Environmental Data Management Best Practices

Geospatial Data Collection
Collection Consistency

This document provides examples of the common manual and sensor-based technologies that an organization could use in their geographic data strategy, and describes best practices for consistent collection of geospatial environmental data.



Overview

Geospatial collection consistency refers to the standards applied when coordinate locations or other geographic topologies are entered into the environmental data management system. Often an organization's data are used to communicate what it does and create a call to action for the organization's partners and stakeholders. While data are predominantly collected and used internally within the organization, the data are often communicated to a wider audience than the organization itself. It is important to have consistent collection methods to ensure transparency and trust in your organization's data quality. The way the data are managed within a geographic information system (GIS), and how the data are displayed and visualized, are addressed in the Geospatial Visualization of Environmental Data subtopic sheet. Geographic data collection consistency ensures that the data can be used for analysis and modeling of scientific data, allowing the user to compare data across different geographic areas and across different boundaries such as state or county lines. The organization must research and be well informed regarding the various data collection unit capabilities and options available to meet their project's needs before data collection begins. Time taken at the beginning of the project to determine the balance between desired project output and the cost of collecting and producing that output will give organization personnel the information they need to make informed decisions about their organization's collection consistency needs.

Using technology to assist in environmental data collection is one way to ensure consistency for data collection quality. Global positioning system (GPS)–enabled technology is now embedded into many devices, including smart phones, tablets, drones, data loggers, and cameras. These GPS-enabled devices or traditional hand-held GPS receivers can capture coordinates, with the accuracy of this measurement dependent on the number of satellites that are accessible from the location of the GPS receiver (see Figure 1). Technology can also include using unmanned aerial systems, remote data loggers, or other remotely controlled sampling systems.

This document will give examples of the common manual and sensor-based technologies that an organization could use in their geographic data strategy, as well as common data processing techniques and processing errors to watch out for.

