BASICS OF GIS & ArcGIS

Esri (/ˈɛzriː/, a.k.a. **Environmental Systems Research Institute**) is an international supplier of geographic information system (GIS) softwares (such as ArcGIS). ArcGIS is a set of tools that allows the visualization and management of geographic information, and also has an extensible architecture through which new functionalities can be added (Olaya, 2011). These are the well-known extensions, among which are Spatial Analyst (raster analysis), 3D Analyst (3D and relief analysis) or Geostatistical Analyst (geostatistics). The company is headquartered in Redlands, California.

Four categories of ESRI's GIS products:

1. **Desktop GIS:** ArcGIS Desktop (hereafter ArcGIS) is the main component of ESRI's ArcGIS application suite and consists of several integrated applications, including ArcMap, ArcCatalog, ArcToolbox, ArcScene, ArcGlobe, and ArcGIS Pro. The product suite is available in three levels of licensing:
    - **Basic (formerly called ArcView):** provides a basic set of GIS capabilities suitable for many GIS applications
    - **Standard (formerly called ArcEditor):** at added cost, allows more extensive data editing and manipulation, including server geodatabase editing
    - **Advanced (formerly called ArcInfo):** at the high end, provides full, advanced analysis and data management capabilities, including geostatistical and topological analysis tools
   - ArcGIS Explorer, ArcReader, and ArcExplorer are basic freeware applications for viewing GIS data
   - ArcGIS for Desktop Extensions are available, including Spatial Analyst for raster analysis, and 3D Analyst for terrain mapping and analysis.
2. **Server GIS:** Server GIS products provide GIS functionality and data deployed from a central environment. ArcGIS for Server is an Internet application service, used to extend the functionality of ArcGIS for Desktop software to a browser based environment.
3. **Mobile GIS:** Mobile GIS conflates GIS, GPS, location-based services, hand-held computing, and the growing availability of geographic data.
4. **Deveoper GIS:** Developer GIS products enable building custom desktop or server GIS applications or embed GIS functionality in existing applications. These focused solutions can then be deployed throughout an organization.
5. **Online GIS:** ArcGIS includes Internet capabilities in all Esri software products.

- Mansi Mungee
- Wildlife Institute of India & IISER, Pune
- mansi[dot]2303[at]gmail[dot]com

Data formats required by ESRI's GIS products:

Despite the heterogeneity of geographic information, there are two basic approaches to simplify and model space: the vector model, usually used to treat discrete geographic phenomena (communications routes, urban fabric, plant cover, etc.), and the raster model, which is generally used to represent continuous phenomena.

1. **Vector model:** The vector data model is based on the assumption that the Earth's surface is composed of discrete objects such as trees, rivers, lagoons, etc. (ESRI, 2010). In this model, there are no fundamental units that divide the collected area, but, rather, the variability and characteristics of this area are collected by means of geometric entities. For each geometric entity the characteristics are constant. The form of these entities is explicitly codified, because it models the geographic space through a series of primitive geometrics containing the most outstanding elements of that space. These primitives are of three types: points, lines and polygons (Olaya, 2014).

    - **Shapefile** – Esri's somewhat open, hybrid vector data format using .shp, .shx and .dbf files.
    - **Enterprise Geodatabase** – Esri's geodatabase format for use in an RDBMS.

The geodatabase is the native data format for ArcGIS. It is a data storage container that defines how data is stored, accessed, and managed by ArcGIS. The term geodatabase combines geo (spatial data) with database (specifically a relational database management system or RDBMS).

    - **File Geodatabase** – Esri's file-based geodatabase format, stored as folders in a file system.
    - **Personal Geodatabase** – Esri's closed, integrated vector data storage strategy using Microsoft's Access MDB format.

