose of the digitizing puck is to feed coordinates into the computer, based on its location on the digitizing table, which would be the source’s coordinate system.

Watch, Manual Digitizing (1:46), for a demonstration of on-screen manual digitizing.

Video Transcript: This video shows an example of on-screen manual digitizing. Note that as we are digitizing the centerlines of the roads, it is being stored into a line vector data set. Additionally, after the road is digitized, it provides an attribute with the name of the road. As we digitize the houses, it is being stored into a separate polygon vector data set. It is also being attributed after the polygon is completely digitized. At the end of the digitizing process, don’t forget to save your work, otherwise all of your digitizing will be lost.

## Automatic Digitizing

### What is Automatic Digitizing?

Automatic digitizing is the computer guided capture from a map image or source. This is different from manual digitizing where it is a human guided capture. The way in which the computer performs automatic digitizing is that it uses algorithms to identify and digitize the features. While these algorithms are often quite powerful, it often requires an initial training period that may take time. Some of the algorithms available in automatic digitizing software can learn how to identify the features and is supposed to trace based on full color scans, instead of requiring an initial binary classification, as you will see in step one.

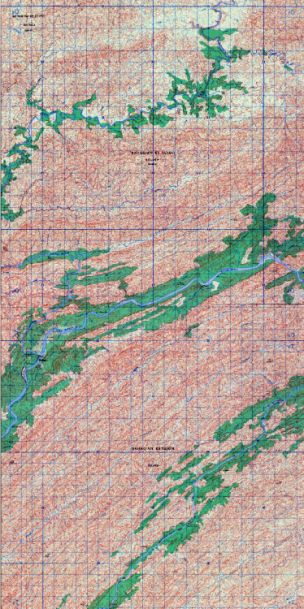
Note - Automatic Digitizing: This course does not include a lab for automatic digitizing so a short case study will be discussed.

### Example of Automatic Digitization

This image is an example of automatically digitizing contour lines from a contour map. What you are seeing in the image is actually three contour maps that have been mosaicked, or pasted together. As you can see there are quite a few red contour lines on this map which would take quite a long time for someone to manually digitize.

In an effort to save time we will automatically digitize the contour lines from this map. This map is a good candidate for automatic digitizing because the contour lines are drawn in a strongly contrasting red color and no other features on the map are the color red. This helps the computer easily identify and isolate the contour lines of the map.

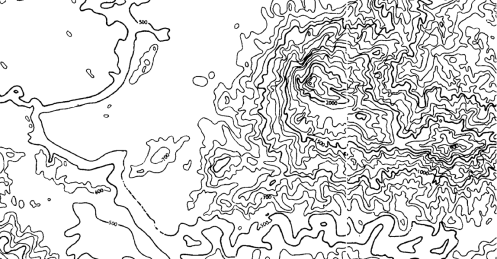
**Example: Automatic Digitization of Contour Lines from Contour Map**



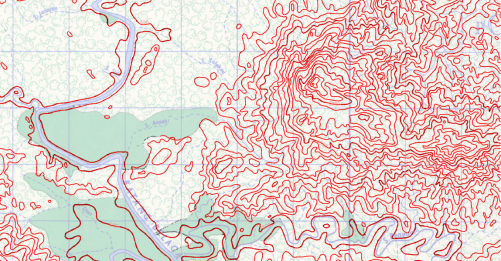
### The Process of Automatic Digitizing

There are four steps outlined in detail to illustrate the process of that occurs when applying the automatic digitizing process. The four steps include: isolating contour lines, running an automatic digitizer, cleaning up contours, and assigning line contour values.

#### Step 1: Isolate Contour Lines

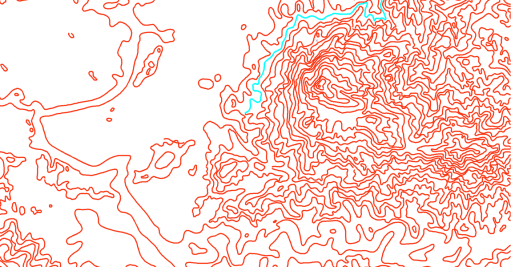
In order to assist the computer and isolating the contour lines the digitized contour map has been classified into a binary image. Wherever there was red on the map, is now black, and where the color red did not exist on the original map, it is now white. This initial step in the automatic digitizing process and is critical for ensuring that the computer can easily distinguish features from the map. It is important to note, that other digitizing software does not need the user to isolate the features for it, as it can work with selected colors.

