Brief bio

I was a professor in the department of Geo-Engineering, College of Engineering (Autonomous), Andhra University Visakhapatnam during 1998-2009. **29 students got Ph.D degrees** under my guidance. 41 research papers are published in international and National journals. I am a fellow of Indian Ggeophysical Union. I am a senior member IEEE,USA since 13 years. I was awarded **"Best Researcher award"** from Andhra University in 1995. I was also awarded **"Best Teacher award"** from Andhra Pradesh State Government in 2005. I have completed 5 research projects from various organization (AICTE, NIRD, DST ,ISRO and MoEF). I edited **three** books one on, Essentials of engineering geology , Applications of RS & GIS for coastal resources management and Applications of mathematical morphology for pattern studies. I authored one book on "*Feasibility and impact of left and right main canals of Godavari and Krishna of river basin transfer using Satellite data* . I did extensive consultancy work on Groundwater, Geophysical studies for Civil Engineering foundations to various Government organizations through University. For the last 10 years I have been engaged in development of flood models and vulnerability mapping of coastal areas. A 4 years Research Project completed on Inter-linking of Peninsular East flowing rivers. After retirement I was a principal of Private engineering college during 2009 to 2010. I was awarded Emeritus Professor by AICTE during 2010-12 and later UGC awarded Emeritus Professor during 2014 to 2016

Introduction to
Geogrphic information system and Formats

Scope

The digital formats described within the geospatial category of this web site can be used by geographic information systems (GIS) or other software applications to access, visualize, manipulate, and analyze geospatial data. Resources in these formats are primarily geospatial, i.e., they focus upon conveying information about the Earth, the location of specific features, and attributes and properties of those geo-located features. As such, the formats included in this geospatial content category will often comprise information in the following three forms: (i) raster or bit-mapped images; (ii) vector images consisting of points, lines, and polygons; and (iii) data values that express attributes associated with geographic locations or features. A defining characteristic for geospatial resources is that they are intended to be used by computer systems that enable spatial analysis.

The intended audience for this web site is the librarian, archivist, and/or data manager responsible for preserving digital resources. This essay is an introduction to geospatial formats and GIS functionality for the generalist or specialist geospatial data manager rather than for a GIS domain expert actively using the geospatial data. See also the summary overview Geospatial Content: Quality and Functionality Factors.

The descriptions of geospatial formats on this web site are intended to support the preservation of the data and its documentation and metadata, as received by a digital archive. The preservation goal is to facilitate future viewing or rendering of the data, and, to the extent known and agreed upon by the geospatial community, enable the re-use, re-analysis, and/or re-compilation of the data in the future.

Characteristics of Geospatial Formats

Geospatial resources are composed of data about geo-located features represented primarily by images (raster or vector) and tables or grids of observed or calculated attributes. Increasingly, geospatial formats include geospatially focused datasets or databases that contain primary information about a geographic location. In addition, ancillary and supplemental data that either are included or can be derived using spatial analysis are considered necessary for the full and effective functioning, interpretation and re-use of the data.

Geospatial formats have been and are continuously being specified and adopted by governmental organizations, software vendors, and standards-making bodies. These formats are often based on specifications or standards that are more general, e.g., for still images (raster and vector) and for datasets. For information about such formats and associated factors for assessing quality and functionality, see Still Images: Quality and Functionality Factors and Datasets: Quality and Functionality Factors. Common to geospatial formats are the capabilities for accurate representation of the described resource's location on the Earth using basic and inherent conceptual mechanisms such as georeferencing, scale, precision and accuracy.

Georeferencing. Georeferencing has been defined as the establishment of a relationship between information (e.g., documents, datasets, maps, images, biographical information) and geographic locations through mechanisms such as the addition of place labels (e.g., place codes or toponyms) or the assignment of geographic coordinates. Georeferencing must be understood as a multi-part process involving the concepts of geographic coordinates and two or three dimensional map projections. All of the terms in the following list bear on georeferencing.

- **Coordinates.** Geographic *coordinates* locate points in space. Two of the most commonly used geographic coordinate systems used are the latitude/longitude systems, and Universal Transverse Mercator (UTM) systems which place points on grids that divide sections of the Earth by various means.
- **Projection.** *Map projection* is necessary to correctly map points on Earth; the term refers to the process of mapping a three dimensional Earth onto a two dimensional planar surface such as a paper map or digital GIS. The Earth's surface is not only curved but is curved irregularly, thus making projection necessary. A declaration of the projection system that was used at time of creation is important although conversion to another projection that will be used in an existing projection framework is often done regardless of the original projection. For a useful overview, see *Projections: What You Need To Know for GIS* [3].
- **Datum.** Different approaches are used to represent the Earth's surface in a regular way that allows for mathematical calculation. The set of parameters and control points that are used to accurately pinpoint the three dimensional shape of the Earth is called a *datum*. Each datum is based on a particular view of the earth's shape as a spheroid or ellipsoid (the science of measuring the shape of the Earth is called *geodesy*). Along with its unit of measurement (e.g., feet or meters), a datum will vary depending upon whether it looks at the Earth from a primarily horizontal point of view (for a horizontal datum) or a vertical point of view (for a vertical datum), and where it begins the grid, i.e., its prime meridian. For a useful overview, see *Datum: What You Need To Know for GIS* [4].

  The fact that a particular datum was used is an essential characteristic of a geospatial resource. A declaration of which datum was used in data creation / collection is GIS metadata that should be represented in the data structure of a geospatial format. Users of a dataset may wish to transform the data from one datum to another for purposes of harmonization to a common view (shared by projects or sites, for instance). This is done using datum transformation methods often facilitated by a GIS or performed by external systems and hopefully documented within the GIS metadata that exists for a given dataset.

