# 1. INTRODUCTION.

**Geography** is the study of Earth's features and patterns of their variations in spatial location and time. Many questions of agricultural production are geographic in nature as the production depends on the environment and prevailing socio economic conditions, both of which vary spatially and in time. Examples are questions related to natural resources management, precision agriculture, agro-ecological classification for land use planning, regional trends and patterns in technology adaptation, agricultural productivity and income, non-Pont source pollution from agricultural lands, etc. Answering these questions requires access to large volumes of multidimensional geographical (spatial) information of weather, soils, topography, water resources, socio economic status, etc. Further, answers to even apparently simple questions require that the data from several sources be integrated in a consistent form. Geographical Information Systems or GIS enable representation and integration of such spatial information.

**Geographic information systems** (GIS) (also known as Geospatial information systems) are Computer Software and hardware systems that enable users to capture, store, analyze and manage spatially referenced data. GISs have transformed the way spatial (geographic) data, relationships and patterns in the world are able to be interactively queried, processed, analyzed, mapped, modelled, visualized, and displayed for an increasingly large range of users, for a multitude of purposes.

In a general sense, the term describes any information system that integrates, stores, edits, analyzes, shares, and displays geographic information. GIS applications are tools that allow users to create interactive queries (user-created searches), analyze spatial information, edit data in maps, and present the results of all these operations. Geographic information science is the science underlying geographic concepts, applications, and systems

It is a special case of information system where the database consists of observations on spatially distributed features, activities or events, which are definable in space as points, lines or area. A geographic information systems manipulates data about these points, lines and areas to retrieve data for ad hoc queries and analyses.

Whether siting a new business, finding the best soil for growing bananas, or figuring out the best route for an emergency vehicle, local problems also have a geographical component GIS will give you the power to create maps, integrate information, visualize scenarios, solve complicated problems, present powerful ideas, and develop effective solutions like never before. GIS is a tool used by individuals and organizations, schools, governments, and businesses seeking innovative ways to solve their problems.

Today, GIS is a multibillion-dollar industry employing hundreds of thousands of people worldwide. GIS is taught in schools, colleges, and universities throughout the world. Professionals in every field are increasingly aware of the advantages of thinking and working geographically.

- ☑ **Geographic:** 80% of government data collected is associated with some location in space.
- ☑ **Information:** attributes, or the characteristics (data), can be used to symbolize and provide further insight into a given location.

☑ **System:** a seamless operation linking the information to the geography – which requires hardware, networks, software, data, and operational procedures.

**GIS is:**

- Not just software!!!
- Not just for making maps!!!

<u>*GIS can be used as synonymous to*</u>:

- **GI Systems**: A computerized tool that helps solve geographic problems.
- **GI Science**: The development of data models, algorithms, and methods for representing geography and spatial relationships in order to support spatial analysis and location-based computing
- **GI Studies**: The systematic study of society's use of geographic information, including institutional, organizational and procedural issues.
- **GI Services**: The business of providing GIS data and analysis tools to GIS users (often by chaining interoperable components in logo-block fashion).

## 1.1. OVERVIEW, HISTORY AND CONCEPT OF GIS

### 1.1.1. Overview

A geographic information system consists of the tools and services necessary to allow one to collect, organize, manipulate, interpret, and display geographic information. A GIS is more than just the hardware and software familiar to most people; it extends to the staff who operate the system, the databases, the physical facilities, and the organizational commitment necessary to make it all work. A GIS can be defined by how it is used (e.g., a land information system, a natural resource management information system), by what it contains (spatially distinct features, activities, or events defined as points, lines, polygons, or raster grid cells), by its capabilities (a powerful set of tools for collecting, storing, retrieving, transforming, and displaying spatial data), or by its role in an organization (a map production system, a spatial analysis system, a system for assisting in making decisions regarding basic geographic questions: where is it? what is it? why is it there?).

GIS can also be defined as geographic information science (GIScience). **GIScience** involves the identification and study of issues that are related to GIS use, that affect its implementation, and that arise from its application (Goodchild 1992). In short, GIScience not only encourages users to understand the benefits of GIS technology in providing a powerful set of analysis tools but also encourages users to view the technology as part of a broader discipline that promotes geographic thinking and problem-solving strategies as being useful to society. The development of GIScience is an outgrowth of the fact that GIS technology is available to more users today than ever before and that spatial categorization and analysis is applicable to many issues and problems of society. Regardless of how a GIS is perceived or used, it is the integration of the various tools and services that leads to a successful GIS. Although other software programs perform GIS-like tasks (e.g., database management, graphics, or computer-assisted drafting [CAD] Software), a GIs is unique in its ability to allow users to create, maintain, and analyze geographic or spatial data. The term spatial data implies that data exist - which not only describe landscape features (e.g., condition, composition, structure of forests) but also reference the location where the features can be found. A GIs allows one to manipulate spatial data and to analyse quickly a large volume of spatial data. A GIS stores spatial data in a digital database files, often referred to as themes, maps, covers, and tables, layers, or GIS databases. The terminology for referring to a GIS database varies, depending on the GIs software program being used. In most GIs software programs, similar landscape features are maintained in a single GIS database. For instance, one may have a soils GIS database, which contains the soil characteristics of a landscape, or a wildlife GIS database, which contains the nest locations of a single species of animal or a group of species of animals. GIs allows the integration and simultaneous examination of multiple GIS databases through a process described as overlay analysis. Overlay analysis represents the essence of what many consider to be a primary role of GIS in natural resource management: the ability to combine two or more GIS databases to assist in making decisions. GIS is related to a number of other fields and disciplines, including computer-assisted drafting (CAD), computer cartography, database management, and remote sensing. In fact, GIs contains certain aspects of each of these fields and, thus, is closely related to each of them. However, the differences between GIS and these other allied fields is notable. For example, most CAD software programs have rudimentary links to a database management system, whereas GIs software programs generally have strong links to a database management system. The field of **computer cartography** emphasizes map production, but although the databases used may be similar to those used in GIS, computer cartography generally puts less emphasis on the nongraphic attributes of spatial landscape features than does GIS.

Database management software programs have the ability to store and manage the location and attribute data of landscape features but generally lack the display capabilities of GIS, CAD, and computer cartography.

### *Types of GIS*

- There are a number of Geographical Information Systems (GIS) (or GIS software) available today.
- They range from high-powered analytical software to visual web applications, and each of those are used for a different purpose.
- These can be categorized into 3 main groups of GIS:
    1. Web-based GIS
    2. Geobrowser
    3. Desktop GIS

**Web-based GIS**

- Web-based GIS, or WebGIS, are online GIS applications which in most cases are excellent data visualisation tools.
- Their functionality is limited compared to software stored on your computer, but they are user-friendly and particularly useful as they not required data download.

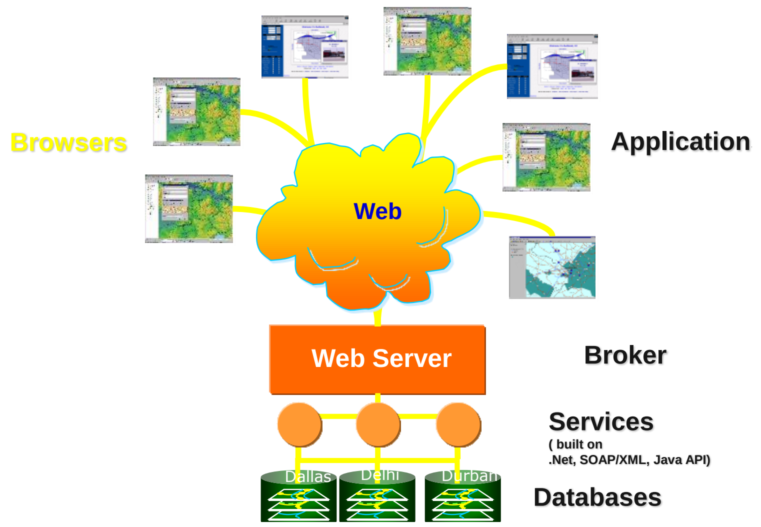



Fig: ***General Web-based GIS Application Architecture***

- There are different map servers and mapping client application that can be used while developing WebGIS applications.

***Web map servers***

- **GeoServer:**
  - ✓ Written in Java and relies on GeoTools. Allows users to share and edit geospatial data.
- **MapGuide Open Source:**
  - ✓ Runs on Linux or Windows, supports Apache and IIS web servers, and has APIs (PHP, .NET, Java, and JavaScript) for application development.
- **Mapnik:**
  - ✓ C++/Python library for rendering - used by OpenStreetMap.
- **MapServer:**
  - ✓ Written in C. Developed by the University of Minnesota.

***Software development frameworks and libraries***

- **GeoBase (Telogis GIS software):**
  - ✓ Geospatial mapping software available as a Software development kit.
  - ✓ Suited for high transaction enterprise environments.
- **Geomajas:**
  - ✓ Open source development software for web-based and cloud based GIS applications.
- **MapFish:**
  - ✓ Aggregates the power of OpenLayers, ExtJS and GeoExt.
- **OpenLayers:**
  - ✓ Open source AJAX library for accessing geographic data layers of all kinds, originally developed and sponsored by MetaCarta.
- **Leafletjs:**
  - ✓ Open-Source JavaScript Library for Mobile-Friendly Interactive Maps

## Geobrowser

- A Geobrowser is better explained with reference to a Web-browser.
- In short, a geobrowser can be understood as an Internet Explorer for geographic information.
- Like the internet it allows the combination of many types of geographic data from many different sources.
- The biggest difference between the World Wide Web and the geographic web however is that everything within the latter is *spatially referenced*.
- Google Earth is the most popular geobrowser available.

## Desktop GIS

- A GIS, or GIS software, allows you to interactively work with spatial data.
- A desktop GIS is a mapping software that needs to be installed onto and runs on a personal computer.
- ArcGIS is what ESRI refer to as a suite of products which can be tailored to your need.
- ArcGIS is used for a vast range of activities, covering both commercial and educational uses.