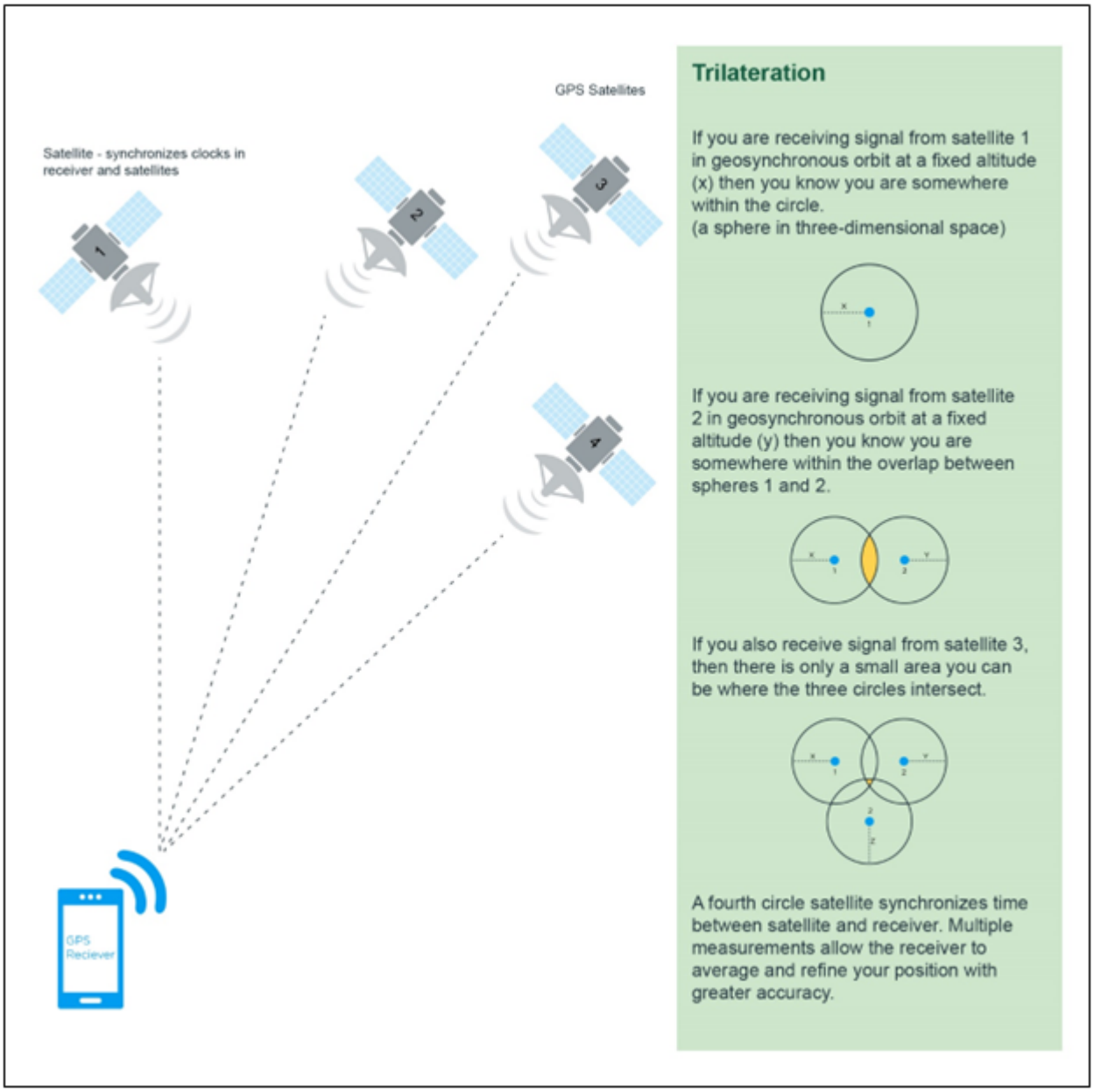


Figure 1. Illustration of how GPS works.

Data Acquisition

If your organization collects location data, you need to have a standard operating procedure for geospatial data collection. It's best practice to be familiar with the operating manual of the equipment and to have standard procedures or a checklist in place that operators must complete for every sample location collected. Ensuring collection consistency can save an organization time and money by not having to resample data because protocols were not followed.

Using technology for field data collection is a growing industry trend. The Geospatial Data Field Hardware subtopic sheet reviews basics for field hardware to ensure collection consistency for the geospatial component of environmental data. The Introduction to and Overview of Field Data Collection Best Practices Fact Sheet describes additional considerations when collecting data in the field. in a survey completed by the *environmental Business Journal* (EBJ 2020), respondents indicated that data collection was the most significant innovation incorporated in environmental practice in the 2010s. A list of available data collection technologies is included below.

- Mobile field data collection efforts use tablet or smartphone devices as the data-logging hardware for capturing GPS data in the field. The strengths of this type of technology include:
    - It replaces hard-copy field notes.
    - It protects data from physical loss.

        - It can save hours and budget in the field and office.
        - It enforces naming conventions for attribution, thereby reducing typing errors.
        - The use of GPS-enabled tablets, drones, and data collectors allows data to be collected real-time in the field. Data can then be mapped sooner, providing a real-time or near real-time reporting of the data.
        - Application changes can be performed remotely.
        - Many users already have a device.
        - Tablet and mobile application software is more cost-effective for budget constraints.
        - The integration of data collection and GIS can provide clients with real-time information (i.e., dashboards) regarding status of their projects. In addition, the utilization of GPS and field data equipment can record geo-located information.
        - Field data are captured and reported electronically (EBJ 2020).

- Digital data collection by phone or tablet allows organizations to:
    - Build data collection apps for these devices, using maps or forms for data entry.
    - Use tools in the organization's software of choice that are available for the mobile device by software as a service (SAAS).

- Data capture methods can be broadly categorized as vector-based or raster-based technologies. These technologies are visually represented in Figure 2 and described as follows:
    - Vector-based data include manual measurements that can be collected by GPS and other land survey methods, or by hand digitization of maps.
    - Raster-based data include imagery (light detection and ranging [Lidar] or satellite), remote sensing data sets (such as thermal, infrared, or hyperspectral), and aerial photography categories.

- Manual data collection methods established by monument, measuring tape and compass, or use of paper maps:
    - A monument should be a stationary reference point (for example, a building corner, nailed board, flagged permanent stake).
    - Line of sight should be established through the compass to the sampling point; degree, direction, and declination, if needed, should be noted.
    - Distance should be calculated from monument.
    - On-screen digitizing of the information from current or historic imagery or from the legal descriptions of the data or location on U.S. Geological Survey 1:24,000-scale topographic map.

- Digital data collection by GPS device
    - Used to navigate and locate field positions and features such as wellheads, sampling points, roads, streams, boundaries, etc.
    - Three levels of GPS units can be used:
        - Recreation-grade GPS—Useful for navigating to locations and providing general location for project items for verification purposes. A recreational-grade GPS device generally has horizontal accuracy of greater than 3 meters. The devices are not designed for high accuracy mapping and production GIS. Depending on your organization standards, the standards for location data quality may allow for the use of recreation-grade GPS.
        - Mapping- or differential-grade GPS—Provides adequate error to maintain the more accurate boundary measures for creating new items or verifying items more accurately. Differential-grade GPS devices generally have a horizontal accuracy of one meter or less.
        - Survey-grade GPS—Used by professional survey crew and has centimeter accuracy. Types of survey-grade GPS include real-time kinematic (RTK), static, and rapid static.