2. **Raster Model:** The raster structure is based on a matrix of cells represented in rows and columns. Each cell can store information about a given variable (precipitation, temperature, relative humidity, solar radiation, radiance, reflectivity, etc.). The raster model does not explicitly collect the coordinates of each cell, but, rather, the values of each cell. It is not necessary to accompany these values with a specific spatial location, since they refer to a particular element of the mesh, which represents a fixed and regular structure. But it is necessary to place this grid in space so that the coordinates of each cell can be calculated (Olaya, 2014)

    - **Esri grid** – binary and metadataless ASCII raster formats.
    - **Mosaic** - data structure for managing and analyzing multidimensional raster and imagery data, including netCDF, GRIB, and Hierarchical Data Format

Mansi Mungee
Wildlife Institute of India & IISER, Pune
mansi[dot]2303[at]gmail[dot]com

***NetCDF** (Network Common Data Form)is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.*

***GRIB** (**GRIdded Binary or General Regularly-distributed Information in Binary form**) is a concise data format commonly used in meteorology to store historical and forecast weather data.*

***Hierarchical Data Format (HDF)** is a set of file formats (HDF4, HDF5) designed to store and organize large amounts of data.*

Glossary of the most relevant geographical terms that will be used in this document:

1. **Band:** Each of the parts into which the electromagnetic spectrum is divided to enable the sensors to capture the electromagnetic radiation. The radiation data (numerical values) captured for each defined band are usually organized as raster files (Moreno Jiménez, 2008).

2. **Layer:** Basic unit of geographic information that can be requested as a map from a server [ISO 19128:2005]. Conceptually, a layer is a portion or stratum of geographic space in a particular area, and could be considered the equivalent of an element of the legend of the map (SENPLADES, 2013).

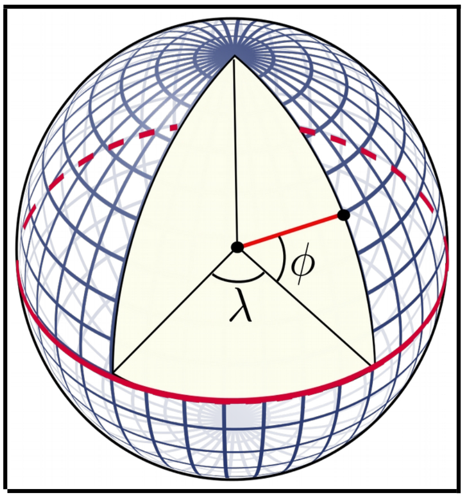
3. **Datum:** Parameter or set of parameters that define the position (A.282) of the origin, scale and orientation of a coordinate system. A **datum** typically defines the surface (eg. radius for a sphere, major axis and minor axis or inverse flattening for an ellipsoid) and the position of the surface relative to the center of the earth. One could have (and often that's the case, especially with the increasing popularity of WGS84/GRS80 ellipsoid) several different datums based on the exact same ellipsoid parameters. The reason for this is that while the WGS84 datum is OK globally since its surface is set to provide minimal average shifts due to tectonic movements across the globe, there's room for improvement on the local scale, where the reference can be fixed to some local reference point or at least to the local tectonic plate (e.g. ETRS, which is fixed to continental Europe)

    - One could explain datum simply as "an agreement on the coordinate system type, shape and its absolute position and orientation relative to some well-known or well-defined real-world reference". The coordinate system doesn't even have to be ellipsoidal (e.g. Vertical datum, which is usually defined by saying that the height of some fixed point is such, and all other heights will be measured relative to this point).

    - To sum it up: A datum is a definition of the size, orientation and position of an ellipsoid used as an approximation of the earths shape. It uses reference points on the surface to define it's placement and orientation, based on a date (which is why a number is in there for the year it was defined to

Mansi Mungee
Wildlife Institute of India & IISER, Pune
mansi[dot]2303[at]gmail[dot]com

account for tectonic plate movements). Datums are used in both spherical long/lat and projected coordinate systems. Consider it a reference point for your coordinates and ellipsoidal heights (ie where's the primemeridian, equator, and what's the height relative to the ellipsoid which isn't the mean sea level). Different datums are used different places because some fit some areas better than others.



- A projection is a formula used to convert long/lat coordinates into a flat coordinate system that you can use on paper or a computer screen. It's usually done from a geographic coordinate system, which in turn uses a datum as it's base definition. So the datum affects all of it. Projecting data creates a lot of distortion of the real world, so it really should only be done when putting your map data on a flat map, or you want to work in a "simpler" coordinate system and can live with the distortions.
- Using the wrong datum could result in your data being offset up to about a mile, so it's quite important to know the datum if you're mixing data together.