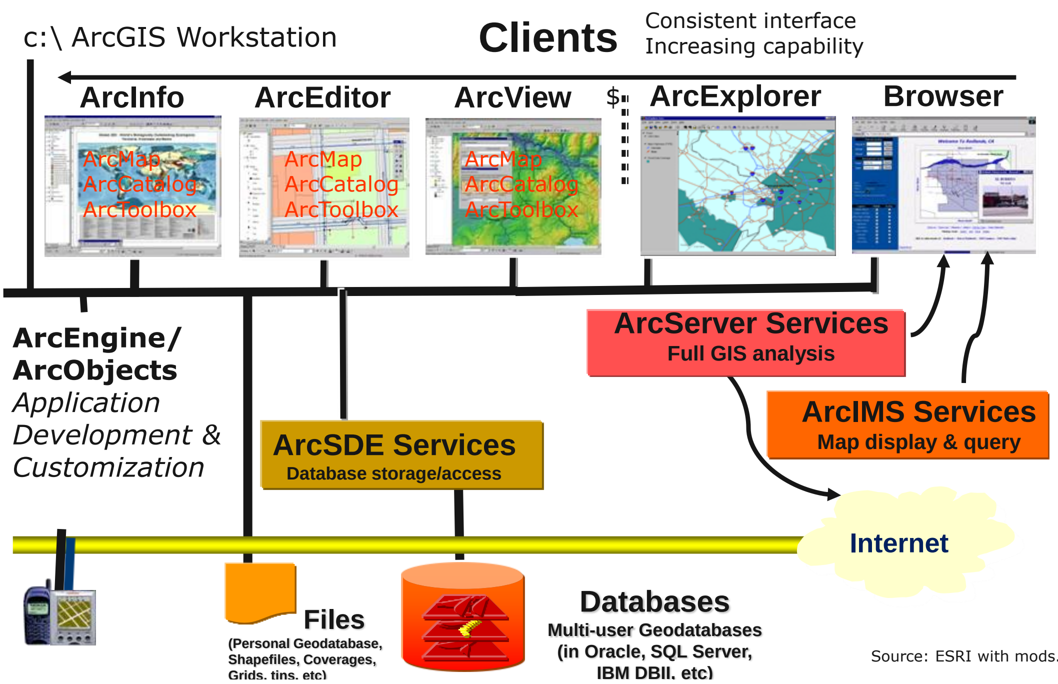



**Fig:** *System Architecture of Arc GIS*

- Companies with high market share
  - **Esri:**
    - ✓ Products include ArcMap, ArcGIS, ArcSDE, ArcIMS, ArcWeb services and ArcGIS Server.
  - **ERDAS IMAGINE by ERDAS Inc:**
    - ✓ Products include Leica Photogrammetry Suite, ERDAS ER Mapper, ERDAS ECW/JP2 SDK (ECW (file format)) are used throughout the entire mapping community (GIS, Remote Sensing, Photogrammetry, and image compression) and ERDAS APOLLO.

- **MapInfo by Pitney Bowes Software:**
  - ✓ Powerful desktop GIS MapInfo Professional is enhanced with many plug-ins including MapInfo Drivetime for route analysis, MapInfo Engage 3D for 3D and statistical analysis, MapInfo Mapmaker for Geocoding

### 1.1.2. History

GIS is unique in its ability to process, map, and analyze spatial data.

**During in 1400:** In Egypt, Spatial data have been collected and maintained with records of property boundary surveys for taxation purposes. Although the term geographic information system was first used in the 1960s, overlay analysis has been demonstrated through manual techniques for over 200 years.

During the American Revolution, French cartographer Louis-Alexandre Berthier overlaid multiple maps to analyze troop movements (Wolf and Ghilani 2002).

**During In 1854:** Dr. John Snow compared the locations of cholera deaths with well locations in London to determine if well water was related to cholera infections.

**During In 1954:** The first written description of how to combine multiple maps precisely through a manual overlay process appeared in a 1954 text titled Town and Country Planning Textbook, by Jacqueline Tyrwhitt (Steinitz et al. 1976).

**During the 1960s:** Organizations in the United States (including the U.S. Geological Survey and the U.S. Department of Agriculture's Natural Resource Conservation Service) began to create GIs databases of topography and land cover (Longley et al. 2001).

Students and researchers began to write computer programs and design hardware devices (such as the precursor to today's digitizing table) that would allow one to trace the outlines of landscape features on hardcopy thematic maps and transfer them into a digital format. These early programs were designed to handle specific tasks and were often limited in scope.

As programmers began to combine these algorithms to create more versatile and powerful software programs, the era of computer mapping applications began.

Early examples of mapping programs include IMGRID, CAM, and SYMAP (Clarke 2001). In conjunction with the development of software programs, other organizations began to assemble GIs databases. The first examples were the GIS databases created by the U.S. Central Intelligence Agency (CIA) called the "World Data Bank." Spatial features in the GIs database included coastlines, major rivers, and political borders from around the world.

**During in 1964:** Ian McHarg described how to use a series of transparent overlays to determine the suitability of areas for development in New York's Staten Island. By using a transparent overlay for each layer of interest (soils, forests, parks, etc.) and blacking-out the areas on each overlay that presented development impediments, the layers could be overlaid and the final suitable areas defined.

McHarg (1969) later published examples of his overlay techniques in his seminal book, Design with Nature.

In the early stages of the development of GIs technology, two facts were evident: there were few geographic or spatial databases to work with, and the technology to store and manipulate the data was rudimentary (by today's standards).

**During 1970:** The U.S. Census Bureau designed a methodology for linking census information to locations in preparation for the 1970 U.S. census. The 1970 U.S. census was the first census that was mailed, and the only piece of information that was returned to reference the location of the respondent was the address. The Census Bureau, however, was faced with the challenge of matching the addresses on the responses to a map, so that the spatial distributions of the responses could be analyzed. To meet this challenge, the Census Bureau developed a system known as **Dual Independent Map Encoding (DIME)** which allowed them to create digital records of all streets and to associate addresses to street locations. The DIME system allowed the Census Bureau to understand which streets were connected to which other streets, as well as what landscape features were adjacent to each street.

The description or characterization of the spatial relationships between landscape features in a GIs database is referred to as topology. Topology is an important concept with respect to GIs applications. Topology allows one to organize and analyze objects according to their location and to their proximity with other objects. The topological characteristics of data structures allow a determination, for example, of which roads in a forest are connected to which other roads and which timber stands are next to which other timber stands.

**During 1988:** The DIME system was the predecessor of *Topologically Integrated Geographic Encoding and Referencing System **(TIGER)*** files, which were introduced by the Census Bureau in 1988 and are still used today to distribute spatially referenced census and boundary data. The availability of TIGER files was instrumental in promoting GIs use in the United States.

Another important contribution to spatial data availability in the United States was made by the **U.S. Geological Survey (USGS)**. The USGS began digitizing features from its 1: 100,000 scale hardcopy maps in the early 1980s. Spatial data from these maps were made available as digital line graphs (DLGs), which, like the TIGER and DIME systems, were also stored in a file format that allowed for the topology of objects to be characterized. The file format was restructured in the early 1990s, and features from finer-resolution 1:24,000 scale maps were later made available for small portions of the country by the USGS. To manage and analyze spatial data for their jurisdictions,

**During 1960's:** Canadian and U.S. organizations began to develop software programs in the 1960s. One of the most ambitious of these systems was the **Canada Geographic Information System (CGIS)**, which was created under the guidance of Roger Tomlinson in 1964. A chance meeting on a plane flight between Tomlinson and Canada's Minister of Agriculture resulted in Tomlinson overseeing the creation of a national effort to inventory Canada's land resources and to develop a software program to quantify existing and potential land uses. The CGIS is recognized as being the first national-level GIS; thus, Tomlinson continues to receive recognition as a GIs pioneer for his efforts.

Other early landmark efforts in the evolution of GIs include the development of the **Land Use and Natural Resource Inventory System (LUNR)** in New York in 1967 and the development of the Minnesota Land Management System in 1969. The success of these early systems and the need for further refinements were recognized by a group of faculty and students at Harvard University's Laboratory for Computer Graphics and Spatial Analysis. The group set forth to create a versatile GIS that would map and track locations as the DIME system does, while possessing the land measurement strengths of the CGIS. From this effort, the Odyssey GIs (containing modules named after parts of Homer's epic work The Odyssey) emerged in 1977 and pioneered the use of a data structure known as the archode or vector data structure. It is important to note that the specifics of the Odyssey vector structure were first published by Peucker and Chrisman (1975), and the structure continues to influence the design of modern GIs software programs.

**During 1980:** Jack Dangermond, a Harvard Laboratory student, founded the **Environmental Systems Research Institute (ESRI)** in 1969, and ArcView and ArcInfo, the most widely used desktop and workstation GIs software programs, are based on the Odyssey vector data structure. ArcInfo, in fact, was introduced in 1981, marking the first major commercial venture into the development of GIs technology. The 1980s also witnessed the proliferation of microcomputers, today's version of the personal computer (PC).

**During 1986:** In response, software manufacturers began to produce GIs software programs that could operate on the microcomputer. In 1986, *MapInfo* Corporation was formed and subsequently developed the world's first major desktop vector GIs software program for the PC. Soon afterward, raster GIs software programs, such as IDRISI, began to appear. Some software programs, such as the raster GIs program GRASS, use a software architecture developed for workstation computer platforms. Other significant developments during the late 1970s and early 1980s were the emergence of GIS related conferences and publications.

The first Auto- Carto Conference was held in 1974 and helped establish the GIs research agenda. One of the first compilations of available mapping programs was published by the International Geographical Union in 1974. Basic Readings in Geographic Information Systems was published in 1984, containing a collection of papers discussing GIs technology.

In 1986, the first textbook written specifically for GIs, Principles of Geographic Information Systems for Land Resources Assessment, was published (Burrough 1986). Finally, the first GIs-related academic journal, the International Journal of Geographic Information Science, was published in 1987. The history of GIs continues to evolve, with GIs users providing a number of challenges. GIs users, for example, have the ability to influence the development of GIs software program features. As new and challenging natural resource management issues arise, users identify and propose processes and functions that will make the task of analyzing potential natural resource decisions more efficient and accurate. In addition, GIs users increasingly expect support and training related to specific GIs software programs, and they expect that the software will be perfected by the time of its release to the general public. Further, because GIs databases are shared data among organizations, the need to standardize data formats is evident, as data transformations can require an extensive commitment of time and resources.

Society is fortunate today, on one hand, to have a variety of GIs software programs from which to choose. On the other hand, evaluating which of these programs best suits the needs of a natural resource management. Organization is problematic. Since organizations (natural resource management as well as academic) may use different GIs software programs, a decision was made to develop as a general reference for describing the typical types of GIs applications faced by 'field-level professionals associated with natural resource management organizations.