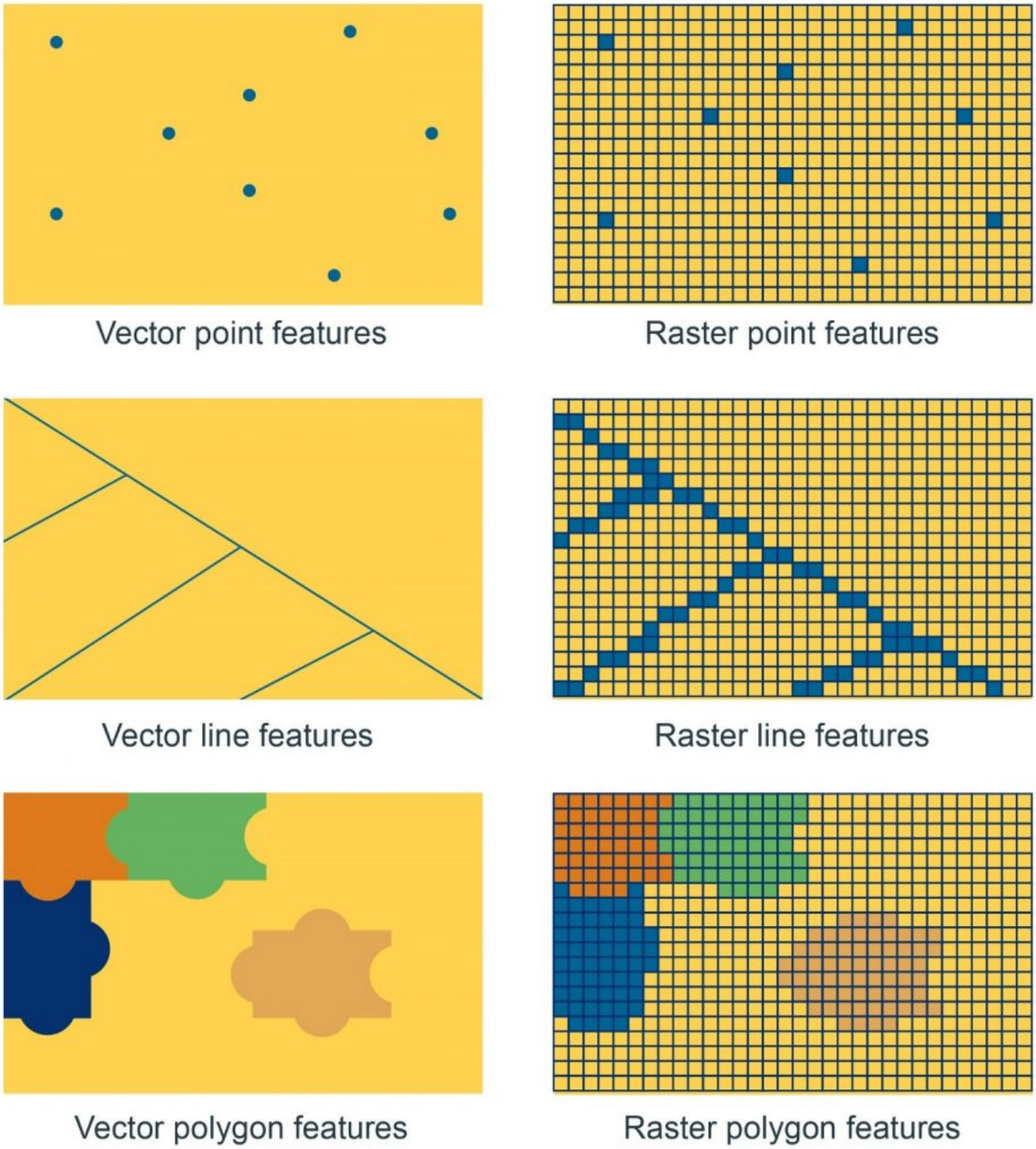


Figure 2. Examples of vector versus raster data.

- Regardless of the GPS device, the vertical error is three times that of the horizontal error. This is important if the elevation of the site is being collected and is important to your project.
- Whenever possible, the GPS user should be in a clear open space with no overhead obstructions; this will give the GPS user the best opportunity to collect quality data.
- The satellites picked up by the GPS should be widespread, not close together, to achieve good geometric dilution of precision (GDOP).
- It is best practice to maintain a position dilution of precision (PDOP) of 3.0 or less on field collection devices.
- The position of the satellites, the atmospheric conditions, and physical features can have an effect on the GPS signal (GISGeography 2021). Minimizing the types of error associated with your collection helps with data integrity, data accuracy, and trust in the data for analysis and reporting purposes. Figure 3 depicts the relationship between user range error and accuracy for GPS technology. If data

collection conditions are not ideal, the data collector should come back another day or at another time to ensure data integrity. Sometimes collecting GPS data at a location may be difficult or impossible due to physical features of the site. Examples of the interference of physical features on GPS accuracy are shown on Figure 4. In such cases, consider using imagery or web-based means to capture coordinate information.

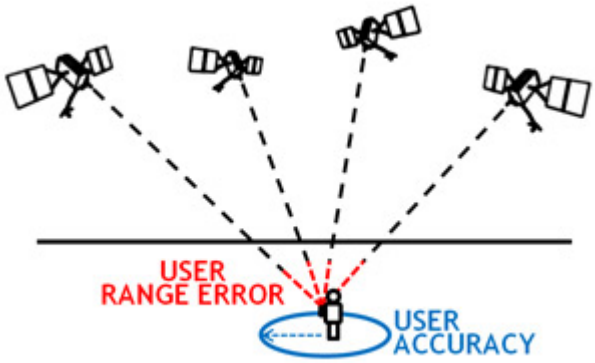


Figure 3. User range error versus user accuracy.
Source: U.S. Space Force, 2022.

