Overview of Projections and Datums (Frequently Asked Questions)

What is a projection?

A projection is the operation used to convert between a location (latitude/longitude) on the surface of a sphere (globe) to a location (x/y) on a plane (map).

Now, you may ask, what's so difficult about converting from spherical coordinates to Cartesian coordinates? Well, try this experiment - using an exacto knife, try to cut a ping-pong ball into pieces that lie flat. Can't be done. Now, also using a knife, cut the peel of an orange into pieces that lie flat. This can be done, sort of, if you are willing to stomp on the pieces until they stretch some. So... you can't convert between the surface of a sphere and a plane without distortion.

Which projection should I use?

The short answer is - Whatever projection your existing maps use.

The long answer is - Whatever projection your existing maps use. If one is mapping an area that has never been mapped (not likely on this planet), one can choose almost any projection. In such a case it would make sense to choose a projection that minimizes distortion over the mapping area. But it most cases, one is working with both digital data and existing paper maps. The easiest way to determining what projection to use, is to examine the labels on your map. Properly documented maps will state clearly what projection (and datum) were used in their creation. For example, U.S.G.S. topo maps do this in small print on their lower-left-hand corners. Poorly made maps do not contain this information, and an educated guess has to be made instead. If your maps don't have projection information on them, try contacting a local surveyor. Every geopolitical area on Earth has one, or a very few, preferred projections, and a local surveyor will know what theses projections are. In those cases where your mapping area falls on the border of two or more projection areas, choose one, and use a good program to convert all of your data to that single projection.

What's a datum?

The short answer is - A datum is a fixed, three-dimensional surface, an oblate spheroid, that is approximately the size and shape of the Earth. From this surface, Latitude, Longitude and Elevation are computed. (Note - because the Earth has an ellipsoidal cross-section, Latitude and Longitude can't be measured, they must be calculated.)

The long answer is - Basically, a datum is an admission of failure. With satellites, Mankind can now measure the shape of the Earth with fair accuracy. But before this, men measured the Earth with very-long surveys. Everyone knew that the estimates derived from these data were wrong. The proof of the pudding was that no two geodesists ever came up with the same figure. In fact, single geodesists couldn't come up with the same figure twice in a row (e.g. Clarke 1866, Clarke 1880). A datum is when surveyors all get together and agree to be wrong. They take a spheroid model of the Earth and fix it to a base point. For NAD27, the U.S.G.S. decided that Clarke 1866 was a good approximation, and they fixed it at Meade's Ranch, Kansas. Unfortunately, because the datum is wrong, and because it is fixed, as one moves away from this point, errors pile up. To eliminate these errors, the surveyor is eventually forced to switch to a different datum. When this happens, maps no longer tie. Very scary things happen when you switch datums. Even (or maybe especially) when you switch datums without moving. If I stand in the middle of the intersection of Baseline Road and County Line Road near Boulder, Colorado. I am standing at exactly 40 N. Latitude, 105 W. Longitude (really, that's why they named it Baseline Road). This is true, as long as I am using NAD27. But when I change to NAD83, without moving an inch, I'm no longer at 40 N, 105 W. In fact, I'm now at 39deg 59min 59.97sec N, 105deg 00min 01.93sec W. This is four feet south and fifty feet west from 40 N, 105 W. That's right, latitude/longitude alone does not uniquely describe a location on the surface of the Earth. Changing the datum changes the lat/long of a point on the surface of the Earth. This is important, let me say it again: Describing a place by lat/long is not good enough.

The ideal solution to all of the above problems would be a spheroidal model that has both the correct equatorial and polar radii, and is then centered on the actual center of the Earth. One would then have a spheroid, that when used as a datum, would accurately map the entire Earth. All lat/longs on all maps would agree. That spheroid, derived from satellite measurements of the Earth, is GRS80, and the datums that match it are NAD83 and WGS84.

Which datum should I use?