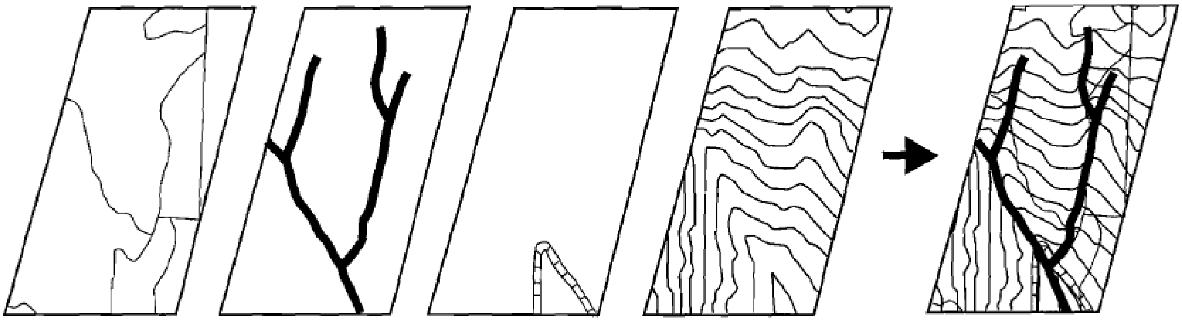



Fig: GIS theme overlay

### 1.1.3. Concept of GIS

#### *How GIS Works?*

A simple five-step process lets you apply GIS to any business or organizational problem that requires a geographic decision. The steps are as follows:

- ✓ **Ask**
- ✓ **Acquire**
- ✓ **Examine**
- ✓ **Analyze**
- ✓ **Act**

**Ask**

- What is the problem you are trying to solve or analyze, and where is it located?
- Framing the question will help you decide what to analyze and how to present the results to your audience.

**Acquire**

- Next you need to find the data needed to complete your project.
- The type of data and the geographic scope of your project will help direct your methods of collecting data and conducting the analysis.

**Examine**

- You will only know for certain that your data is appropriate for study after thoroughly examining it.
- This includes how the data is organized, how accurate it is, and where the data came from?

**Analyze**

- Geographic analysis is the core strength of GIS.
- Depending on your project, there are many different analysis methods to choose from.
- GIS modeling tools make it relatively easy to make these changes and create new output.

**Act**

- The results of your analysis can be shared through reports, maps, tables, and charts and delivered in printed format or digitally over a network or on the web.
- You need to decide on the best means for presenting your analysis, and GIS makes it easy to tailor the results for different audiences.

While creating any GIS applications you should follow the above five steps very carefully and thoroughly. Each steps are very important and have its own significance on overall application development. If you don't pay attention to any one of the above steps the overall result that you are supposed to get will be different.

## 1.2. SCOPE AND APPLICATION AREAS OF GIS

- GIS technology can be used for scientific investigations, resource management, asset management, archaeology, environmental impact assessment, urban planning, cartography, criminology, geographic history, marketing, logistics, Prospectivity Mapping, and other purposes.
- For example, GIS might allow emergency planners to easily calculate emergency response times in the event of a natural disaster, GIS might be used to find wetlands that need protection from pollution, or GIS can be used by a company to site a new business location to take advantage of a previously under-served market.
- Uses of GIS range from indigenous people, communities, research institutions, environmental scientists, health organizations, land use planners, businesses, and government agencies at all levels.
- Uses range from information storage; spatial pattern identification; visual presentation of spatial relationships; remote sensing - all sometimes made available through internet web interfaces, involving large numbers of users, data collectors, specialists and/or community participants.
- One of the primary services provided by a GIS project is the geo-referencing of various data layers for mapping projection, involving the use of satellite image data for GIS mapping including:
  - ☑ Mineral Mapping
  - ☑ Pipeline Corridor Mapping
  - ☑ Defence Mapping
  - ☑ Airport Mapping
  - ☑ Land Cover Classification
  - ☑ Urban Development
  - ☑ Pre and post 2D/3D seismic surveys
  - ☑ Environmental Impact Studies (EIS)
  - ☑ Coastal erosion studies
  - ☑ Cadastre Mapping
  - ☑ Disaster Analysis

### *Mineral Mapping*

Satellite Imagery and aerial photography have proven to be important tools in support of mineral exploration projects. They can be used in a variety of ways. Firstly they provide geologists and field crews the location of tracks, roads, fences and inhabited areas. This is important for mapping out potential access corridors for exploration areas and considering the environmental impact of large project. These images are also useful for mapping outcrops and regolith systematics and vegetation cover across exploration blocks and over regional areas.

### Pipeline Corridor Mapping

Improve Safety and Security for Pipeline and Transmission Surveys: Satellite imagery and GIS data have significant potential to reduce a number of safety and security issues for pipeline corridor planning as well as supply managers with solutions through spatial representation of data for land, lease management, exploration, production, transmissions, environmental, financial and facilities management. This information is required to make decisions that will significantly impact the operator's ability to provide the services demanded by their customers.

### Defence Mapping

Defence and Security

Satellite imagery and GIS maximizes security programs which can enable local governments to better assess and understand how to develop programs to save lives, protect property and enhance the future economic stability of their communities. The current threats to a country range from incidents of terrorism and information attacks on critical infrastructure to the potential use of weapons of mass destruction and the spread of infectious diseases. Each one of these threats could cause massive casualties and disruption to a country.

### Airport Mapping

Airport Mapping Database

Satellite Imaging Corporation (SIC) provides 3D airport mapping using high resolution stereo satellite imagery to support airport pre-planning and design, airport layout plans (ALPs), navigational mapping, and airport security and aviation safety operations.

3D Digital Surface Models (DSM's) and Digital Terrain Models (DTM's) can be created to support airport and aviation operations to provide details and data for the construction of airport runways, airport terminals, airport layout design, airspace analysis, obstruction surveys, facility mapping, taxiways, aprons/parking areas, 3D flight simulation for pilot training, aircraft operations, and GIS database development.

### Land Cover Classification

Satellite Imagery and GIS for Land Cover and Change Detection

Satellite imagery and GIS maps for land cover, land use and its changes is a key to many diverse applications such as environment, forestry, hydrology, agriculture and geology. Natural Resource Management, Planning and Monitoring programs depend on accurate information about the land cover in a region. Methods for monitoring vegetation change range from intensive field sampling with plot inventories to extensive analysis of remotely sensed data which has proven to be more cost effective for large regions, small site assessment and analysis.

Evaluation of the static attributes of land cover (types, amount, and arrangement) and the dynamic attributes (types and rates of change) on satellite image data may allow the types of change to be regionalized and the approximate sources of change to be identified or inferred.

### Urban Development

Satellite imagery for urban and land development can be used to gather strategic planning information pertaining to a district or an entire city. High resolution satellite imagery and LiDAR incorporated into a GIS (Geographic Information Systems) and CAD (Computer Aided Drafting) has gained popularity among Planners, Developers and Engineers for large scale mapping of any region for most urban and land development applications.

Information from satellite images when combined with GIS mapping is used for analysis in evaluating construction costs as well as environmental impact of alternative routes for utility and transport corridors; land cover and land use classification; identifying population groups at risk where human intervention is most needed to limit and prevent hazards during development stages.

### *Pre and post 2D/3D seismic surveys*

No matter how remote, Satellite Imaging Corporation (SIC) can retrieve satellite images from the most difficult-to-photograph areas of the world. For heavily forested areas, we provide medium-to-high resolution "Bare Earth" DEMs. This provides weather independency, allowing us to map large areas of terrain in limited timeframes, independent of the weather and solar illumination conditions. We are also familiar with specialized retrieval methods used for satellite imagery in remote areas, highly developed areas and areas of persistent heavy cloud cover such as the tropics.

### *Environmental Impact Studies (EIS)*

Satellite Imaging Corporation (SIC) provides satellite image data at different spatial, spectral, and temporal resolutions by using the appropriate combination of bands to bring out the geographical and manmade features that are most pertinent to your project for detecting and monitoring environmental changes.

Satellite imagery and GIS have greatly expanded opportunities for data integration, analysis, modelling, and map production for environmental monitoring and assessment. As populations grow, as countries boost their economies, as landscapes change, governments have increasingly relied on up-to-date satellite imagery and other geospatial data for applications such as environmental planning, land registration, disaster response, public health, agricultural biodiversity conservation and forestry

### *Coastal erosion studies*

Many coastal managers are changing the way they manage coastal problems. Instead of only undertaking corrective measures, officials are moving toward prevention. Using potential models with satellite imaging technology and land cover data through GIS, managers can create scenarios for future development, as well as permitting and land use scenarios, to estimate the impacts on sensitive water bodies.

Satellite images can provide coastal management researchers and scientists with data for assessment and analysis of water temperature, salinity, phytoplankton, hydrology, shoreline changes, bathymetry, soil moisture and potential threats to our coasts.

Assessments and predictive capabilities through satellite imagery incorporated with GIS are needed to predict onset of events that may significantly affect human health, critical wetlands and ecosystems, and economic development.

### *Cadastre Mapping*

Satellite images which forms the base for the generation of action plan maps, if used in the background of intelligent cadastral vector data, can improve the details of the thematic maps as well as action plan maps. It also helps in the monitoring of land cover changes that can be identified by detailed change detection processing procedures and implemented in the GIS cadastre mapping project.

### *Disaster Analysis*

Satellite imagery and GIS maps can give emergency and disaster response officials a wealth of information for assessment, analysis and monitoring of natural disasters such as hurricanes, tornadoes and cyclone damage from small to large regions around the globe.

Estimates of the particular land cover classes that may be inundated by a natural disaster can enable operators and planners to better assess their region's risk and vulnerability. This information will allow for prioritizing target mitigation and preparedness activities for their area.

The use of multispectral satellite imagery is therefore critical for the separation of constituent materials within an image and for the interpretation of images of damage for pre or post-disaster assessment. View Before and after example satellite images here.

## 1.3. PURPOSE AND BENEFITS OF GIS

### 1.3.1. Purpose of GIS

**What does GIS do?**

- What does GIS actually do? Several things, in fact:
- It allows users to map multiple different sources of geographic data within a single computerized environment.
- Different data sources are usually treated as layers, which may be reordered and switched on and off at will, set to varying transparencies, and manipulated through tools such as zooming, panning, and sometimes rotating.
- It allows users to employ many different & powerful tools to analyze the spatial distribution of their data.
- This spatial analysis can provide a route into discovering and unlocking previously unseen patterns in our data, shedding new light on unknown aspects of the past.
- It also allows users to produce paper and electronic maps for inclusion in their work and for the dissemination of their results to the wider archaeological, historical and public communities.
- Depending on the GIS software used, this might include animations or interactive maps delivered over the internet.

Thus, essentially, GIS provides the tools to integrate spatial data, to analyze spatial data, and to present spatial data to a wider audience.

**What does GIS do?**

- GIS is not a panacea that can solve all of our problems.
- It has several drawbacks and difficulties that must be borne in mind by any user who wishes to produce good quality, authoritative results:
    - It is easy to disguise poor quality data by entering it into a GIS, resulting in maps that convey an undue authority.