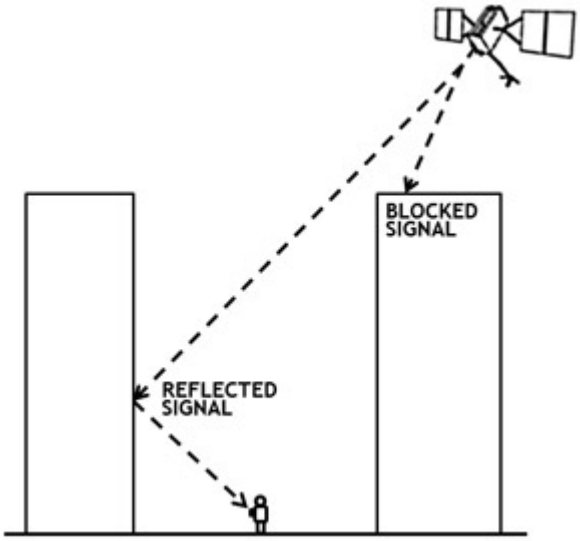


Figure 4. How overhead objects affect GPS position accuracy.
Source: U.S. Space Force, 2022.

- Digital data collection by lidar
    - Lidar is used for the high level of accuracy and precision of locations both horizontally and vertically. This is important when accuracy and precision are critical aspects of a project. Potential projects for which these technologies can be used include:
        - Development or verification of property boundaries
        - Determining locations of existing buildings or physical features
        - Locating existing utilities
        - Identifying and creating easements
        - Creating a digital elevation model (DEM) and contour lines
        - Collecting hydrographic or bathymetric information
        - Recording the final locations of features as they are built during construction, referred to as an as-built survey
        - Wetland delineation

    - One drawback to these more accurate technologies is the cost. Keep in mind that as locational

accuracy increases, so does the cost. The organization needs to balance the cost with the amount of acceptable accuracy. The organization may need to obtain current data at a lower level of accuracy while planning for the potential that future data are collected at a higher level of accuracy.

- Digital data collection by UAV/UAS/drone
    - Many organizations use unmanned aerial vehicles (UAV) or unmanned aerial systems (UAS), commonly referred to as drones. Their popularity is due to ease of use and their ability to capture features in locations difficult for people to access. As an example, an organization can upload the XY coordinates along a flight path to the UAV, and the UAV will navigate to the sights based on the order of the coordinates. Some considerations to keep in mind when contemplating purchasing and using a UAV would be:
        - UAVs are regulated by Federal Aviation Administration (FAA) and local ordinances.
        - There is specific testing and certification that one must receive in order to fly a UAV; see the FAA's Become a Drone Pilot website. It is important to understand the regulations for operating a UAV.
        - UAVs can collect an enormous amount of data in a single flight. A data management plan must be in place for storage, maintenance, and use of the data collected from a UAV. The discussion of a UAV data management plan is beyond the scope of this team. Existing ITRC guidance for Advanced Site Characterization Tools describes and provides recommendations for drone technology (see Section 6.2 of Implementing Advanced Site Characterization Tools).

    - Potential uses of UAVs include:
        - The operator can take site-specific orthoimages if the UAV is equipped with a camera.
        - The user can perform volumetric analysis based on the site data collected at each location, such as cut/fill and stockpile volume.
        - UAVs can be used to collect data for ground surface models.

    - The data quality is such that a person can produce localized terrain analysis and land cover analysis.
        - For example, the UAV can be used to document remediation/construction progress to ensure permit compliance.
        - Inspectors can use UAVs to perform inspections in remote areas that may be difficult to navigate (for example, stacks, structures).
        - Specially equipped UAVs can be used for a laser scan for a more accurate as-built survey of an oil and gas field.

- Data collection by professional land surveyor
    - Ensures that the most accurate collection of data locations (points, lines, and polygons) is used.
    - Good metadata about the location accuracy is followed for the use of the data within the organization's GIS.
    - Ensures that the data has projection file or world file with the complete projection information that the surveyor used to collect the data.
    - The licensed surveyor is signing off that the locations are accurate to a certain degree.

It is important to know the accuracy of your data in order to stay within that level of accuracy for your use of the data. If the data are collected at a 1:100,000 scale, it is inappropriate to use it to model at a 1:12,000 scale and report the accuracy of the map to be 1:12,000.

Data Storage

Cloud computing and related technologies have made large-quantity geospatial data storage more affordable; which has been incredibly beneficial for geographic data projects. Much of the data a GIS professional works with is in the gigabits or terabits size; this can grow to petabits if working with geographically expansive (covering a large geographic area) data sets or crowd-sourced data. Here are some considerations for data storage needs and what an organization should watch for specifically regarding geospatial data sets:

- Do we need cloud-based or on-site storage?
    - Many cloud-based storage companies offer a wide range of storage scenarios to meet the needs of small to large organizations. Have a clear understanding of what those offerings are by company.