4. **Vertical datum:** Datum describing the relationship of height related to gravity or depth with the Earth. NOTE: In most cases, the vertical datum is related to the mean sea level. The geodesic heights are treated in relation to a three-dimensional ellipsoidal coordinate system that refers to a geodetic datum [ISO 19111:2007].

5. **Ellipsoid:** Surface formed by the rotation of an ellipse around a main axis. NOTE: In this international standard, the ellipsoids are always oblong, which means that the axis of rotation is always the minor axis [ISO 19111:2007].

Length of One Degree of Longitude (on WGS 84 Ellipsoid)			Length of a Degree of Latitude (on the WGS 84 Ellipsoid)		
Latitude	Kilometres	Miles	Latitude	Kilometre s	Miles
0°	111.32	69.17	0°	110.57	68.71
10°	109.64	68.13	10°	110.61	68.73
20°	104.65	65.03	20°	110.70	68.79
30°	96.49	59.95	30°	110.85	68.88
40°	85.39	53.06	40°	111.04	68.99
50°	71.70	44.55	50°	111.23	69.12
60°	55.80	34.67	60°	111.41	69.23
70°	38.19	23.73	70°	111.56	69.32
80°	19.39	12.05	80°	111.66	69.38
90°	0.00	0.00	90°	111.69	69.40

At the equator, the distance of between one degree of latitude of WGS 84 is 111.32 km. This is close to 111.17 km. (less than one percent error) Proving that the circular assumption is valid.

6. **Scale:** Relationship between the magnitudes of the elements represented on the map and those that they have in reality. It defines the reduction to which the earth's surface must be submitted to be able to represent it in a document or on a map, whose size is much smaller. The representation of the scale on a map can be graphical or numerical (López Trigal, 2015).

Mansi Mungee
Wildlife Institute of India & IISER, Pune
mansi[dot]2303[at]gmail[dot]com

7. **Geopositioning:** Retrieving the geographic position of an object [ISO/TS 19130:2010].

8. **Georeferencing:** Operation of obtaining and assigning geographical coordinates to information (usually a layer) that lacks it. Usually applied to place Earth images or events associated with postal addresses (Moreno Jiménez, 2008). **Georeferencing** assigns locations (in three dimensions!) to points on a spheroid. Georeferencing is done with a **datum**.

9. **Image:** Raster type coverage whose attribute values are a numeric representation of a physical parameter [ISO 19115-2:2009]

10. **Digital terrain model and digital elevation model:** A digital terrain model is a spatial representation of a quantitative and continuous variable, such as temperature, or atmospheric pressure. In particular, when the variable to be represented is the elevation of the terrain, it is called the Digital Elevation Model or DEM. Therefore, a digital elevation model is a numerical data structure that represents the spatial distribution of the altitude of the terrain surface (Mancebo, Ortega, Martín, & Valentín,2008).

11. **Slope:** Relation of elevation change with respect to the length of the curve [ISO 19133:2005].

12. **Cartographic projection:** Geometric operation that allows the curved surface of the earth (three-dimensional) to be represented on a flat (two-dimensional) surface. This procedure is used to transform the angular coordinates with which the location of the geographic objects is determined on the terrestrial globe into plane coordinates that allow the cartographic representation on a surface of two dimensions (López Trigal, 2015). A **Projection** is a series of transformations which convert the location of points on a curved surface(the reference surface or datum) to locations on flat plane (ie transforms coordinates from one coordinate reference system to another). **Projecting** is an operation that mathematically distorts and shrinks a portion of the spheroid onto flat paper. Projecting can be undone ("inverted"). "Unprojection" expands a feature on a map and plasters it back onto the spheroid.It, too, is a mathematical operation.

13. The **spheroid** models the shape of the earth's surface. It is an idealization that does not account for local changes in topography.

14. **Remote sensing:** In a broad sense it can be defined as the acquisition of information about an object at a distance, that is, without there being material contact between the observed object or system and the observer (Sobrino, 2000).