    - Users should make very clear any concerns that they have about data quality, either in the description of any published map, and preferably also through the use of appropriate symbology.
    - For example, any uncertainties associated with objects can be graphically expressed through careful choice of symbols (e.g. half-filled squares for possible Roman fort sites), or through the intelligent use of transparency.
- Electronic maps output from a GIS may need tweaking in picture editing software to produce the best results for publication.
- Many of the tools provided by GIS packages can be applied to data where their use would not be appropriate.
- GIS tools can be used to support distinctly spurious ideas and to cloud what might normally be seen as unconvincing conclusions.
- Finally, the most widespread GIS packages only express a fraction of the true spatial complexity of the world around us.
- This is because the third dimension is only just starting to be properly represented and, furthermore, time is entirely absent from the majority of conventional GIS packages.
- Time is what separates geography from geometry, so current GIS software will remain incomplete until their developers begin to integrate temporality

**GIS can be used to:**

- Explain events
- Planning Strategies
- Integrate Information
- Solve complicated problems
- Predict outcomes
- Create “smart” maps
- Visualize scenarios
- Present powerful ideas

**Who Uses GIS?**

- Police and Law Enforcement Agencies
- Planning Strategies
- Foresters
- Industry
- Environmental Engineers
- Real Estate Professionals
- Telecommunications Professionals
- Emergency Response Organizations
- Local and Federal Government
- Health
- Transportation
- Geographers
- Market Developers

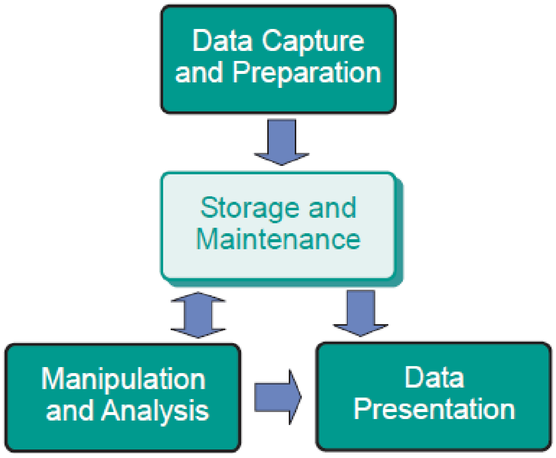
### *Top Five Benefits of GIS*

GIS benefits organizations of all sizes and in almost every industry. There is a growing interest in and awareness of the economic and strategic value of GIS, in part because of more standards-based technology and greater awareness of the benefits demonstrated by GIS users. The number of GIS enterprise solutions and IT strategies that include GIS are growing rapidly.

The benefits of GIS generally fall into five basic categories:

1. Cost savings resulting from greater efficiency. These are associated either with carrying out the mission (i.e., labour savings from automating or improving a workflow) or improvements in the mission itself. A good case for both of these is Sears, which implemented GIS in its logistics operations and has seen dramatic improvements. Sears considerably reduced the time it takes for dispatchers to create routes for their home delivery trucks (by about 75%). It also benefited enormously in reducing the costs of carrying out the mission (i.e., 12%-15% less drive time by optimizing routes). Sears also improved customer service, reduced the number of return visits to the same site, and scheduled appointments more efficiently.
2. Better decision making. This typically has to do with making better decisions about location. Common examples include real estate site selection, route/corridor selection, zoning, planning, conservation, natural resource extraction, etc. People are beginning to realize that making the correct decision about a location is strategic to the success of an organization.
3. Improved communication. GIS-based maps and visualizations greatly assist in understanding situations and storytelling. They are a new language that improves communication between different teams, departments, disciplines, professional fields, organizations, and the public.
4. Better geographic information recordkeeping. Many organizations have a primary responsibility of maintaining authoritative records about the status and change of geography (geographic accounting). Cultural geography examples are zoning, population census, land ownership, and administrative boundaries. Physical geography examples include forest inventories, biological inventories, environmental measurements, water flows, and a whole host of geographic accountings. GIS provides a strong framework for managing these types of systems with full transaction support and reporting tools. These systems are conceptually similar to other information systems in that they deal with data management and transactions, as well as standardized reporting (e.g., maps) of changing information. However, they are fundamentally different because of the unique data models and hundreds of specialized tools used in supporting GIS applications and workflows.
5. Managing geographically. In government and many large corporations, GIS is becoming essential to understand what is going on. Senior administrators and executives at the highest levels of government use GIS information products to communicate. These products provide a visual framework for conceptualizing, understanding, and prescribing action. Examples include briefings about various geographic patterns and relationships including land use, crime, the environment, and defence/security situations. GIS is increasingly being implemented as enterprise information systems. This goes far beyond simply spatially enabling business tables in a DBMS. Geography is emerging as a new way to organize and manage organizations. Just like enterprise-wide financial systems transformed the way organizations were managed in the '60s, '70s, and '80s, GIS is transforming the way that organizations manage their assets, serve their customers/citizens, make decisions, and communicate. Examples in the private sector include most utilities, forestry and oil companies, and most commercial/retail businesses. Their assets and resources are now being maintained as an enterprise information system to support day-to-day work management tasks and provide a broader context for assets and resource management.

## 1.4. FUNCTIONAL COMPONENTS OF GIS




GIS in the wider sense consists of software, data, people, and an organisation in which it is used. We should note that organisation factors will define the context and rules for the capture, processing and sharing of geo-information, as well as the role which GIS plays in the organisation as a whole.

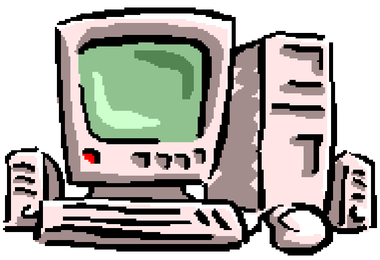
GIS consists of several functional components- components which support key GIS functions. These are data capture and preparation, data storage, data analysis, and presentation of spatial data. The system should not be called a geographic information system if any one of these components is missing.

### *The Components of GIS*

A Geographic Information System is a system of computer software, hardware and data, and the personnel that make it possible to enter, manipulate, analyze, and present information that is tied to a location on the earth's surface.

The components of a GIS fall into three main categories.

- Computer Hardware and Software
- Spatial Data from the "Real World"
- Trained Personnel



- **Hardware/Software:** Hardware is the computer on which a GIS operates. The software runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations. GIS software provides functions and tools needed to input and store geographic information. It also provides query tools, performs analysis, and displays geographic information in the form of maps or reports. All GIS software packages rely on an underlying database management system (DBMS) for storage and management of the geographic and attribute data. The GIS communicates with the DBMS to perform queries specified by the user.

- **Data:** Data is one of the most important, and often most expensive, components of a GIS. Geographic data, which is comprised of geographic features and their corresponding attribute information, is entered into a GIS using a technique called digitizing. This process involves digitally encoding geographic features, such as buildings, roads or county boundaries. Digitizing is done by tracing the location, path or boundary of geographic features either on a computer screen using a scanned map in the background, or a paper map that is attached to a digitizing tablet. The digitizing process can be very tedious and time consuming, especially when capturing large datasets such as soil polygons, streams or topographic contours. Fortunately, much of the data GIS users need has been created by government agencies or commercial operations, and is available for free or for purchase from the data provider or from a spatial data clearinghouse.

- **People:** The real power of a GIS comes from the people who use them. Over the past decade, computers have become much easier for people to use and more affordable for companies, schools and organizations to purchase. Given this fact, the number of GIS users has increased rapidly, and no longer includes only GIS specialists. Today GIS is being used by people, in many different fields, as a tool that enables them to perform their jobs more effectively. Police use GIS to solve crimes, Emergency 911 operators use GIS to send emergency personnel to a person in distress, biologists use GIS to protect plant and animal species, teachers use GIS to teach lessons in geography, history or engineering. The list of GIS users in the 21st century goes on and on. Whatever the application, the user is the key to a successful GIS.

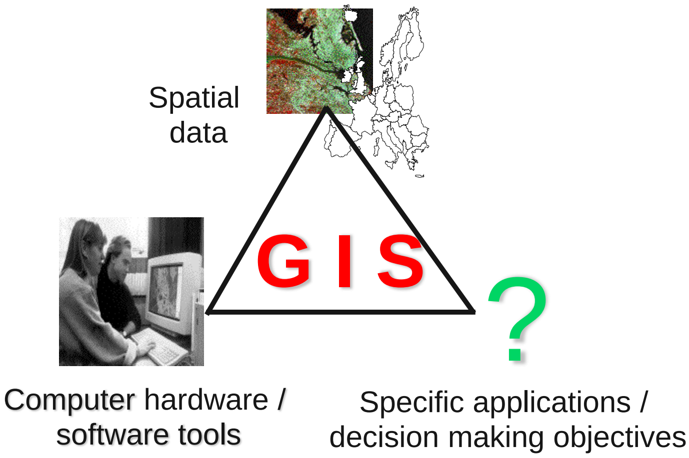



**Fig:** *Components of GIS*

## 1.5. IMPORTANCE OF GPS AND REMOTE SENSING DATA IN GIS

The widespread use of computers has led to the development of new technologies, collectively known as geographical information sciences (GISci), for mapping and monitoring features on the surface of the Earth. Foremost for exploration and fieldwork among these technologies are: geographical information systems (GIS), which can take digital datasets and produce maps showing features of interest in matter of seconds; the global positioning system (GPS), which allows positions to be determined to ±10 m anywhere on the Earth's surface; and methods of observing features from a distance, such as photography or infra-red scanning, known as remote sensing. These GISci techniques complement the surveys and sampling that are at the heart of scientific exploration: they greatly enhance the types of fieldwork that can be carried out, reduce the amount of time needed for many tasks and improve the quality of results. Geographical information sciences and expedition fieldwork. A fundamental objective of most exploration is to observe and record information about the part of the world being studied, for instance by field surveys, photography, or questionnaires. The development of ever-cheaper and more powerful computers, GIS software and GPS kit, along with low-cost satellite pictures of the Earth, has greatly improved the potential of expeditionary fieldwork to record, analyse and present data that may help us to improve conditions on this beleaguered planet. Remote sensing provides us with a means of recording the distribution of features on the surface of the Earth and changes in those features over time: it is often the only source of new data about a region that will be available to you, prior to you going there to collect field data. Your GPS will tell you where you are in your study region and allows you to input your sample locations into a GIS. A GIS is a means of combining existing data and new data from fieldwork or the interpretation of remotely sensed images. GIS-generated maps greatly reduce the original amounts of data and can be designed to focus on specific themes of interest to your research.