- What is the immediate and long-term cost of the data storage needed?
- Plan storage based on geospatial data needs.
    - The more detailed information that is collected, the more space that will be needed for storage. The common slogan in the GIS world is “vector is better, but raster is faster.” Data needs will determine whether to collect and use vector or raster data, and this decision will affect your storage needs.

- Know the internal security offerings that are provided to the organization for both storing and accessing data.
    - Does the organization offer firewall protection, secured services, encryption, security certificates, or password protection to store and to access data?
    - With the possibility of billions of hits on your organization’s web services, attacks can come from anywhere. Some of the most vulnerable areas for these cyberattacks are data storage locations. These storage locations provide excellent places to hide malware or gigabytes of non-organization digital data that can affect the system’s performance both internally and externally.

- Request demonstrations of the offered storage scenarios to see the technology at work.
- It is best to discuss data storage needs during the initial project planning phase before data are collected so you can ensure your agency has a plan in place detailing the storage capabilities needed for all data collected, along with the associated security requirements for storing and accessing the data. Planning goes a long way to achieving your ideal data storage needs; it is nearly impossible to overplan for data management. A lot of planning in the beginning of a project can help avoid problems that can arise because of the lack of storage or sufficient security. Problems that can be avoided by planning ahead include:
    - Lack of storage space for caching and tiles created to optimize raster rendering
    - Lack of available space for running geoprocesses for web services and internal processes
    - Lack of space for multiple versions of program data that have multiple data edits
    - Slow rendering speeds due to low available space on the server
    - Insufficient or suboptimal space for all services running on the server
    - Inability to expand servers to meet unforeseen space requirements as more services and data are added to the server.

Data Processing and Maintenance

After you have collected the data, the next steps involve bringing the data into the organization’s internal software environment, making it available to the organization’s GIS for use by both internal staff and, if authorized, in services used by the public. There are many ways to go about this process. The purpose of this section is to provide a list of some common procedures and software used to make collected data available to internal staff and the public. These suggestions are not the only options, but can provide a starting point for your organization. You may choose to use different procedures or software.

- Post-processing is the act of refining field collection of global navigation satellite system (GNSS) raw data

selected and filtered via a post-processing software that allows the conversion of the raw data into a usable format for your organization's GIS (for example, Trimble Pathfinder Office).
    - Many GPS receivers collect data that need no further processing. The data can be imported and used directly in the GIS. However, if data need to be post-processed, there are some considerations.
    - The University NAVSTAR Consortium (UNAVCO) provides a list of possible GPS/GNSS post-processing software at https://www.unavco.org/software/data-processing/postprocessing/postprocessing.html.

- For data collected with a GPS unit, collected points will be designated and saved in a manufacturer-defined file format, and should be archived in the projects folder in case of loss of the data or data collection device.

After the points are post-processed to a format usable in GIS (that is, converted to a file type that is compatible with your organization's GIS platform), the data may still need to be reviewed and further geographically processed for analysis and consumption in your GIS applications. The type of data captured and intended final use of the data will determine the geographic adjustments that you need to complete. The more common types of geographic adjustments include:

- Spatial adjustment:
    - Spatial adjustment is the process of using spatial tools within the GIS software to stretch or contract your spatial data to fit closer to the data's true location. Spatial adjustment should be used only if there is an inherent error with the data, or no projection or coordinate information (see the "Importance of coordinate systems and datums in geospatial data collection" discussion below in the Quality section of this document) is associated with the data. These situations are likely to occur in historical data, or digitized features from images or paper maps.

- Spatial transformation of digital vector data:
    - This is the process of stretching or contracting the data based on one of many different transformation algorithms that are available from the National Geodetic Survey or built into your GIS software. The organization may collect the data in a certain format, but then transform the data to another format for consumption in their GIS.

- Edge matching:
    - The use of snapping techniques to properly align adjacent mapping data sets along edges of the data sets. This is useful if your project's data collection goes over multiple days. The operator may have to use this tool to properly align the stopping and starting location between days or sampling periods.

- Attribute transfer:
    - The transferring of nonspatial attributes to a feature in a layer can be performed by joining data tables based on a unique identifier and then exporting the feature layer as a new layer. The joined data are transferred to the new feature layer. Also, the GIS analyst can perform a spatial join based on one layer's spatial location in conjunction to another spatial layer. Both techniques are equally effective; the analyst just must figure out which is best for their particular needs.

Data Sharing and Transfer

Data sharing and usage of the collected geographic data is a vital step of your geographic data collection process. If the data are not usable then the time spent collecting and processing the data was wasted. The final data usage will determine the steps that need to be undertaken to prepare the data for that use. This section discusses the factors to consider for sharing the organization's collected data.

The data and services need adequate metadata so users know how to use the data properly. See the Geospatial Metadata subtopic sheet for a complete discussion of metadata standard protocol.

Review the Organization Standards for Geospatial Environmental Data Management subtopic sheet for specifics and examples to follow for using an organization's data.