#### Step 2: Run Automatic Digitizer

The second step of automatic digitizing is to run the automatic digitizing algorithm. The red lines now shown on the map are the actual digitized vector lines. We cannot see the original raster representation of the contour lines, as they are underneath the digitized lines; this means that the automatic digitizing algorithm did a good job of tracing the contour lines.

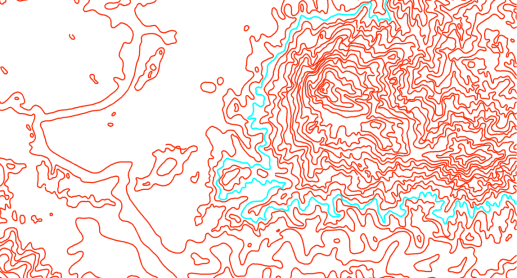
Initial Results of Automatic Digitization

#### Step 3: Cleanup

The third step in automatic digitizing is to clean the results of step two. On this particular map, if the contour lines make sharp turns, the contour lines become disconnected. For instance, the blue contour line should connect with the red contour line that it touches toward the center of the screen.

However, as this is a significant change of angle, the automatic digitizer considers this to be separate contour lines. Therefore, it is up to the user to either: modify digitizing parameters so that this may be caught, and run step two again, or, manually connect the contour lines.

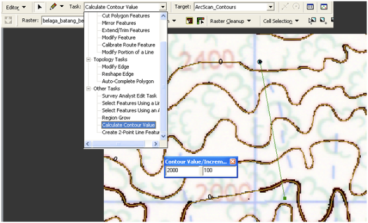
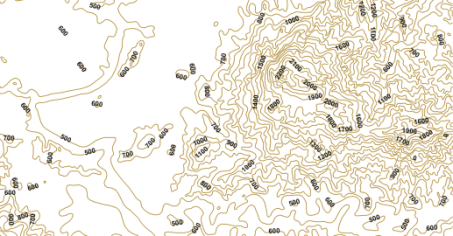
Example of Cleanup Task Connecting Disconnected Contours image 1

This is what the corrected contour line looks like. In this case, the user opted to manually connect the contour lines. Now that that error has been corrected, the user would continue to review the automatic digitizing results to find the other errors, and correct those.

Example of Cleanup Task

Connecting Disconnected Contours

#### Step 4: Assign Contour Line Values

The fourth step is to assign the contour line values into the attribute table of the newly digitized contour lines. In this case, a tool was downloaded that makes it easy to assign multiple contour lines at once, by drawing a line across multiple contour lines, and then specifying the first contour line it crosses contour value, and how much it should increment each additional contour line that passes over. So, in this case, the contour line at the bottom would be assigned a value of 2000, the line above it 2100, line above that in the line above that 2220 and 2300 respectively, and finally, 2400 as the last, top contour line is crossed.

This is the result of the automatically digitized contour lines with the contour values labeled.

This image shows the completed digitized contour lines overlaid on top of the original contour map. As you can see, the digitized contour lines follow the original contour lines extremely close and it does a good job of saving the manual digitizer time.

### Manual vs. Automatic Digitizing

Now that we’ve seen both the automatic and manual digitizing methods, let us take a moment to discuss positive and negative aspects of the two methods.

#### Manual Digitizing

## For most of your digitizing work, if it is sufficiently small in scope, manual digitizing will be the preferred method.

Positive Aspects

* Manual digitizing is typically sufficiently accurate as humans can interpret maps easier than computers.
* As digitizing, is not typically a job that requires highly skilled operators, the training period is often short, and the labor is cheap.
* For most of your digitizing work, if it is sufficiently small in scope, manual digitizing will be the preferred method.