### *Remote sensing GPS Surveys & sampling*

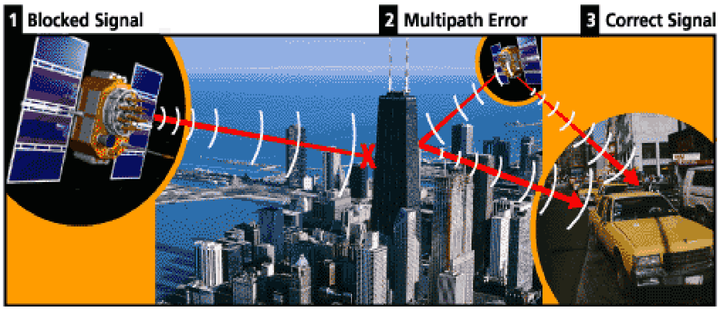
This manual aims to provide expeditions with details of fieldwork techniques, from 'traditional' compass-based surveying, through to the use of GIS to show GPS-located sites on satellite images displayed on a laptop screen in, say, Amazonia or the Himalayas. There are many ways in which geographical information sciences can help with fieldwork projects, these are just a few of the possible applications:

- Logistics: planning routes and navigation
- Research: mapping vegetation, wildlife, urbanisation, soils and geological features
- Monitoring: data logging of fire extents, forest loss, river channel changes
- Conservation applications: assessing biodiversity, park zonation, impact assessment
- Technology transfer: training local technical staff, donating hardware and shareware
- Education: maps for displays, involving schoolchildren with fieldwork.

### 1.5.1. GPS

- The Global Positioning System (GPS) is a satellite-based navigation system made up of a network of 24 satellites placed into orbit by the U.S. Department of Defense.
- GPS was originally intended for military applications, but in the 1980s, the government made the system available for civilian use.
- GPS works in any weather conditions, anywhere in the world, 24 hours a day.
- There are no subscription fees or setup charges to use GPS.
- Accuracy can be pinpointed to within one (1) meter with special military-approved equipment.

- GPS equipment is widely used in science and has now become sufficiently low-cost so that almost anyone can own a GPS receiver.




**GPS Technology**

- The 24 satellites that make up the GPS space segment are orbiting the earth about 12,000 miles above us.
- They are constantly moving, making two complete orbits in less than 24 hours.
- These satellites are travelling at speeds of roughly 7,000 miles an hour.
- GPS satellites are powered by solar energy & have backup batteries. Why?
- They have backup batteries onboard to keep them running in the event of a solar eclipse, when there's no solar power.
- Small rocket boosters on each satellite keep them flying in the correct path.

**GPS Facts**

- The first GPS satellite was launched in 1978.
- A full constellation of 24 satellites was achieved in 1994.
- Each satellite is built to last about 10 years.
- Replacements are constantly being built and launched into orbit.
- A GPS satellite weighs approximately 2,000 pounds and is about 17 feet across with the solar panels extended.
- Transmitter power is only 50 watts or less.

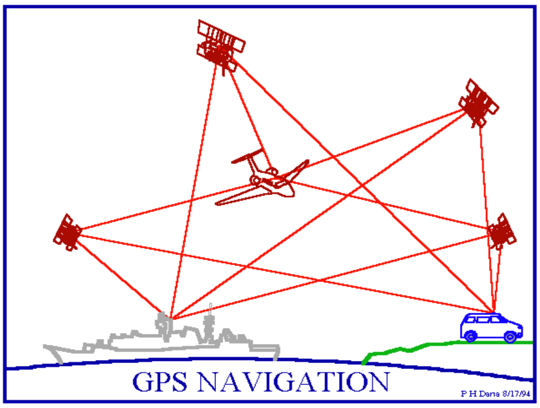



**Fig:** *GPS Navigation*

**How GPS works?**

☑ GPS satellites broadcast radio signals providing their location, status, and precise time {$t_1$} from on-board atomic clocks.

- ☑ The GPS radio signals travel through space at the speed of light {c}, more than 299,792 KM/Second.
- ☑ A GPS device receives the radio signals, noting their exact time of arrival {$t_2$}, and uses these to calculate its distance from each satellite in view.
- ☑ Once a GPS device knows its distance from at least 4 satellites, it can use geometry to determine its location on Earth in three dimensions.

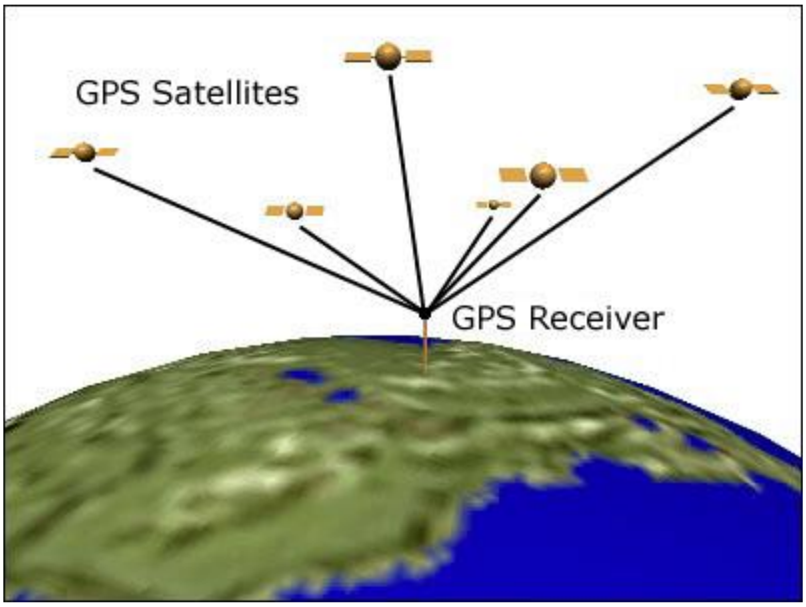
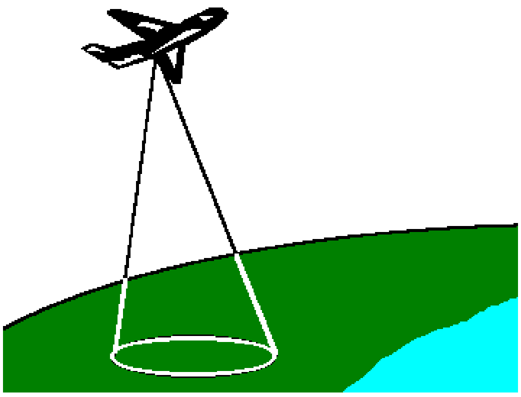



**Fig:** *Working of GPS*

### 1.5.2 Remote Sensing

- Remote Sensing is methods of observing features from a distance, such as photography or infra-red scanning
- Remote sensing technologies are used to gather information about the surface of the earth from a distant platform, usually a satellite or airborne sensor.
- Most remotely sensed data used for mapping and spatial analysis is collected as reflected electromagnetic radiation, which is processed into a digital image that can be overlaid with other spatial data.
- For example, chlorophyll strongly absorbs blue (0.48 mm) and red (0.68 mm) wavelength radiation and reflects near-infrared radiation (0.75 - 1.35 mm).
- Leaf vacuole water absorbs radiation in the infrared region from 1.35 - 2.5 mm (Samson, 2000).
- The spectral properties of vegetation in different parts of the spectrum can be interpreted to reveal information about the health and status of crops, rangelands, forests and other types of vegetation.



**Satellite Remote Sensing**

- In satellite remote sensing of the earth, the sensors are looking through a layer of atmosphere separating the sensors from the Earth's surface being observed.
- Hence, it is essential to understand the effects of atmosphere on the electromagnetic radiation travelling from the Earth to the sensor through the atmosphere.
- The atmospheric constituents cause wavelength dependent absorption and scattering of radiation.
- These effects degrade the quality of images. Some of the atmospheric effects can be corrected before the images are subjected to further analysis and interpretation.
- Remote sensing systems are often designed to operate within one or more of the atmospheric windows.
- These windows exist in the microwave region, some wavelength bands in the infrared, the entire visible region and part of the near ultraviolet regions.
- Although the atmosphere is practically transparent to x-rays and gamma rays, these radiations are not normally used in remote sensing of the earth

**Optical and Infrared Remote Sensing**

- In Optical Remote Sensing, optical sensors detect solar radiation reflected or scattered from the earth, forming images resembling photographs taken by a camera high up in space.
- The wavelength region usually extends from the visible and near infrared (commonly abbreviated as VNIR) to the short-wave infrared (SWIR).
- Different materials such as water, soil, vegetation, buildings and roads reflect visible and infrared light in different ways.
- They have different colours and brightness when seen under the sun.
- The interpretation of optical images require the knowledge of the spectral reflectance signatures of the various materials (natural or man-made) covering the surface of the earth.

**Remote Sensing Images**

- Remote sensing images are normally in the form of digital images.
- In order to extract useful information from the images, image processing techniques may be employed to enhance the image to help visual interpretation, and to correct or restore the image if the image has been subjected to geometric distortion, blurring or degradation by other factors.
- There are many image analysis techniques available and the methods used depend on the requirements of the specific problem concerned.
- In many cases, image segmentation and classification algorithms are used to delineate different areas in an image into thematic classes.
- The resulting product is a thematic map of the study area.
- This thematic map can be combined with other databases of the test area for further analysis and utilization.