Whatever datum your existing latitude/longitude values use. Lat/long values aren't absolute, they're calculated. The lat/long value of a place on the Earth's surface depends upon the datum used in its calculation. In other words, for a point on the Earth's surface, if one uses a different datum, one calculates a different lat/long value. The easiest way to determining what datum to use, is to examine the labels on your map. Properly documented maps will state clearly what datum (and projection) were used in their creation. For example, U.S.G.S. topo maps do this in small print on their lower-left-hand corners. Poorly made maps do not contain this information, and an educated guess has to be made instead. If your maps don't have datum information on them, try contacting a local surveyor. Every geopolitical area on Earth has one, or a very few, preferred datums, and a local surveyor will know what these are. In those cases where you have mixed datum lat/long data, choose one, and use a good program to convert all of your data to reference a single datum.

Why don't my GPS lat/longs match my existing lat/longs?

Almost certainly because your GPS receiver is set to output lat/long values referenced to a different datum than your existing lat/long values. GPS receivers calculate their locations referenced to the WGS84 datum. Most GPS receivers by default display WGS84 lat/long values. Most existing mapping data reference a local (non-WGS84) datum. Most GPS receivers have built-in datum transform software, and can be configured to output referenced to whatever datum the user requires.

What are NAD27 and NAD83?

The literal answer to this question is that NAD27 and NAD83 are two datums in common use in North America (NAD is short for North American Datum). The more responsive answer is that most people use the terms to refer to two sets of projections used in the United States and Canada. The NAD27 Coordinate System is intimately linked to the spheroid (Clarke 1866) used in its projection computations, and the associated datum (NAD27) used to generate those lat/long values that are to be converted into grid coordinates. That is to say, only lat/long points calculated using the NAD27 datum should be converted to x/y's using the NAD27 Projection System. But, lat/long's that are calculated in the NAD27 system are valid only in reference to the NAD27 datum, and do not tie to lat/long's referenced to any other datum (i.e. everywhere outside of North

America). It is now possible, with satellite data, to create a spheroid/datum that is valid throughout the entire World. That datum is WGS84, which is basically the same as GRS80, which is for all intents and purposes the same as NAD83. It has been proposed (by The National Geodetic Survey I believe) that the NAD27 datum be replaced by the NAD83 datum in North America. This is obviously a good thing (having all lat/long coordinates agree with each other throughout the World).

It is further proposed (by whom I am not sure) that the U.S. State Plane Coordinate system be 'updated' along with the change from using the NAD27 datum to using the NAD83 datum. These changes are dramatic. California drops a zone (from 7 to 6), Montana drops two zones (from 3 to 1), Nebraska, South Carolina and Puerto Rico/Virgin Islands each drop a zone (from 2 to 1). The Arizona system will change X/Y units from Survey Feet to International Feet, all other zones change from Survey Feet to meters (it's always Arizona that has to be different). All False Eastings and False Northings will change. For Transverse Mercator zones, seven will change Latitude of Origin, seven (but not the same seven) will change Scale Factors, and one will change its Central Meridian. For Lambert Conic zones, Montana changes everything, Latitude of Origin, Longitude of Origin and both Standard Parallels in going from 3 zones to 1 zone, as do Nebraska and South Carolina in going from 2 zones to 1 zone. Six other zones change their Latitude of Origin, and three will change their Longitude of Origin.