Negative Aspects

* The negative aspects of manual digitizing, is that the map scale impacts the accuracy. Different operators may work better, or worse, at different scales, as some maps may have such fine detail, that the operator has a hard time accurately digitizing the intended feature.
* The hardware that the operators are using may also impact accuracy, if the operators are not provided with capable hardware and software.
* Different operators will digitize at different qualities, and may interpret what is being digitized differently. This may lead to inconsistent digitizing results.

#### Automatic Digitizing

## If the digitizing work covers an extremely large area, and the source has easily distinguishable features, automatic digitizing can provide a large time, and money savings.

Positive Aspects

* Automatic digitizing is good for large projects that would take a large amount of staff, and time to complete.
* There is no initial operator error, but the results will require manual editing, as we just saw.
* Automatic digitizing can work well with a large number of elements on a map, so long as those elements are easily distinguishable by the computer.

Negative Aspects

* Software can be very expensive; however, if the project is sufficiently large, the expensive software may ultimately be cheaper than the labor that would be required for manual digitizing.
* Automatic digitizing can also only translate and trace, but cannot interpret what it is actually looking at.
* Automatic digitizing may be more susceptible to miss digitizing areas in the map where the scanner messed up, or had a smudge.

## Global Navigation Satellite Systems (GNSS)

### What is GNSS?

## A global navigation satellite system (GNSS) is a satellite-based positioning system composed of three components. Global navigation satellite systems are in widespread use today and provide continuous, worldwide positioning.

## 

### Two GNSS Systems

There are two functioning GNSS systems in the world today: NAVSTAR Global Positioning System commonly known as a GPS and is created by the United States of America. The second functioning GNSS is the GLONASS which was created by Russia. There are two GNSSs in development: Galileo by Europe, and BeiDou Navigation Satellite System by China. For the remainder of this lesson we will focus on the NAVSTAR Global Positioning System.

### NAVSTAR GPS

NAVSTAR GPS provides worldwide, continuous, accurate, three-dimensional position and velocity. It also provides precise time. The GPS system is a passive location technology which means that the GPS satellites only transmit information and the GPS receivers held in the user’s hand does not transmit their location to anyone else. It was designed this way so that no one could intercept your location. GPS is provided by the United States of America Department of Defense which originally created GPS for military operations, but has been opened up for public use.

A GPS provides:

* + Worldwide, continuous, accurate, three dimensional position and velocity
  + Precise time
  + Passive location technology

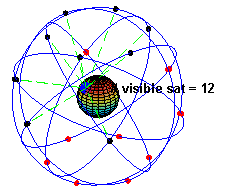
GPS provided by:

* + USA Department of Defense
  + Developed for military operations

### GPS Segments

NAVSTAR GPS is composed of three segments: the space or satellite segment which consists of a constellation of satellites orbiting Earth, the control segment which is terrestrial-based tracking, communications, data gathering, integration, analysis, and control facilities, and the user segment which is composed of individuals using GPS receivers. Attribute each one of these three segments a little more detail.

#### Space Segment

The space segment is composed of a constellation of GPS satellites. There are 24 active satellites orbiting the earth on six orbital planes on a 12 hour orbit. The GPS satellites are at an elevation of 20,200 km. With the 24 active satellites on the six orbital planes and a 12 hour orbit, we are guaranteed to have at least four satellites visible above the horizon and all times anywhere in the world.

The GPS satellites broadcast to carrier waves in the L band which is the radio band. The GPS satellites broadcast with less than 50 W of power at two frequencies. The first frequency, known as the L1 frequency, broadcast that 1575.43 MHz and contains the coarse acquisition code in the P1 code. The second frequency, known as the L2 frequency, broadcast at 1227.60 MHz and contains the P2 code. The P1 and P2 code are used for determining position, while the coarse acquisition code is used for satellite identification and general information.

#### Control Segment

The control segment is composed of Earth-based monitoring stations that track, update the position of and calibrate the clocks on the GPS satellites. There are five monitoring stations spread roughly equally around the world with the master station being in Colorado.