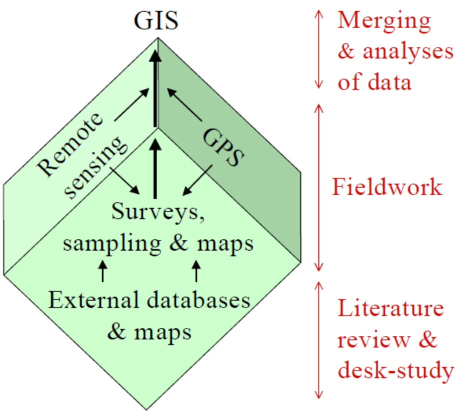
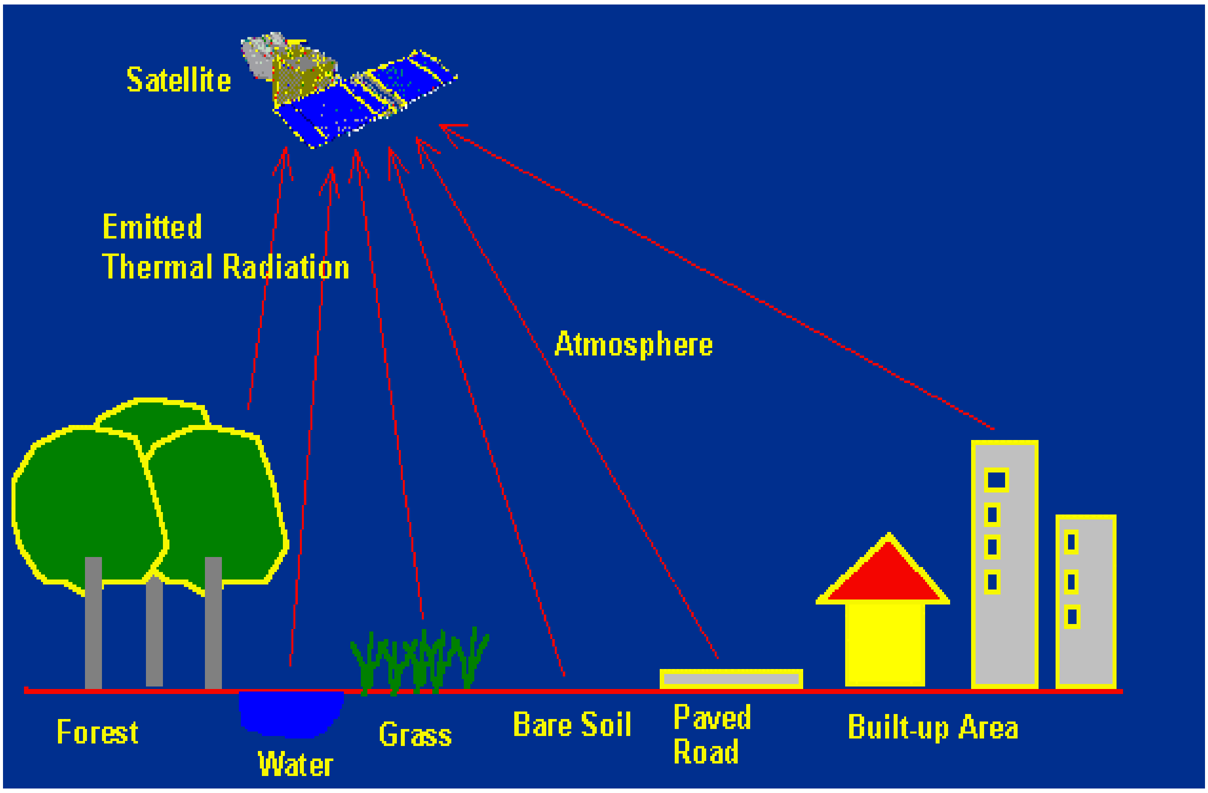



**Fig:** Data sources for expedition fieldwork. Arrows in the diamond indicate how data from one source can feed into another

**How Remote Sensing works?**



# 2. DIGITAL MAPPING CONCEPT

## 2.1. MAP CONCEPT: MAP ELEMENTS, MAP LAYERS, MAP SCALES AND REPRESENTATION

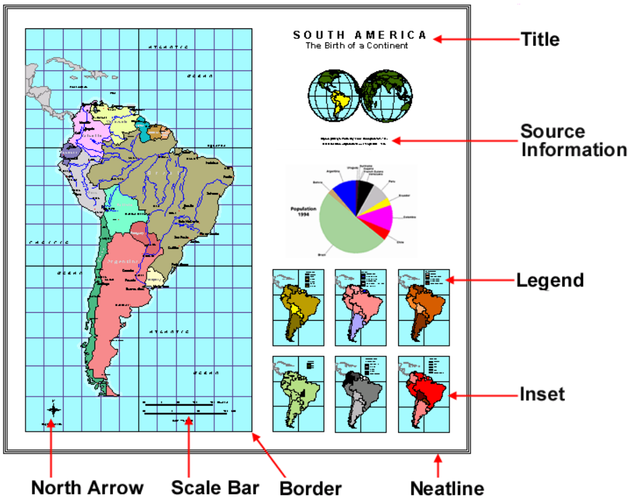
### 2.1.1. MAP elements

Maps are the primary tools by which spatial relationships are visualized. Maps therefore become important documents. There are several key elements that should be included each time a map is created in order to aid the viewer in understanding the communications of that map and to document the source of the geographic information used.

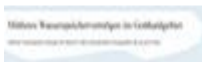
#### *1. Marginalia*

The elements of MAP which also called Marginalia refers to any supporting information or elements on the map.

- **Title:** It should be the largest text on the map but not over power the main body
- **Legend:** A figure to help the reader to interprate the map body.
- **North Arrow**
- **Scale Bar**
- **Borders and Neatlines**
- Source information and other text including projection information
- **Inset maps**




A map usually contains the following elements:

- **Title (and subtitle)**: Usually draws attention by virtue of its dominant size; serves to focus attention on the primary content of the map. Should be an answer to "What? Where? When?".

**Tips**: Never underline a title (or a subtitle), and never put a colon after a title.

- **Legend**: The principal reference to the map symbols; subordinated to the title. However, this is still a key element for map reading; describing all unknown or unique map symbols used. Tips: Only the word "Legend" should be written on your map (and not "Map Legend", or "Switzerland Legend", etc.).

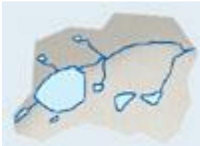

- **Map Scale:** Provides the reader with important information regarding linear relations on the map. A scale can be numerical (for example 1:50000) or graphical.


Tips:The dimension and thickness of a graphical scale has to be adapted to the map content.

- **Credits:** Can include the map source, the author, indication of the reliability of accuracy of the map, dates, or other explanatory material.

Tips: Credits should always be written smallest as possible (but nevertheless readable) and be placed in a box without a frame.

- **Mapped Areas:** Objects, land, water, and other geographical features important to the purpose of the map.



- **Map Symbols:** Wide variety of forms and functions; the most important element of the map, along with the geographic areas rendered.

- **Place name and Labelling:** The chief means of communicating with maps; serve to orient the reader on the map and provide important information regarding its purpose.

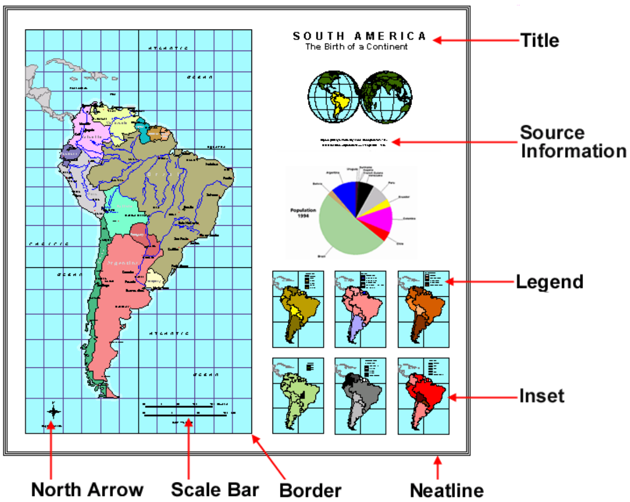

**Tips:**Use the same font for the map frame, the map layout, and the map content.

- **North arrow:** According to the rules, each map should have a north arrow. But if the map is north oriented, or if the geographical co-ordinate are already on the map the north arrow can be omitted.


Tips:The north arrow must be well readable, but not be too dominant on the map.

- **Border and Neatlines:** Both optional; borders can serve to restrain eye movements. Neatlines are finer lines than borders, drawn inside them and often intra-parallelism, rendered as part of the graticule; used mostly for decoration.

- **Graticule:** Often omitted in maps today; should be included if the location information is crucial




to the map purpose, e.g. into topographical maps.

## 2. Map Templates

- If you are making a series of maps set up the Marginalia and save the project as a template.
- You can use the template to give the series a consistent look and feel.

## 3. Map Design Objectives

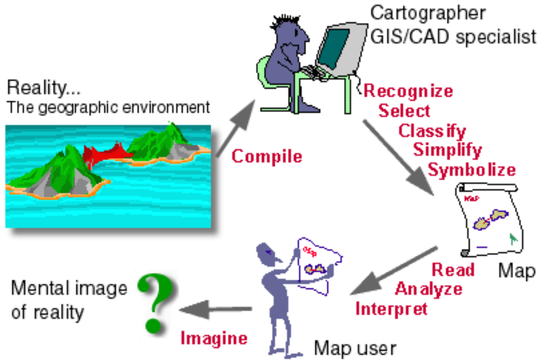
- Fulfill Intended Purpose
- Share Information
- Highlight Relationships
- Illustrate Results

## 4. Communicating with Maps

- Determine what the purpose of the map is
- Who is the intended audience
- What features are needed
- What is the best way to symbolize the map

## 5. Factors Controlling Cartographic Design

- Map Objective: Will the map be in a book, hang on a wall, be folded or flat, black and white or color, etc?
- Audience: Is the audience a group of scientist or the general public.
- Scale: Controls how much detail can be on the map.
- Technical Limits: Minimum line width, Limited color palette, lack of color
- Mode of use: Will the map be used in the field? Place on a wall? Used while driving?

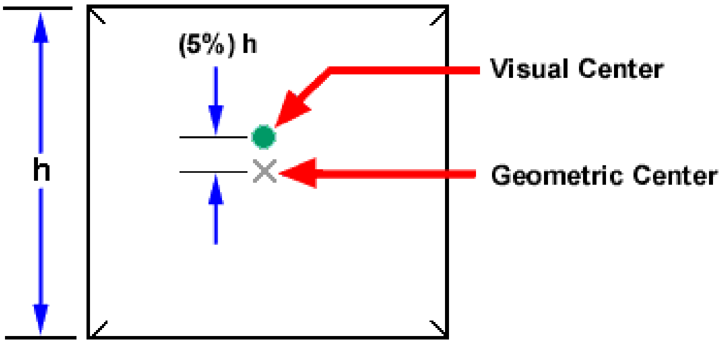



## 6. Standard Symbolization

- People are conditioned to recognize standard symbols.
- Roads Red
- Streams Blue
- Geologic Symbols

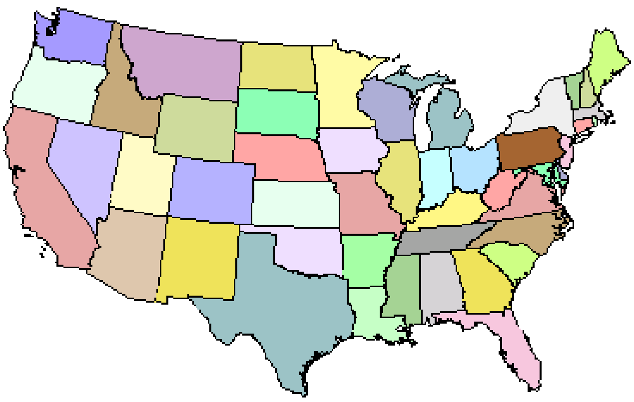
## 7. Visual Balance

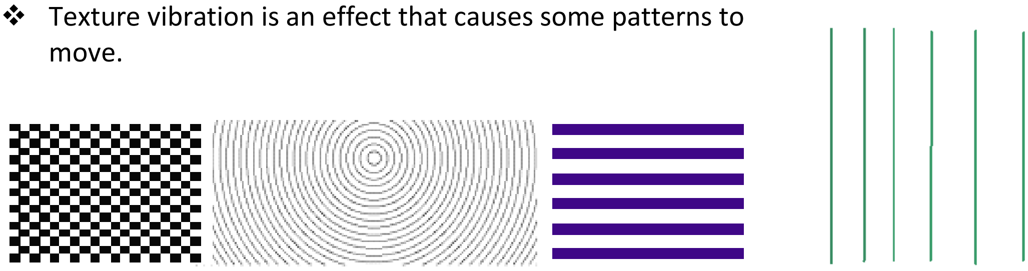
❖ Most people will focus on a point slightly above the image's geometric center




## 8. Graphic Perception

- The human eye has difficulty deciphering more than 12 colors in one view
- This map has 48 colors.
- Can you tell the difference between California and Nebraska
- The human eye can decipher no more than seven or eight shades form the 256 shades of one color
- All of the lines differ by 1.9%
- 5 to 7% of the population is color blind
- Your map may be reproduced in Black and white.
- Texture vibration is an effect that causes some patterns to move.





