Date / /

Geographical data: Nature, types and Sources; Scale of measurement (Nominal, ordinal, Interval and Ratio.)

The data are defined as numbers that represent measurements from the real world. Datum is a single measurement. In otherwords any figure or name or any ~~quantity~~ quantity of an object is a data. It is transformed into the information when it is meaningful. It becomes geographical in nature when it conveys some spatial or temporal information.) In general, due to the highly variable character of a numerical data it is known as a variate or variable. We often read the news like 20 cm of continuous rain in some place or 35 cm of rain ~~in~~ in a place in 24 hours OR information like New delhi – Mumbai distance via Kota – Vadodara is 1385 km and via Itarsi – Mumma is 1542 km by train. This numerical information is called data. There are large volume of data available around the world today. However at times, it is difficult to derive logical conclusion from these data if they are in raw form. Therefore it is important to ensure that the measured information is algorithmically derived and/or logically deduced and/or statistically calculated from multiple data. (Here information is defined as either a meaningful answer to a query or a meaningful stimulus that can cascade (a process whereby something, typically information or knowledge is successively passed on) into further queries.

In "geographical data", the first term conveys two meanings — "earth" and "geographic Space". By "earth" it is implied that all data is pertinent to earth's features and resources, including human activities, based on or associated with these features and resources. By "geographic Space", it is meant that the commonality of both the data and the problems that the systems are designed to solve is 'geography', i.e. location, distribution, pattern, and relationship within a specific geographical reference framework. The second term refers to a set of facts and figures, relating to places, people, events, things and concepts. (Lo and Yeung 2002). These are usually represented in four fundamental forms —

(a) Numerical value, (b) Alpha-numerical characters or texts, (c) Symbols or Signs and (d) Signals.

Thus geographical data record the locations and characteristics of natural features or human activities that occurs on or near the earth's surface. It has two important properties —

(1) The reference to geographic Space, which means the data is registered to an accepted geographical co-ordinate system of earth's surface so that data from different sources can be spatially cross-referenced and integrated.

(2) The representation at geographical scale,

which means that the data is normally recorded at relatively small scales and as a result, must be generalized and symbolised.

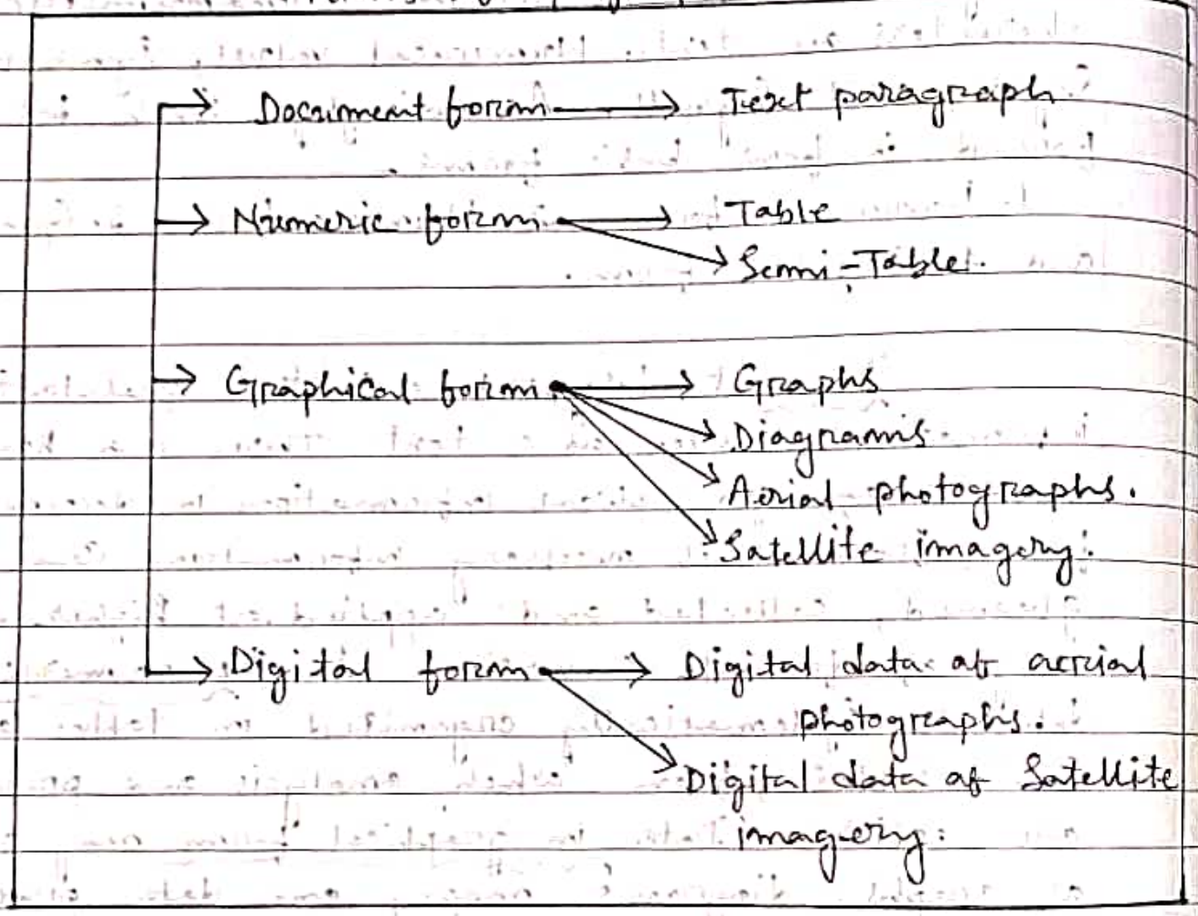
Geographical data is fundamentally recorded using four different means - alphanumeric characters as texts, Numerical values, Symbols as Signs, and Signals. Accordingly, Such data is found in four basic forms —