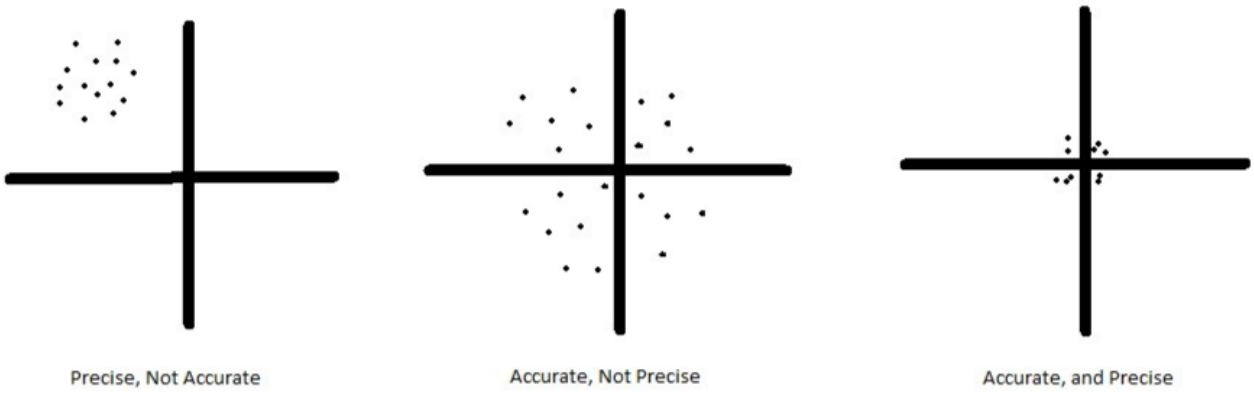
For web services, ensure that the Web Feature Service (WFS) and Web Map Service (WMS) capabilities are set to share the data across the open-source community. Using WFS and WMS allows the data to be used in the public's analysis and modeling applications without giving them the underlying data. Use of web services increases an organization's productivity and improves response times for data requests. The web services can be published to the organization website, increasing the accessibility of the data to the public.

Quality

The level of geospatial data quality feasible for a task or project will be determined by project data, data quality standards, organization standards, national standards, and overall cost. See the Using Data Quality Dimensions to Assess and Manage Data Quality subtopic sheet for additional discussion. Here are a few items to consider when assessing the quality of the geospatial data:

- Accuracy and precision
    - These terms describe how much to trust the collected data. The closer the data are to the true location the lower the collection error for mapping purposes and the lower the accumulated error will be for analytical purposes when combined with multiple data sets. Depending on the type of GPS (recreational versus survey-grade) used, the GPS will collect the error data. If using the basic recreational GPS, the user may have to take multiple readings at a known site near the project area, then average the distance from the known location to get the collection error. Accuracy and precision of location data are correlated with the orientation of satellites in the operator's available satellite array. Figure 5 shows the difference between accuracy and precision of location data.

Figure 5. Difference between accuracy and precision for location data.



- Second, accuracy and precision are affected by the surrounding landscape. High mountains and buildings cause a lot of bounce of the signal received from the satellites. Keep in mind the general area of the sampling efforts for the program in order to purchase an appropriate GPS receiver for the organization's location needs. See the Data Acquisition section above for more discussion on types of location data equipment and procedures.

- Consistency
    - Ensure that proper protocols are followed each time that a feature location is gathered. Proper protocols can include using the organization's standard operating procedures (SOPs) for data collection. If there are no SOPs, follow the operating manual of the data collection device—hold the receiver at the same height, ensure a minimum number of satellites are available each time a feature location is collected, for example. This will assist with post-processing and resampling