#### User Segment

The third segment of NAVSTAR GPS is the user segment. A user is anyone or anything that uses a GPS receiver. Depending on the investment, there are three user types. The typical user can expect an accuracy of about 15 m using the eponymous GPS method and investing about $100 for the GPS receiver. The autonomous GPS method means that the GPS receiver is receiving positioning information only for the GPS satellites.

The typical Marine user can expect anywhere between ½ to 5 m accuracy using a differential GPS method with an investment from $100 to $5,000. The differential GPS method means of the GPS receiver is receiving positioning information from the GPS satellites and a secondary source. The surveyor user can expect accuracy between .05 mm and 20 mm using differential phase GPS method and investing more than $5000 typically reaching $30,000-$60,000 in investment. The differential phase GPS method means the GPS receiver is receiving positioning information from the GPS satellites and a secondary source and performs complex mathematical calculations to improve its accuracy.

|  |  |  |  |
| --- | --- | --- | --- |
| **User** | **Accuracy** | **GPS Method** | **Price** |
| Typical | ~15 meters | Autonomous | $100 |
| Typical\* /Marine | ~0.5 – 5 meters | Differential | $100-$5,000 |
| Surveyor | ~.05 – 20 mm | Differential Phase | >$5,000 |

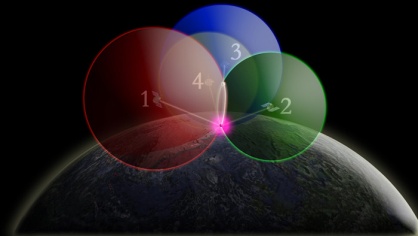
### How does a GPS Work?

Now you know about the three segments of GPS, it is time to find out how GPS works. Two codes are broadcasting the GPS satellites: the coarse acquisition code and the P code. The coarse acquisition code is 1 ms long and is used by both the civilian and military sector.

The coarse acquisition code provides a pseudorandom code describing the ephemeris data and GPS almanac. Ephemeris means the location of the satellites. The P code takes seven days to broadcast one time and is used to determine the location of the GPS receiver. The P code is used by the civilian sector and encrypted Y code is used by the military sector.

Perhaps, the best way to explain how GPS works is to use the analogy of singing a song. Imagine that both the GPS receiver and the GPS satellite have a copy of the same song. The GPS receiver and GPS satellite agree to start playing that song at exactly midnight of a certain day.

Since it takes a little while for the signal to get from space to earth, the GPS receiver will probably be a little ways into the song before it appears the first word to the song from the GPS satellite. The GPS receiver uses his information to determine how “off” the two sequences are from each other which tells us how long it took the signal to travel from space to earth. We can then take that travel time multiplied by the speed of light to get the distance of the GPS receiver is from the GPS satellite.

Continuing the song analogy, if we are singing the song with at least four GPS satellites simultaneously, then we can determine our position relative to all four satellites which provides a unique position on the face of the earth. The idea is that with one satellite we know where we are on the surface of a sphere at a set distance from the satellite. With two satellites, we know we are somewhere in the area were those two spheres intersect, with three satellites we know our position can be either one or two spots, and with the fourth satellite, we know our position to be at a single location.

### Sources of Error

GPS has many potential sources of error and it is important that you understand the sources of error before you go out and perform a data collection. There are five sources of error in GPS: atmospheric, multipath, clock, ephemeris, and constellation geometry. Consult the table to see how much the error could be for each source. We will cover each one of these sources of air and a little more detail.

|  |  |
| --- | --- |
| **Source** | **Error** |
| Atmospheric | 5 meters |
| Multipath | 1 meter |
| Clock | 2 meters |
| Ephemeris | 2.5 meters |
| Constellation Geometry | 1.6-10 meters |