## *9. Annotation*

- Map labels are useful but lack control.
- When you need more control use Annotation
- It is possible to create feature linked annotation in 8.3 (ArcEditor, ArcInfo)
- New product at 9.0 called Maplex has rules based labeling

## *10. Geographic Features*

- The map body is the main focus for the map. It should be prominently displayed. The other elements of the map should not direct attention away form it.

## 2.2. MAP LAYERS

Each map layer is used to display and work with a specific GIS dataset. A layer represents geographic data, such as a particular theme of data. Example map layers include streams and lakes, terrain, roads, political boundaries, parcels, building footprints, utility lines, and orthophoto imagery.

A layer references the data stored in geodatabases, coverages, shapefiles, rasters, and so on, rather than actually storing the geographic data. Thus, a layer always reflects the most up-to-date information in your database.

Layers have a number of properties you can work with and set. You can right-click a layer in the table of contents and click Properties to view the Layer Properties dialog box, where you can set symbology, labeling, drawing rules, and other options. For example, you can specify that streams are drawn with all blue lines, parcels are drawn based on their land-use code, parks are drawn using a green pattern fill and are labeled with the park name, digital elevation is portrayed as a shaded relief, and so on. In addition, other properties include defining the scales at which they can draw, which features to draw from the data source, where that data is located in your database, attribute properties, joins, and relates for working with the tabular information.

Like the thematic layers in a map, GIS datasets represent logical collections of individual features with their geographic locations and shapes, as well as descriptive information about each feature stored as attributes.

Prior to GIS, mapmakers created a series of map layers that were used to geographically describe and characterize a location. They often used transparencies that could be overlaid on a light table. These integrated displays were used to visualize spatial relationships and gain insight about relevant characteristics of a place. Practitioners would use these to make interpretations and to draw interesting conclusions.

One visionary who used this process for planning was Dr. Ian McHarg, a landscape architect and renowned writer on regional planning using natural systems. His seminal book *Design with Nature* was published in 1969 and articulated concepts for ecological planning, which applied these map overlay principles. You can learn more at Wikipedia about Ian McHarg and his work.

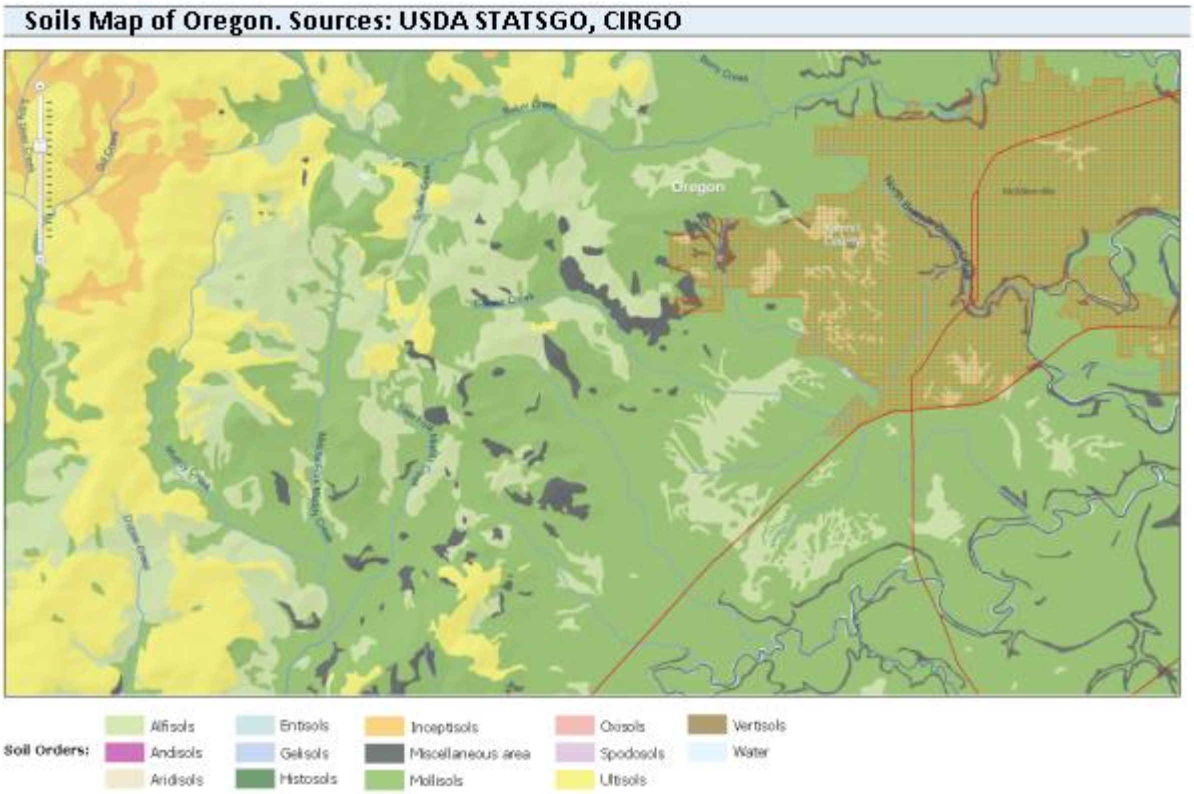
Around this same time, Dr. Roger Tomlinson, known as the father of GIS, developed his early ideas for GIS. Among other aspects of GIS, he further articulated the concept of thematic layers and overlays as a cornerstone for GIS.

These early GIS practitioners thought about how geographic information could be partitioned into a series of logical information layers—as more than a random collection of objects. They envisioned homogeneous collections of representations that could be managed as layers. These GIS users organized information in individual data themes that described the distribution of a phenomenon and how each theme should be portrayed across a geographic extent. They found that they could use relatively simple GIS data types (points, lines, polygons, and rasters). These simple data layers could be combined through location—that is, georeferencing enabled datasets to be combined in a map or overlaid using geoprocessing operations such as polygon overlay.

These pioneers also provided a protocol for data collection and how to manage these collections as geographic data layers. Here is one example for representing soils.

Each and every area (polygon) in a specified extent could be assigned a dominant soil type, and the soil types could be consistently classified and described using properties or attributes of each polygon. In the case of soils, very involved sets of properties are typically recorded for each soil polygon.

A theme could be defined to delineate various areas representing the dominant soil type (that is, a layer collection of soil type polygons and their descriptions as attribute values).

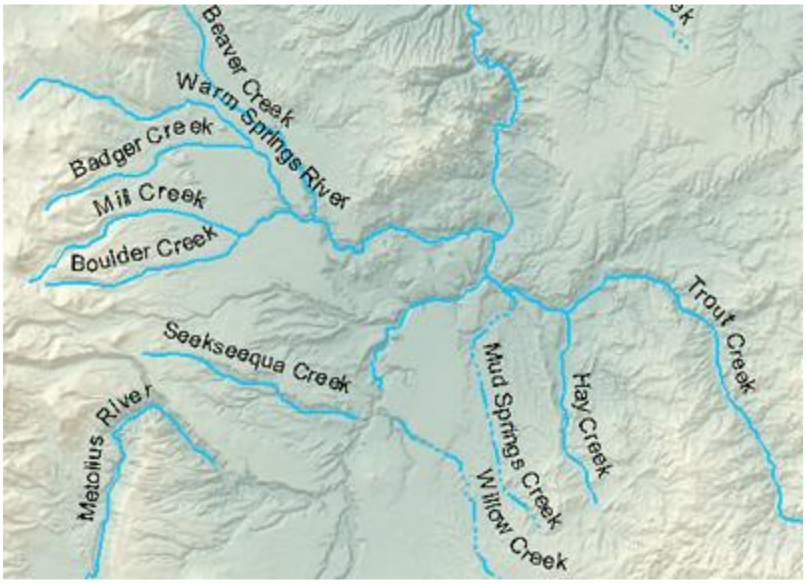



This organizing principle of geographic layers became one of the universal GIS principles that provided the foundation for how GIS systems represent, operate on, manage, and apply geographic information.

## SYMBOLIZING DATA IN A LAYER

- When you add a dataset to a data frame in your map, by default ArcMap draws all the features using the same symbol and color. If you open the properties of a layer and click the Symbology tab, you can change the way that information is drawn. Depending on the type of layer that is created, you will find different ways to symbolize the data contained within the layer. Most often, information contained within the dataset itself is used to determine how the data is symbolized.

  Sometimes all the features in a dataset are drawn using a common symbol or color. For example, all streams are drawn using blue lines and labeled in the same font and size.




However, chances are that you will often want some layer displays to be more sophisticated and will want to assign symbols based upon attribute values. For example, the circle symbols representing cities in the map below are classified into sizes based upon population.


Categorical data such as this is drawn using unique symbols for each category. You can work with continuous numeric data fields by using classification tools to create a series of class ranges. Then, you can assign symbols to draw each class or use a color ramp to assign color values to each class.

Here are examples of three frequently used methods for symbolizing layers.

## DIFFERENT TYPES OF LAYERS

There are different kinds of layers. Some layers represent a particular type of geographic feature, while others represent a particular type of data. Each layer type has different mechanisms for displaying and symbolizing their contents, and specific operations that you will perform against them. Many layers have special sets of tools for working with the layer and its contents. For example, you can use the Editor toolbar to manipulate feature layers, the Topology toolbar to work with the contents of a topology layer, and the Annotation toolbar to create and edit annotation (text) layers.

Here are a few of the common layer types:

- Feature layer: a layer that references a set of feature (vector) data that represents geographic entities as points, lines, and polygons. A feature layer's data source can be stored in geodatabases, shapefiles, ArcInfo Coverages, CAD files, and so on.
- Raster layer: a layer that references a raster or image as its data source.
- Graphics layer: a layer used to display graphics, text, and annotation—typically on top of other map layers.
- Service layer: a layer used to display ArcIMS, ArcGIS Server, WMS services, and other Web services.
- Geoprocessing layer: a layer that displays the output of a geoprocessing tool.

