**Lecture - 10**

**Special Data Accuracy and Quality**

**Spatial accuracy** refers to quantifying errors in the locations of boundaries. Spatial accuracy generally relates to the spatial resolution of the data set. Thematic and Spatial Accuracy are related but are usually addressed separately.

**Data quality** is the degree of data excellency that satisfy the given objective. In other words, completeness of attributes in order to achieve the given task can be termed as Data Quality. Production of data by private sector as well as by various mapping agencies assesses the data quality standards in order to produce better results. Data created from different channels with different techniques can have discrepancies in terms of resolution, orientation and displacements.

[Data quality is a pillar in any GIS](https://www.gislounge.com/keys-to-successfully-managing-gis-data-health-and-quality/) implementation and application as reliable data are indispensable to allow the user obtaining meaningful results. Spatial Data quality can be categorized into Data completeness, Data Precision, Data accuracy and Data Consistency.

* **Data Completeness:** It is basically the measure of totality of features.  A data set with minimal amount of missing features can be termed as Complete-Data.
* **Data Precision:** Precision can be termed as the degree of details that are displayed on a uniform space. More about precision:
* **Data Accuracy:** This can be termed as the discrepancy between the actual attributes value and coded attribute value.
* **Data Consistency:** Data consistency can be termed as the absence of conflicts in a particular database.

**LOCATION ERRORS**

Location errors refer to the geometric inaccuracies of digitized features, which can vary by the data source used for digitizing.

**Location Errors Using Secondary Data Sources**

If the data source for digitizing is a secondary data source such as a paper map, the evaluation of location errors typically begins by comparing the digitized map with the source map. The obvious goal in digitizing is to duplicate the source map in digital format. To determine how well the goal has been achieved, we can plot the digitized map on a transparent sheet and at the same scale as the source map, superimpose the plot on the source map, and see how well they match and if there are any missing lines. How well should the digitized map match the source map? There are no federal standards on the threshold value. A geospatial data producer can decide on the tolerance of location error. For example, an agency can stipulate that each digitized line shall be within 0.01-inch (0.254-millimeter) line width of the source map. At the scale of 1:24,000, this tolerance represents 20 feet (6 to 7 meters) on the ground.

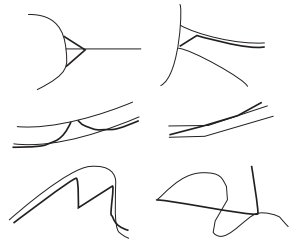
Spatial features digitized from a source map can only be as accurate as the source map itself. A variety of factors can affect the accuracy of the source map. Perhaps the most important factor is the map scale. The accuracy of a map feature is less reliable on a 1:100,000 scale map than on a 1:24,000 scale map. Map scale also influences the level of detail on a published map. As the map scale becomes smaller, the number of map details decreases and the degree of line generalization increases (Monmonier 1996). As a result, a meandering stream on a large-scale map becomes less sinuous on a small-scale map.

**Causes of Digitizing Errors**

Discrepancies between digitized lines and lines on the source map may result from three common scenarios.

**The first** is human errors in manual digitizing. Human error is not difficult to understand: when a source map has hundreds of polygons and thousands of lines, one can easily miss some lines, connect the wrong points, or digitize the same lines twice or even more times. Because of the high resolution of a digitizing table, duplicate lines will not be on top of one another but will intersect to form a series of tiny polygons.

**The second** scenario consists of errors in scanning and tracing. A tracing algorithm usually has problems when raster lines meet or intersect, are too close together, are too wide, or are too thin and broken. Digitizing errors from tracing include collapsed lines, misshapen lines, and extra lines (**Figure)**. Duplicate lines can also occur in tracing because semiautomatic tracing follows continuous lines even if some of the lines have already been traced.



**Figure**: Common types of digitizing errors from tracing. The thin lines are lines on the source map, and the thick lines are lines from tracing.

**The third** scenario consists of errors in converting the digitized map into real-world coordinates. To make a plot at the same scale as the source map, we must use a set of control points to convert the newly digitized map into realworld coordinates. With erroneous control points, this conversion can cause discrepancies between digitized lines and source lines. Unlike seemingly random errors from the first two scenarios, discrepancies from geometric transformation often exhibit regular patterns. To correct these types of location errors, we must redigitize control points and rerun geometric transformation.

**Location Errors Using Primary Data Sources**

Although paper maps are still an important source for spatial data entry, use of primary data sources such as global positioning system (GPS) and remote sensing imagery can bypass printed maps and map generalization practices. The resolution of the measuring instrument determines the accuracy of spatial data collected by GPS or satellite images; map scale has no meaning in this case. The spatial resolution of satellite images can range from less than 1 meter to 1 kilometer. The accuracy of GPS point data can range from several millimeters to 10 meters