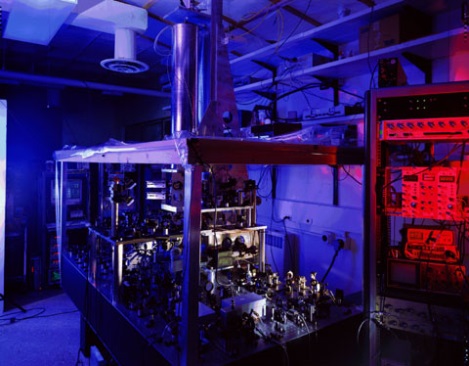
#### Atmospheric Error

Atmospheric error refers to the fact that the speed of light is not constant once it leaves a vacuum. As the radio signal enters the Earth’s atmosphere, its speed is varied because it slows down or is deflected as it works through the atmosphere. Since we rely so heavily on the speed of light in our calculations, any deviation from the speed of light directly affects our distance calculation.

#### Clock Error

Clock error refers to the fact that the GPS satellite and the GPS receiver are not perfectly synchronized in time. The GPS satellites contain atomic clocks while the receivers contain high precision but much deeper clocks. This location is determined in large part by temporal calculations, even small errors and timing can translate to large errors in positioning.

To illustrate the difference between the GPS satellite clocks the clock in the receiver, one picture shows the NIST-F1 atomic clock at the national Institute of standards and technology. This is an extremely precise, accurate, and expensive clock. The receiver clock uses much cheaper components and is prone to error and drift. The GPS receiver does try to synchronize its clock with the GPS satellite clock; however there is often error in the synchronization.

*NIST-F1* Atomic Clock

at the National Institute of

Standards and Technology

Receiver Clock

(Your clock)

#### Ephemeris Error

## Ephemeris error is residual of the satellites exact position in the position to which it reports to be. In the satellite is not the location which believes it to be, it will affect position calculations.

#### Multipath Error

Multipath errors when the signal travels a further distance than the shortest path between the GPS satellite in the GPS receiver. Any additional signal travel time produces position offset as it increases the time it takes for the signal to travel between the GPS satellite in the GPS receiver.

This is most commonly a problem in urban areas and heavily force it areas as the GPS signal will bounce off of houses, trees, leaves, or other large objects before it reaches the GPS receiver, thereby increasing the time it takes for the signal to get from the satellite to the receiver.

#### Constellation Geometry Error

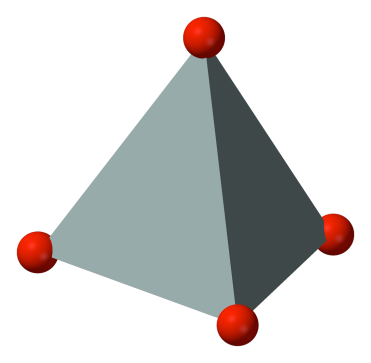
The constellation geometry error refers to ranges created when an area of uncertainty is increased due to the poor distribution or satellites in the sky above. If the satellites are grouped closely together, the area of uncertainty increases as the GPS satellites are further from being perpendicular from each other. Widely spaced satellites are complementary because they result in a smaller area of uncertainty.

### Dilution of Precision (DOP)

GPS satellite constellation geometry can be summarized by using a number called the dilution of precision. There are three kinds of dilution of positions: the vertical dilution of precision (VDOP), the horizontal dilution of precision (HDOP), and the positional dilution of precision (PDOP). The vertical and horizontal dilution of precisions refers to how well the GPS constellation geometry is positioned with respect to providing an accurate vertical or horizontal position. The positional dilution of precision refers to how well the GPS constellation geometry is positioned in respect to providing an accurate position in whole.

### Positional DOP

The positional dilution of precision is the most use of the three. The positional dilution of precision is the ratio of the volume of a tetrahedron created by the most for widespread satellites to the volume of an ideal tetrahedron.

A tetrahedron is created by having points spaced at 120° intervals around the horizon and having the fourth point directly above. As this is a ratio, the lower the positional dilution of precision number is the better, with 1.0 being perfect. In general, if you’re performing high precision measurements, you should keep the positional dilution of precision below 3.0.

### Correcting for Errors

The lesson has explained these errors so now the lesson will focus on how you can minimize errors in the GPS system. For the atmospheric error, you can use differential GPS or do atmospheric modeling to determine the best times to use the GPS receiver. For multipath, you should stay away from buildings or from underneath heavy canopy.