15. **Latitude:** Usually represented by the greek letter φ (capital letter fi), latitude is the angle formed from the center of the Earth on the meridian plane by the normal to the ellipsoid at the point considered and the plane of the equator. All points on the Earth's surface with equal latitude define the parallel lines. The measure is expressed in sexagesimal notation starting from the equator, positive towards the north (0° to 90°) and negative towards the south (0° to -90°) (Del Bosque González et al. 2012).

16. **Longitude:** Usually represented by the greek letter λ (lambda), longitude is the dihedral angle, formed from the center of the Earth on the plane of the equator, between the meridian containing the point and

the meridian taken as the origin. The plane of the equator is that which passes through the center of the Earth and is perpendicular to the axis of rotation. The meridians are lines formed by all points of equal length, and represent the intersection of the Earth's surface with planes perpendicular to the plane of the equator and containing the axis of rotation (Del Bosque González et al. 2012).

- The Equator represents 0° latitude, while the North and South Poles represent 90° North and 90° South latitudes. In addition to the Equator, there are four other major latitudes that are usually found on maps and globes. The positions of these latitudes are determined by the Earth's axial tilt.
- The ***Arctic Circle*** **is** the latitude 66° 34′ North. All locations falling North of this latitude are said to be in the Arctic Circle.
- The ***Antarctic Circle*** on the other hand, is the latitude 66° 34′ south. Any locations falling south of this latitude are said to be in the Antarctic Circle. Places in both the Arctic and Antarctic circles experience extreme weather, and experience the Midnight Sun.
- The latitude 23° 26′ North is also known as the ***Tropic of Cancer.*** It marks the northern-most position on the Earth, where the Sun is directly overhead at least once a year. This happens during the June Solstice, when the Earth’s Northern Hemisphere is tilted towards the Sun.
- ***The Tropic of Capricorn*** is the latitude that lies at 23° 26′ South of the Equator. It is the southern-most position on the globe, where the sun is directly overhead during the December Solstice.
- Sometimes, latitudes north of the Equator are denoted by a positive sign. Latitudes south of the Equator are given negative values. This eliminates the need to add whether the specified latitude is north or south of the Equator.
- Half of a longitudinal circle is known as a *Meridian*. Meridians are perpendicular to every latitude.
- Unlike, latitudes, there is no obvious central longitude. However, in order to measure the position of a location based on the longitude, cartographers and geographers over the course of history have designated different locations as the main longitudinal reference point. Today, the meridian line through Greenwich, England is considered as the reference point for longitudes. This line is also known as the ***Prime Meridian.***
- The Prime Meridian is set as 0° longitude and it divides the Earth into the Eastern and the Western Hemisphere. All the other longitudes are measured, and named after the angle they make with respect to the center of the Earth from the intersection of the Meridian and the Equator.
- Since a sphere has 360 degrees, the Earth is divided into 360 longitudes. The meridian opposite the Prime Meridian (on the other side of the Earth) is the 180° longitude and is known as the ***anti meridian.***

Mansi Mungee
Wildlife Institute of India & IISER, Pune
mansi[dot]2303[at]gmail[dot]com

- Modern timekeeping systems use longitudes as references to keep time. Time zones are defined by the Prime Meridian and the longitudes.

All spatial data is created in some coordinate system, whether it is points, lines, polygons, rasters, or annotation. The coordinates themselves can be specified in many different ways, such as decimal degrees, feet, meters, or kilometers—in fact, any form of measurement can be used as a coordinate system. Identifying this measurement system is the first step to choosing a coordinate system that displays your data in its correct position in ArcGIS Pro, in relation to your other data.

Coordinate systems

Coordinate systems enable geographic datasets to use common locations for integration. A coordinate system is a reference system used to represent the locations of geographic features, imagery, and observations, such as Global Positioning System (GPS) locations, within a common geographic. Units of measurement is typically feet or meters for projected coordinate systems or decimal degrees for latitude-longitude framework. Data is defined in both horizontal and vertical coordinate systems. Horizontal coordinate systems locate data across the surface of the earth, and vertical coordinate systems locate the relative height or depth of data. Several hundred geographic coordinate systems and a few thousand projected coordinate systems are available for use. In addition, you can define a custom coordinate system.