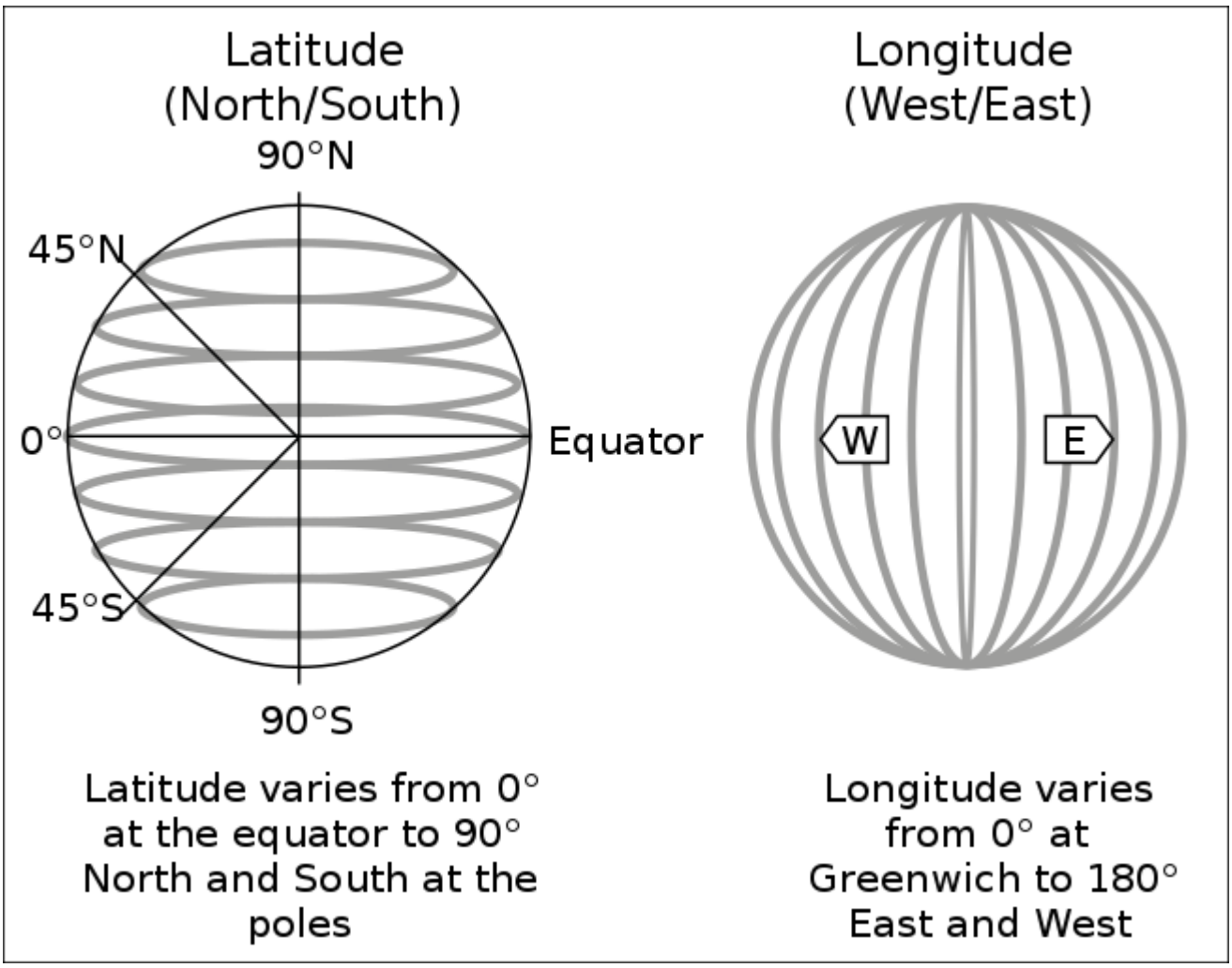
needs.

- Validation
    - Quality control checks are vital to ensuring that the organization has a trustworthy and accurate data set. The more accurate the features are, the more accurate the end products will be.

- Importance of coordinate systems and datums in geospatial data collection:
    - Know the limitations of the projections that your organization uses. Know which are appropriate based on location versus those that are appropriate for displaying across a state or a region. Know which coordinate system is better if you are located on or near projection system boundaries. There are two main categories of projections.
    - The geographic coordinate systems define where the data are located on the Earth's surface (recorded in angular units such as degrees). A projected coordinate system mathematically describes how to draw the data on a flat surface, such as your computer screen or a map. There are hundreds of projected coordinate systems. You will need to use a projected coordinate system to accurately display your geospatial data. Discuss with a GIS professional and refer to software documentation if you need to determine how to complete the data transformation between geographic coordinate systems and projected coordinate systems.
    - Geographic coordinate system (latitude/longitude; see Figure 6), three-dimensional surface:
        - Latitude and longitude coordinates are most commonly collected from a GPS receiver. The data are easily brought into a GIS and converted to the coordinate system of choice for the project. The geographic coordinate system is an algebraic grid system. North America falls in the northwest quadrant so the X-value (longitude) needs a (-) in front of the value in order to fall in the proper quadrant. The Y-value (latitude) will be a positive value. Incorrect negative sign is a common but very easily solvable quality control problem when loading geospatial data into a GIS.
        - One of the most common geographic coordinate systems is World Geodetic System 1984 (WGS84). This coordinate system is the standard used by the U.S. Department of Defense for geospatial information and the default reference system used by GPS receivers.

Figure 6. Illustration of latitude versus longitude.

Source: FedStats, Undated.



- Projected coordinate system (for example, State Plane or Universal Transverse Mercator (UTM)), two-dimensional surface:
    - There are many different projected coordinate systems, so it is important to know if your coordinate system of choice is better for local, statewide, regional, continental, or global representation of the geospatial data. Projected coordinate systems will affect coordinate accuracy and distortion of data. The organization will want to select the coordinate system that has the smallest effect on the feature locations to ensure data accuracy.
    - It is important to note that State Plane coordinate systems are used only in the United States. The State Plane coordinate systems divide each state into multiple zones depending on the size and shape of the state. Distances in these coordinate systems are generally measured in feet. The U.S. Geological Survey provides more information on the State Plane coordinate system and best-use scenarios.
    - The UTM coordinate system can be used globally. It divides the world into sixty north–south zones, and distances in these coordinate systems are measured in meters. The USGS provides more information on the UTM coordinate system and best-use cases.
    - Projected coordinate systems have boundaries where the system changes from one zone within the system to another. For example, the UTM projected coordinate systems has 10 different grids or zones across the United States from Zone 10 on the west coast to Zone 19 in New England. The projected coordinate system you choose may be dependent on where the point is in relationship to the projected coordinate system boundary. Distances between points on or near the edge of coordinate system boundary can be distorted. It is best practice to select projected coordinate systems that preserves the properties most important to the user for creation of maps and geographic analysis.
- Datums—A datum is a point of reference from which the distance around the earth is measured, and as such is a line within a geographic coordinate system. The datum is not a coordinate system. You