If you need to record a position next to a building, for instance, take a GPS coordinate away from the building, and then use a bearing and distance measurement to calculate the coordinate. There is not much you can do for clock or ephemeris error is that should be handled by the control segment of the GPS. For constellation geometry, you should use a low speed up possible. It is possible to look at a GPS almanac which will tell you what the positional DOP will be throughout the day.

|  |  |
| --- | --- |
| **Source** | **Correction** |
| Atmospheric | Differential GPS  Atmospheric modeling |
| Multipath | Stay away from buildings  Distance offset |
| Clock | Control segment |
| Ephemeris | Control Segment |
| Constellation Geometry | Use lower PDOP |

### Differential GPS

The most common GPS method is autonomous. Autonomous GPS means that the GPS satellite is broadcasting directly to the GPS receiver and no other communications are involved. To increase the accuracy of your position, you can leverage differential GPS. Differential GPS is used primarily to remove most of the range errors inherent to the GPS.

In order to use differential GPS, you must establish a base station receiver at a known coordinate point. The base station and roving receiver should be generally close to each other and should be collecting data from the same satellites. As the base station knows where it is supposed to be because it was set atop a known coordinate point, it can broadcast to the rover receiver the difference between its known location and the location the GPS is identifying. The rover receiver can take this information and apply it to correct its location.

In a differential system, the base station is placed above a known point and does not move. The rover is transportable as what is used to collect points at features of interest.

### Differential Processing Methods

The base station can broadcast the corrections in real time at both the base station and the rovers have some sort of communication capability such as a radio: this is known as real-time kinematic GPS. If neither the base station nor the rover have communication capabilities, and the GPS base station corrections can be applied to the rover data at a later time using a computer which is known as post-processing.

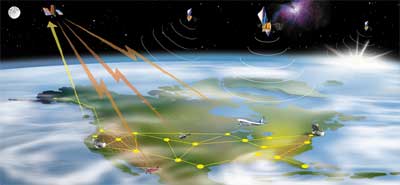
All post-processing does not allow for high accuracy in the field, it is suitable for most situations and does not require the purchase of communication information. Additionally, some vendors and government agencies have base stations set up around the nation that you can use instead of having to establish your own base station.

### Continuous Operation Reference Stations (CORS)

One such set of base stations set up around the nation is the Continuous Operation Reference Stations, commonly referred to as CORS. CORS was established by the United States Coast Guard and the stations are used to provide either real time or post-processing GPS observations depending on the capability of each CORS station.

### Wide Area Augmentation System (WAAS)

A second differential positioning method is the Wide Area Augmentation System, commonly referred to as WAAS. WAAS was established and is operated by the United States Federal Aviation Administration and is designed to provide real-time accuracies within 7 meters or less. Unlike the previous base stations we were discussing, WAAS uses geostationary satellites and a network of reference stations throughout the United States. The geostationary satellites broadcast the corrections to GPS receivers that are able to receive the WAAS signal.



### Uses of GNSS

There are many uses of GNSS. As a GNSS provides accurate positions worldwide, that position can be combined mobile mapping software to allow for field digitizing. The GPS receiver can also be attached to animals, people, or assets, among other things, to provide tracking capabilities that can be stored in a geospatial data set.

## Geocoding

### What is Geocoding?

Geocoding is the process of spatially referencing features based on its address to which it relates on the face of the earth. For example, the GPS unit in your car performs geocoding every time you enter a street address. The GPS unit takes the street address, and associates it with road data sets, to determine the coordinate where that addresses exists.

There are two typical types of geocoding: linear geocoding, and area geocoding. Linear geocoding assumes that addresses vary linearly along a feature which is typically a line. Area geocoding assigns the geocoded location to the entire area that matches the input address.

### Linear Geocoding

An example of linear geocoding would be to imagine a geospatial data set that has the centerlines of roads. Between intersections, each road segment has a beginning address, and an ending address. If we geocode using that data set, since we do not know exactly where each address is on the block, we assume that the addresses are spaced evenly between the start and end range of addresses in the block, and therefore estimate where the addresses exists.