Types of coordinate systems

There are two categories of coordinate systems –

1. **Horizontal Coordinate system:** 3 types: geographic, projected, and local. A physical location will usually have different coordinate values in different geographic coordinate systems. You can find out what coordinate system your data is in by examining the layer's properties.

   a) **Geographic Coordinate System (GCS):** A global or spherical coordinate system such as latitude-longitude. These lines encompass the globe and form a gridded network called a graticule. GCS uses a three-dimensional spherical surface to define locations on the earth. A point is referenced by its longitude and latitude values. Latitude and longitude values are traditionally measured either in decimal degrees or in degrees, minutes, and seconds (DMS).

   b) **Projected Coordinate System (PCS):** such as **Universal Transverse Mercator (UTM)**, Albers Equal Area, or Robinson, all of which (along with numerous other map projection models) provide various mechanisms to project maps of the earth's spherical surface onto a two-dimensional Cartesian coordinate plane. Projected coordinate systems are referred to as map projections. Coordinates are expressed using linear measurements rather than angular degrees. Finally, some data may be expressed in a local coordinate system with a false origin (0, 0 or other values) in an arbitrary location that can be anywhere on earth. Local coordinate systems are often used for large-scale (small area) mapping. The false origin may be aligned to a known real-world coordinate or not, but for the purposes of data capture, bearings and distances may be measured using the local coordinate system rather than global coordinates. Local coordinate systems are usually expressed in feet or meters. A projected coordinate system (PCS) is defined on a flat, two-dimensional surface. Unlike a GCS, a PCS has constant lengths, angles, and areas across the two dimensions. A PCS is always based on a GCS that is based on a sphere or spheroid. In addition to the GCS, a PCS includes a map projection, a set of projection parameters that customize the map projection for a particular location, and a linear unit of measure.

2. **Vertical coordinate systems :** A vertical coordinate system defines the origin for height or depth values. Like a horizontal coordinate system, most of the information in a vertical coordinate system is not needed unless you want to display or combine a dataset with other data that uses a different vertical coordinate system. Perhaps the most important part of a vertical coordinate system is its unit of measure. The unit of measure is always linear (for example, international feet or meters). Another important part is whether the z-values represent heights (elevations) or depths. For each type, the z-axis direction is positive "up" or "down," respectively.

    - You cannot define a vertical coordinate system on a dataset without a corresponding geographic or projected coordinate system.

    - Vertical coordinate systems are either gravity-based or ellipsoidal. Gravity-based vertical coordinate systems reference a mean sea level calculation. Ellipsoidal coordinate systems reference a mathematically derived spheroidal or ellipsoidal volumetric surface.

Spatial Referencing

A spatial reference is a series of parameters that define the coordinate system and other spatial properties for each dataset in the geodatabase. It is typical that all datasets for the same area (and in the same geodatabase) use a common spatial reference definition. A spatial reference includes the following:

- The coordinate system
- The coordinate precision with which coordinates are stored (often referred to as the coordinate resolution)
- Processing tolerances (such as the cluster tolerance)
- The spatial extent covered by the dataset (often referred to as the spatial domain)

Mansi Mungee
Wildlife Institute of India & IISER, Pune
mansi[dot]2303[at]gmail[dot]com

Transformations

After defining the coordinate system that matches your data, you may still find you want to use data in a different coordinate system. This is where transformations are useful. Transformations are required to convert data between different geographic coordinate systems or between different vertical coordinate systems. Unless your data lines up, you'll face difficulties and inaccuracies in any analysis and mapping you perform on the mismatched data.

1. **Geographic (datum) transformations:** If two datasets are not referenced to the same geographic coordinate system, you may need to perform a geographic (datum) transformation.This is a well-defined mathematical method to convert coordinates between two geographic coordinate systems.