- **Map Scale.** The concept of *map scale* is also important to be able to judge how to view the data within a geospatial format. Defined as the ratio between linear distance on a resource and the corresponding distance on the surface being mapped, it can involve both *grain* (the size of a pixel and the smallest resolvable unit) and *extent* (the size of the study area and the largest resolvable unit). (See the glossary in Linda Hill's *Georeferencing: The Geographic Associations of Information* [2], p. 231, and *Scale in GIS: What You Need To Know for GIS* [5].) Information about scale where present and about the readability by humans or machines should be recorded and preserved to facilitate normal functionality.

  Map scale is also related to the important concepts of *accuracy* and *precision* as they are reflected in the data, and built into the data structures of a given geospatial format. Each of these factors is important to understand about geospatial data in order to judge its appropriateness for a planned use.
- **Accuracy.** *Accuracy* in the geospatial context generally measures how close an observed and recorded value is to the true value. Spatial accuracy is measured in four primary ways: *positional accuracy*, *attribute accuracy*, *logical consistency*, and *completeness*. It is by *positional accuracy* that one knows how close the recorded location is to the real location. *Attribute accuracy* allows one to measure (by means of error and percentage calculations, usually) how close to reality are the attributes recorded for a described location. *Logical consistency* reflects the presence, absence or frequency of inconsistent data, usually determined by comparison of themes or juxtaposition of facts. Finally, *completeness* describes how fully the data describes the location and features about the location that it is intending to represent.
- **Precision.** *Precision* in this context refers to the consistency of a measurement method, measured by how often the same results are achieved when measuring. It is usually defined in terms of how dispersed a set of repeat measurements are from the average measurement.

GIS Metadata and Data Documentation

What makes geospatial formats different from other formats is the fundamental capability of placing information in relation to the surface of the Earth either both horizontally and vertically, or horizontally with height or depth implicit. Important characteristics to be recorded in metadata include geographic coordinates, projection, scale, and datum since the digital resources are not comprehensible or usable with normal functionality without it. Even a simple raster image that is a product of a dataset or other format, such as a simple PDF or bit-mapped image of a map should also include information on the map that would allow the viewer to determine how accurately the map corresponded to its geographic "reality," or determine its accurate location from the data behind the map.

Content Standards. Community based content standards exist for geospatial data such as the U.S. Federal Geospatial Data Committee's *Content Standard for Digital Geospatial Metadata* (FGDC) [7] and the broader ISO standard for geographic information, ISO 19115:2003 [8]. In the U.S., FGDC's content standard elements have has been adopted more widely as a result of being incorporated into commonly used software products by such domain giants as ESRI and GeoMedia. (ESRI uses a "profile" of the FGDC standard rather than a native FGDC XML schema.) ISO 19115:2003 is slowly being adopted by more U.S.

federal agencies and international agencies and is beginning to be incorporated into common software packages.

Quality Factors. For some geospatial resources such as satellite data, it is critical to proper understanding and use of the data to include information about the quality of the data as well as its provenance and lineage. For example, for satellite images the percentage of cloud cover is a significant quality characteristic. Such information is critical not only to understanding what the data says, but also to understand appropriate and inappropriate uses of the data.

Levels of Data Quality. There can be many levels at which data quality is and should be documented. For example at the product level, it is key to know how closely the data represents the actual geophysical state given the output from different instruments. Another quality level would be at the pixel level where the algorithms used to create the data points are noted as well as an assessment of the usability of those data points. At the granule-level, statistical roll-up of pixel-level data is compiled. This kind of computation could be important to validate the model used. For example, climate change data models can have grids of contiguous data tagged with uncertainty statistics for each grid cell, thus providing the means to assign quantitative risk factors or uncertainty levels to different mitigation scenarios. Examples of data quality reports for a data set can be found at NASA's *NASA Surface Meteorology and Solar Energy: Accuracy* [12]. The assessment of bias is a key data quality factor, i.e., bias that is generated from the instruments used (instrumental bias), or the type of sampling or observations made that provide the view of the data produced. In addition, an assessment of appropriate and/or inappropriate use is often considered to be an important data quality consideration.

Provenance and Lineage. Documentation about the *provenance* of data in terms of factual establishment of its authorship is usually considered to be quite important in order to determine the authenticity of the data, and to some extent its accuracy. For example, knowing the name of the organization and/or person(s) responsible for the creation and/or collection of data may help ascertain whether or why certain features are or are not present, such as roads or buildings on a map of a city. A data consumer would have more confidence in the accuracy of such a map if it had been created by the city's data center rather than by a student at a local college or university. Another example of describing the provenance of data is the tracking of what instrumentation was used to generate or record the data and the algorithms used to calculate the data output.

There are many applications of GIS.

Facility management, planning, environmental monitoring, population census analysis, insurance assessment, and health service provision, hazard mapping and many other applications. The following list shows few applications in natural resource management:

- Agricultural development

- Land evaluation analysis
- Change detection of vegetated areas
- Analysis of deforestation and associated environmental hazards
- Monitoring vegetation health
- Mapping percentage vegetation cover for the management of land
- Crop acreage and production estimation
- Wasteland mapping.
- Soil resources mapping
- Groundwater potential mapping
- Geological and mineral exploration
- Snow-melt run-off forecasting
- Monitoring forest fire
- Monitoring ocean productivity etc.
- GIS application in Forestry.

With the rise of World Wide Web, new Internet protocols such as the Hypertext Transfer Protocol (HTTP), as well as easy to use interfaces (browsers), tools and languages (HTML, XML, and Java), the Internet has become a hub for GIS functionalities from the client side without even any GIS software. The GIS field is still evolving and it will be the major force in various walks of life dealing with geographic information.