must indicate the datum of the coordinate system no matter which coordinate system you use. As an example, the operator could use the WGS84 Geographic Coordinate System with the North American Datum of 1927 (NAD27), North American Datum of 1983 (NA83), WGS84, or another datum. The operator can use whichever coordinate system and datum that they choose; it is important to know which was used when post-processing the locations in your GIS. Not inputting the correct system and datum used for collection will put the data in the wrong location within the GIS.

Scalability and Deployment Considerations

Critical pieces of information to ensure collection consistency for a project include:

- resolution requirement
- horizontal accuracy
- vertical accuracy
- epoch date of data collection

These four pieces of information are important because they can increase the cost of equipment needed to meet the project requirements. As a general rule of thumb, as accuracy increases, so does the cost of the unit. For example, the cost of a recreational-grade GPS with a +10-foot horizontal error starts at about $150. A survey-grade GPS that measures millimeters of movement of the earth's plates will cost upward of $100,000. It is important to know the accuracy requirements for your organization's needs. The accuracy is also important when using the data for analysis and modeling purposes. The lower the coordinate error, the better the trust in the data; this is especially true in projects that are more localized in nature. Positional accuracy is how close the locations are to their true locations on the earth's surface, and it becomes increasingly important the finer the scale of your data. For example, the location accuracy at 1:12,000-scale is more important than at 1:2,000,000-scale (Kerski and Clark 2012).

Another aspect of scale that the organization needs to consider is scale of the project—for example, local or statewide scale. The organization needs to choose the appropriate scale in the project planning phase, and then collect data that are at the appropriate scale for the projects, analyses, or modeling efforts that will be required. This type of scalability will determine if 1:24,000-scale is appropriate or if 1:500,000-scale will suffice. The smallest scale of data will affect the spatial error and the overall scale for your project and analyses. For example, if you are modeling at 1:100,000-scale, but you add a layer that has been collected at 1:1,000,000-scale, the actual scale of your model and analysis will be 1:1,000,000-scale, the smallest scale of the entire project. It would be incorrect to then report your findings at 1:100,000-scale when your true scale is 1:1,000,000.

Vertical accuracy is difficult to work with, because vertical error is roughly three times greater than that of horizontal error. Increasing the vertical accuracy of GPS equipment adds significantly to the cost of that equipment. The organization will need to balance the horizontal and vertical accuracy in the project planning phase. One of the questions to ask is, "does the amount of vertical accuracy gained outweigh the amount of increased cost for the equipment?" As an example, does the cost of a 1-foot horizontal error with a 3-foot vertical error outweigh the cost for a unit with a 3-foot horizontal error and a 9-foot vertical error? The difference between these two measurements could be substantial.

Epoch date can be recorded if a data collection location is professionally surveyed. The epoch date refers to the date the geospatial data is referenced; it can be the date the data was collected or a standardized date (Gakstatter 2013). The epoch date, if it has been collected, should be preserved and maintained as part of the attributes of the location data. The epoch date of the datum is also important because of the movement of the Earth's plates changing over time, particularly along the West Coast of the United States; the eastern boundary of the Pacific Plate is moving to the northwest at about 4 centimeters per year (cm/yr) while the North American Plate is moving southeast at about 2 cm/yr. The coordinates in states closer to the center of the North American continent stay relatively the same over time (Maher 2020). Over long periods of time or when conducting studies over multiple decades, the coordinate locations can be different as plates move, especially on the coasts of North America. The organization needs to be aware of this issue within their coordinate data and note if epoch has been preserved in the original data collection. If collected the epoch indication must be preserved if the same data point is used for sample collection over time. This information will be vital for data collection to take place at the correct location over time. If the epoch date has been accurately preserved, then it is possible to return to the same location for data collection over time after bringing coordinates into the same epoch realization.

The organization must research and be well informed of the various data collection unit capabilities and options available to meet their project's needs before data collection begins. Time taken at the beginning of the project to determine the best balance between desired project output and the cost of collecting and producing that output will give organization personnel the information they need to make informed decisions about their organization's collection consistency needs.

Resources

Related Links:

- https://community.esri.com/t5/arcgis-collector-ideas/epoch-date-support-in-arcgis-pro-explorer-and/idc-p/1092870#M1674
- https://committee.iso.org/sites/tc211/home/projects/projects—complete-list/iso-19111.html
- https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/gcs_vs_pcs/