2. **Map projections:** Whether you treat the earth as a sphere or a spheroid, you must transform its three-dimensional surface to create a flat map sheet. This mathematical transformation is commonly referred to as a map projection. Representing the earth's surface in two dimensions causes distortion in the shape, area, distance, or direction of the data. Different projections cause different types of distortions. Some projections are designed to minimize the distortion of one or two of the data's characteristics. A projection could maintain the area of a feature but alter its shape. The extent, location, and property you want to preserve must inform your choice of map projection. There are over 4000 coordinate systems in the ArcGIS platform, so it is likely that you can find one to match your data. In the event that you cannot, you can create a custom coordinate system to display the data.

    - ArcGIS Pro reprojects data on the fly so any data you add to a map adopts the coordinate system definition of the first layer added. As long as the first layer added has its coordinate system correctly defined, all other data that has correct coordinate system information reprojects on the fly to the coordinate system of the map. This approach facilitates exploring and mapping data, but it should not be used for analysis or editing, because it can lead to inaccuracies from misaligned data between different layers. Data is also slower to draw when it is projected on the fly. If you intend to perform analysis or edit the data, then first project it into a consistent coordinate system and the same coordinate system shared by all your layers. This creates a new version of your data.

Projection parameters

A map projection by itself is not enough to define a PCS. You can state that a dataset is in Transverse Mercator, but that's not enough information. Where is the center of the projection? Was a scale factor used? Without knowing the exact values for the projection parameters, the dataset cannot be reprojected.

- You can also get some idea of the amount of distortion the projection has added to the data. If you're interested in Australia but you know that a dataset's projection is centered at 0,0, the intersection of the equator and the Greenwich prime meridian, you might want to think about changing the center of the projection.

- Each map projection has a set of parameters that you must define. The parameters specify the origin and customize a projection for your area of interest. Angular parameters use the GCS units, while linear parameters use the PCS units.

Linear parameters

- False easting is a linear value applied to the origin of the x-coordinates. False northing is a linear value applied to the origin of the y-coordinates.
- False easting and northing values are usually applied to ensure that all x- and y- values are positive. You can also use the false easting and northing parameters to reduce the range of the x- or y- coordinate values. For example, if you know all y- values are greater than 5,000,000 meters, you could apply a false northing of –5,000,000.
- Height defines the point of perspective above the surface of the sphere or spheroid for the Vertical Near-Side Perspective projection.

Angular parameters

- Azimuth defines the centerline of a projection. The rotation angle measures east from north. It is used with the azimuth cases of the Hotine Oblique Mercator projection.
- Central meridian defines the origin of the x-coordinates.
- Longitude of origin defines the origin of the x-coordinates. The central meridian and longitude of origin parameters are synonymous.
- Central parallel defines the origin of the y-coordinates.
- Latitude of origin defines the origin of the y-coordinates. This parameter may not be located at the center of the projection. In particular, conic projections use this parameter to set the origin of the y-coordinates below the area of interest. In that instance, you do not need to set a false northing parameter to ensure that all y- coordinates are positive.
- Longitude of center is used with the Hotine Oblique Mercator center (both two-point and azimuth) cases to define the origin of the x-coordinates. It is usually synonymous with the longitude of origin and central meridian parameters.
- Latitude of center is used with the Hotine Oblique Mercator center (both two-point and azimuth) cases to define the origin of the y-coordinates. It is almost always the center of the projection.
- Standard parallel 1 and standard parallel 2 are used with conic projections to define the latitude lines where the scale is 1.0. When defining a Lambert Conformal Conic projection with one standard parallel, the first standard parallel defines the origin of the y-coordinates.

Mansi Mungee
Wildlife Institute of India & IISER, Pune
mansi[dot]2303[at]gmail[dot]com

Unitless parameters

- Scale factor is a unitless value applied to the center point or centerline of a map projection. The scale factor is usually slightly less than one. The UTM coordinate system, which uses the Transverse Mercator projection, has a scale factor of 0.9996. Rather than 1.0, the scale along the central meridian of the projection is 0.9996. This creates two almost parallel lines approximately 180 kilometers, or about 1°, away where the scale is 1.0. The scale factor reduces the overall distortion of the projection in the area of interest.
- X and y scales are used in the Krovak projection to orient the axes.

Mansi Mungee
Wildlife Institute of India & IISER, Pune
mansi[dot]2303[at]gmail[dot]com