1. Document form, 2. Numeric forms, 3. Graphical form and 4. Digital form.

Document data refers to the qualitative information, written as a text. There is a huge amount of geographical information in document forms from which necessary information can be gleaned, collected and applied at higher levels by way of formats conversion. In the numeric form, data is systematically organised in tables and semi-tables, from which analysis and presentation are easier. Data in graphical form may occur as graphs, diagrams, maps, and data products. Remotely Sensed data, such as aerial photographs and Satellite images consists of a 2-dimensional array of ~~dist~~ discrete pixels (i.e. picture elements) or grid cells, which are ordered, in rows and columns. Each pixel has a digital number (DN that represents the intensity of the received Signal reflected or emitted by a given area of the earth surface. The size of the area belonging to a pixel is called the Spatial resolution. In digital form, such data can be stored, managed manipulated and analysed in three basic models

Raster (using grids / cells as reference in 2D), Vector (using points with rectangular Co-ordinates in 2D) and Surface (in 3D Co-ordinate System).

Forms of Geographical data.



Nature of Geographical data :

Geographical data concerns facts about the geographic reality and geographical facts are always inter-subjective and reliable. Therefore geographical facts are simply the objective records of observation. Thus, a geographer's data matrix (GDM) essentially refers to a matrix of individuals (objects or events) against attributes (various observations made on the attributes of

these individuals). Basically, a geographical individual is one identified by way of a space-time language. It may be a point (0-dimensional), a line (1-dimensional), an area (2 Dimensional), a volume (3-dimensional), a space-time volume (4-Dimensional) and so on.

Essentially, geographical data has spatial reference, i.e., attributes of the variables are collected on the basis of areal units. Generically, these units may be natural, e.g., plains, plateaus, lakes, etc. It may be artificial also, e.g., recording/monitoring-stations or simply the enumeration units of administration. It may be singular unit (e.g., individual households or landholdings) or collective units (e.g., a region made up of many such landholdings). The areal units are often arranged into some sort of hierarchical structure, e.g., India contains a set of States, each one of which is divided into a set of districts, that is again divided into a set of community development blocks (CDB) and each CDB contains a set of gram panchayats (GP); each one of which comprises a No. of landholdings and household units. At each level these have different but nested [(of similar objects of graduated sizes) placed or stored one inside the other) / Arranged in hierarchical structure] hierarchic identification code that makes them unique specially when managing the voluminous national level data like population census.]

In such a nested hierarchical situation, comparisons can only be made between similar

individuals occupying the same level in the hierarchy, viz - GP level, block level, district level, or State level. Inferences (a conclusion reached on the basis of evidence and reasoning) drawn about the relationships at one level can not be extended to any other level. Thus GP level inferences may not hold good for the State level. Thus, the nature of the analysis is contingent (subject to change) upon whether the individuals being compared or analysed are at the same or different levels. Individuals can be directly compared when they are at the same level of hierarchy. High-level to low-level analysis produces a contextual relationship and a low-level to high-level analysis produces an aggregate relationship. These are all interesting situations and each requires its own particular mode of thought about data collection, storage and procedure for manipulation and analysis.

Geographical data is characteristic of multivariate situations. It is often the failure of a geographer or a geographical data analyst not being able to comprehend the situation and act according to its philosophical basis. The attributes should be precisely defined and a finite number of significant attributes out of an infinite possibility should be carefully chosen for analysis. The decision regarding such choices, however depends upon the objectives and purpose of study and

Date / /

ultimately upon the theary itself.

Real world features occures in two fundamental ~~terms~~ forms - objects and phenomena. objects are discrete and definite, while phenomena are continuously distributed over a geographical space. Hills, rivers, forests, mines, highways, cities, bio-reserves etc. are examples of objects, and terrain, Slope, rainfall, temperature, pollution level and other environmental indices are phenomena. Spatial objects may be "exact objects" when they have well-defined boundaries (eg. land holding, building, etc.) and "inexact objects" or fuzzy entities, when Such boundaries are not well defined (eg. landform features and natural resource features). Graphically, Spatial objects are represented as three geometric elements - points, lines and polygons.

There are two distinct ways of the real world representation in geographical databases - the object based model and the field based model. The first one views geographic space as populated by Sets of Spatial objects, for which data is ~~of~~ obtained both by field Surveying and laboratory analysis. In Such Spatial databases, attributes are arranged againest locations ~~of~~ defined by co-ordinate lists or vector lines. Hence it is called the "Vector data Structure". The Second model on the otherhand, views geographic Space as populated by one or more Spatial phenomena. As Spatial phenomena take no obvious or Specific

Date / /

extent, these are basically represented as Surfaces and can be conceptualised as being made up of a number of Spatial data units. Data for Spatial phenomena are normally arranged against locations defined by elemental tessellation (an arrangement of shapes closely fitted together, especially of polygones in a repeated pattern without gaps or overlapping) in the form of Square or rectangular cells. Hence, Such data Structures are called "Raster data models". The field based Spatial data can be obtained either directly or indirectly. Data is acquired directly from aerial photography, Satellite imagery, map-Scanning and field measurements at Selected or Sampled locations. Data can also be generated indirectly by applying mathematical functions, e.g. - Contours and digital elevation models.