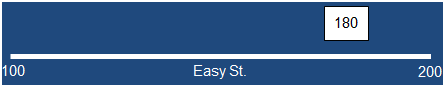
So for the example on the slide, this block of Easy Street starts at the address of 100, and ends at the address of 200. We wish to know where the address 180 Easy Street is located, we first identify the feature named easy Street, then narrow down the results to the block where are input address exists.

Next, we estimate how far down Easy Street that addresses assuming equal spacing of addresses. The major downside to linear geocoding, is at the location may be misplaced if the addresses are linearly spaced. The major upside the linear geocoding, is at the data sets are commonly, and often freely, available, and are easy to construct.

Estimated location

(180 – 100) / (200 – 100) = .8 = 80%

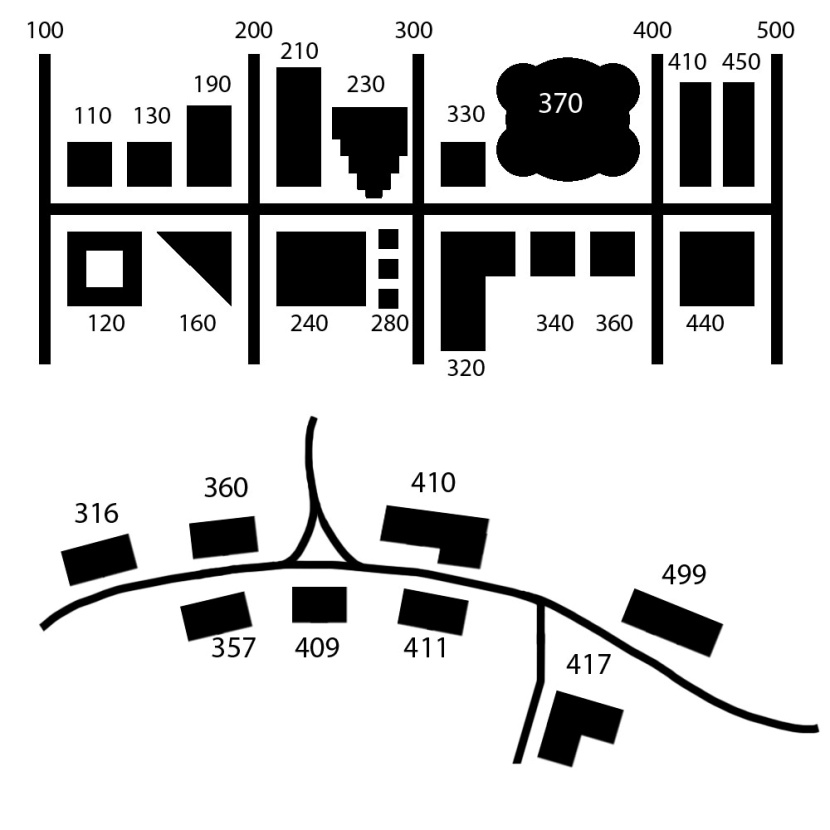
80% down the length of the street



### Area Geocoding

Now let’s look at an example of area geocoding. This illustration shows for street segments, from address 100 to 500, for each block has the address range of 100. Here, the building footprints have been stored in a polygon geospatial data set, and each polygon has the address of that building. As each building has a slightly different shape and size, you can easily see that the addresses are not evenly distributed along the blocks.

With area geocoding, we do not need to worry about whether the addresses are evenly spaced, as we can point to the exact area that the address exists. This is a much more accurate way to geocode then linear geocoding. The major downside to area geocoding is at the data sets are not commonly available, and are very expensive to construct.



## SUMMARY

This lesson provided information about data sources. You learned about creating geospatial data and the various types of data available. This lesson also explained both automatic and manual digitizing with information, images and a comparison of the two methods. The lesson wrapped up with an explanation of Global Navigation Satellite Systems (GNSS).

## ASSIGNMENTS

1. Lab 7: Data Sources
2. Quiz 7: Data Sources