As a personal note/rant - I believe that this last idea is a bad thing. The U.S. State Plane Coordinate System is only used in the United States. It has been in use for several decades, and, just as nearly all lat/long data in the U.S. are referenced to the NAD27 Datum, nearly all x/y data are referenced to the State Plane System. The State Plane System and NAD27 datum are archaic systems of limited areal applicability, but they are critically important in the U.S. They are recognized by all States as the system for definition of legal boundaries of all sorts. They have been exclusively used in the generation of all U.S. location data in common use. These data have tremendous economic value, so long as there is no question as to their frame of reference. However, the instant that the frame of reference of location data comes into question, those data lose all of their value. I applaud the use of a Universal Datum in the creation of some maps, and encourage all U.S. Mapping Agencies to create maps using a Universal Datum. However, I believe that the person or people who came up with the idea of updating the State Plane Coordinate System to NAD83, should be making minimum wage flipping hamburgers somewhere, as they are obviously too stupid to be trusted in positions of importance. It seems clear to me that if someone wants to make a map of the United States that ties to globally valid spherical coordinates, that someone should use a projection system that is also globally valid (UTM comes to mind, but I'm open to suggestions). Updating an archaic, limited-use, projection system to an ECEF datum benefits no one, and has the potential to harm all people in the United States whose income depends upon accurate location data (which directly or indirectly is probably everybody).

What is the Earth's 'real' shape?

It's not a sphere, it's only sort of a spheroid. The actual shape of the Earth depends upon who you talk to. There are two surfaces that people refer to as the 'shape' of the Earth. The first is the topographic surface of the Earth. This one's easy to understand. It's where the Lithosphere meets either the Atmosphere or the Hydrosphere. The second is the Mean Sea Level (MSL) Geoid. It is necessary to preface 'The Geoid' (in capital letters) with MSL, as there are an infinite number of geoids. A geoid is an equipotential surface with respect to the gravitational acceleration of the Earth. That is to say, an object anywhere on the surface of the geoid will have the same potential energy. Another aspect of an equipotential surface is that the acceleration due to gravity is always perpendicular to the geoid. Or, gravity always points straight down towards the geoid. Or, an equipotential surface is 'flat' with respect to gravity. A picture of the MSL Geoid appears rugose, (rippled) from above, and in fact is rugose due to local variations in the Earth's gravitational field (See the NOAA picture of the MSL Geoid under North America). To a person walking however, a geoid appears 'flat'. More importantly, to a person surveying, the geoid is 'flat'. So, if the MSL Geoid is the 'true shape of the Earth', how well does the WGS84 datum match this surface? It turns out to be an amazingly good fit. For most of the Earth the deviation between the MSL Geoid and the WGS84 Datum is within +/- 40 meters. The two exceptions are a Northern Mid-Atlantic high in the MSL Geoid, centered on the Mid-Atlantic Ridge, where the geoid reaches slightly more than 60 meters above the datum, and a Southern India low, just off the Southern tip of India, where the geoid reaches slightly more than 100 meters below the datum.

What is a spheroid / ellipsoid?

There are exact mathematical definitions of these two terms, which I won't go into too deeply here, because they aren't what map makers mean when they use them. It is enough to say that in mathematics, a spheroid is a type of ellipsoid, one that is made by rotating an ellipse, in the third dimension, around either its long (semimajor) or short (semiminor) axis.

When rotated about its semiminor axis, a spheroid is called an oblate spheroid. The Earth is (roughly) elliptical in cross-section, rotates about its short axis, and is therefor (approximately) an oblate spheroid. Map makers tend to be less anal retentive than mathematicians, and are quite happy using the terms 'ellipsoid' and 'spheroid' interchangeably. When a map maker uses either term, he (or she) is referring to one of a number of oblate-spheroidal models of the Earth. Many people confuse spheroid and datum. It's probably easiest to remember that a spheroid is an Earth model, a datum is the practical application of the model. Much of the confusion arises from the fact that spheroid values go into both calculating lat/long values from a datum, and calculating x/y values with a projection. To make things even more confusing, there is nothing that requires that the datum spheroid be the same as the projection spheroid, and they are often different.

What are False Easting (False X) and False Northing (False Y)?

In order to minimize distortion, most projections are centered in the mapping area. The unfortunate side effect of this is that with the x/y origin in the mapping area, negative x or y values are likely. To eliminate this possibility, False X and False Y values are chosen and added to all x and y values so that all x/y values are positive