### 2.2.1. Map Scale and Representation

Naturally it is impossible for real world features to be drawn on the map as large as their true size. Therefore in order to represent the real world, maps are made to a specific scale. Map scale is defined as the ratio of the distance between two points on the map to the corresponding distance on the ground. Maps come in a variety of scales. Large scale maps cover a small area with great detail and accuracy, while small scale maps cover a large area in less detail.

map scale including bar, verbal and fractional scales As shown in this image, map scales can be expressed as a verbal statement, as a fraction or ratio and finally as a graphic or bar scale. Such scale expressions can be used to find the ground distance between any features from conversion of the corresponding map distance measurement.

Verbal Scale:

"1 centimetre on the map represents 500m on the ground" is a verbal scale. Clearly here a distance of 1cm on the map corresponds to 500m on the earth's surface. So if you plan a route with a total distance of 22cm on the map, that would imply that you'll be traveling (22cm x 500m) / 1cm = 11000m or 11km on the ground.

Representative Fraction (RF) - Fractional Scale - Ratio Scale:

1:50000 represents the map scale as a mathematical ratio or fraction, thus the name ratio scale or fractional scale. 1:50000 can be shown as 1/50000 as well. Here such a scale means that one unit of measurment on the map is equal to 50000 of the same unit on the ground. Such a unit can be anything such as centimetre, meter, feet, inches, your finger length, half a lenght of a pencil, etc. Also we can say that any distance on the map is 1/50000 of its true value on the ground. Therefore 1cm on the map is equal to 50000cm on the ground, that is 1cm on the map is equal to (50000cm x 1m) /100cm = 500m or 0.5km on the ground. Again a 22cm route on the map can be calculated to be equal 22 x 50000cm = 1100000cm on the ground or (1100000cm x 1m) / 100cm = 11000m.

## 2.3. MAP PROJECTION: COORDINATES SYSTEM AND PROJECTION SYSTEM

### 2.3.1. Map projection

#### *1. Brief about Map Projections*

The shape of the Earth is represented as a sphere. It is also modeled more accurately as an oblate spheroid or an ellipsoid. A globe is a scaled down model of the Earth. Although they can represent size, shape, distance and directions of the Earth features with reasonable accuracy, globes are not practical or suitable for many applications. They are hard to transport and store. Globes are not suitable for use at large scales, such as finding directions in a city or following a hiking route, where a more detailed image is essential. They are expensive to produce, especially in varying sizes (scales). On a curved surface, measuring terrain properties is difficult, and it is not possible to see large portions of the Earth at once.

Maps do not suffer from the above shortcomings and are more practical than globes in most applications. Historically cartographers have tried to address the challenge of representing the curved surface of the Earth on a map plane, and to this end have devised map projections. A map projection is the transformation of Earth's curved surface (or a portion of) onto a two-dimensional flat surface by means of mathematical equations. During such transformation, the angular geographic coordinates (latitude, longitude) referencing positions on the surface of the Earth are converted to Cartesian coordinates (x, y) representing position of points on a flat map.

For simplicity most of this article assumes that the surface to be mapped is that of a sphere. In reality, the Earth and other large celestial bodies are generally better modeled as oblate spheroids, whereas small objects such as asteroids often have irregular shapes. These other surfaces can be mapped as well. Therefore, more generally, a map projection is any method of "flattening" into a plane a continuous curved surface.

## 2. MAP projections

A map projection is one of many methods used to represent the 3-dimensional surface of the earth or other round body on a 2-dimensional plane in cartography (mapmaking). Map projection is a systematic transformation of the latitudes and longitudes of locations on the surface of a sphere or an ellipsoid into locations on a plane. Any mathematical function transforming coordinates from the curved surface to the plane is a projection. Map projections are necessary for creating maps.

All map projections distort the surface in some fashion. Depending on the purpose of the map, some distortions are acceptable and others are not; therefore different map projections exist in order to preserve some properties of the sphere-like body at the expense of other properties. There is no limit to the number of possible map projections.

More generally, the surfaces of planetary bodies can be mapped even if they are too irregular to be modeled well with a sphere or ellipsoid. Even more generally, projections are the subject of several pure mathematical fields, including differential geometry and projective geometry.

The creation of a map projection involves three steps:

1. Selection of a model for the shape of the Earth or planetary body (usually choosing between a sphere, ellipsoid, oblate ellipsoid or geoid). Because the Earth's actual shape is irregular, information is lost in this step.
2. Transformation of geographic coordinates (longitude and latitude) to Cartesian (x, y) or polar plane coordinates. Cartesian coordinates normally have a simple relation to eastings and northings defined on a grid superimposed on the projection.
3. Reduce the scale (in manual cartography this step came second, in digital cartography it comes last)

Maps can be more useful than globes in many situations: they are more compact and easier to store; they readily accommodate an enormous range of scales; they are viewed easily on computer displays; they can facilitate measuring properties of the terrain being mapped; they can show larger portions of the Earth's surface at once; and they are cheaper to produce and transport. These useful traits of maps motivate the development of map projections.

However, Carl Friedrich Gauss's Theorema Egregium proved that a sphere's surface cannot be represented on a plane without distortion. The same applies to other reference surfaces used as models for the Earth. Since any map projection is a representation of one of those surfaces on a plane, all map projections distort. Every distinct map projection distorts in a distinct way. The study of map projections is the characterization of these distortions.

### 3. Selection of a model for the shape of the Earth

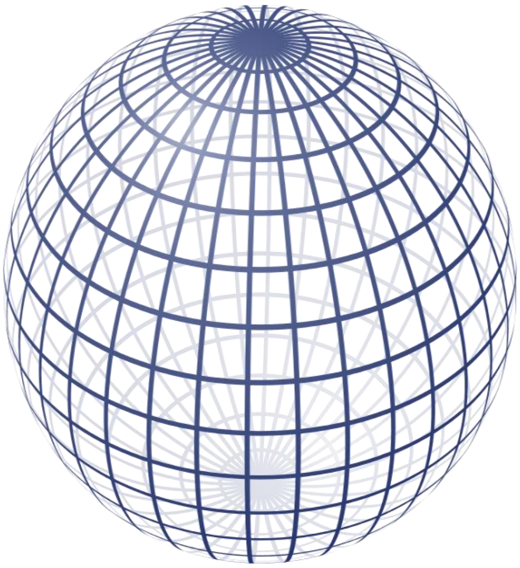
Projection construction is also affected by how the shape of the Earth or planetary body is approximated. The earth is taken as a sphere in order to simplify but the Earth's actual shape is closer to an oblate ellipsoid.

Selecting a model for a shape of the Earth involves choosing between the advantages and disadvantages of a sphere versus an ellipsoid. Spherical models are useful for small-scale maps such as world atlases and globes, since the error at that scale is not usually noticeable or important enough to justify using the more complicated ellipsoid. The ellipsoidal model is commonly used to construct topographic maps and for other large- and medium-scale maps that need to accurately depict the land surface.

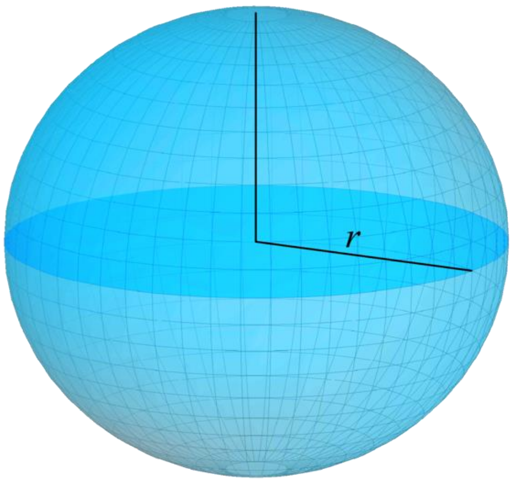
A third model is the geoid, a more complex and accurate representation of Earth's shape coincident with what mean sea level would be if there were no winds, tides, or land. Compared to the best fitting ellipsoid, a geoidal model would change the characterization of important properties such as distance, conformality and equivalence. Therefore in geoidal projections that preserve such properties, the mapped graticule would deviate from a mapped ellipsoid's graticule. Normally the geoid is not used as an Earth model for projections, however, because Earth's shape is very regular, with the undulation of the geoid amounting to less than 100 m from the ellipsoidal model out of the 6.3 million m Earth radius. For irregular planetary bodies such as asteroids, however, sometimes models analogous to the geoid are used to project maps from.

**Sphere**

A sphere is a perfectly round geometrical and circular object in three-dimensional space that resembles the shape of a completely round ball. Like a circle, which, in geometric contexts, is in two dimensions, a sphere is defined mathematically as the set of points that are all the same distance r from a given point in three-dimensional space. This distance r is the radius of the sphere, and the given point is the center of the sphere. The maximum straight distance through the sphere passes through the center and is thus twice the radius; it is the diameter.



A two-dimensional perspective projection of a sphere




r – radius of the sphere

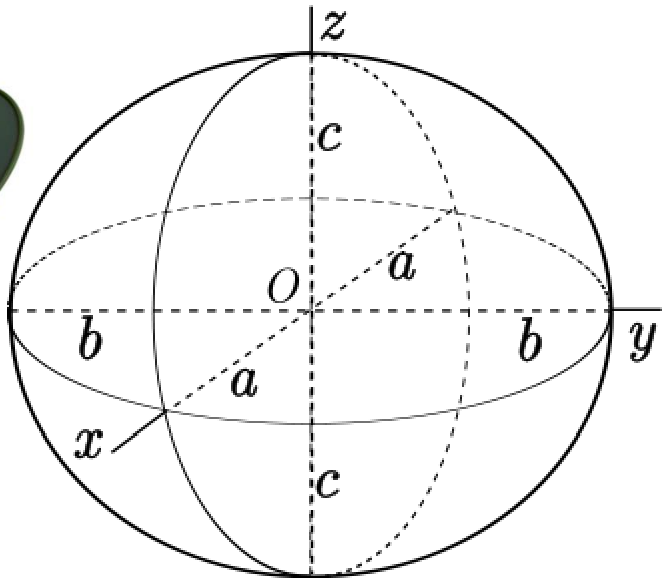
**Ellipsoid**

An ellipsoid is a closed quadric surface that is a three-dimensional analogue of an ellipse. The points (a,0,0), (0,b,0) and (0,0,c) lie on the surface and the line segments from the origin to these points are called the semi-principal axes of length a, b, c. They correspond to the semi-major axis and semi-minor axis of the appropriate ellipses.

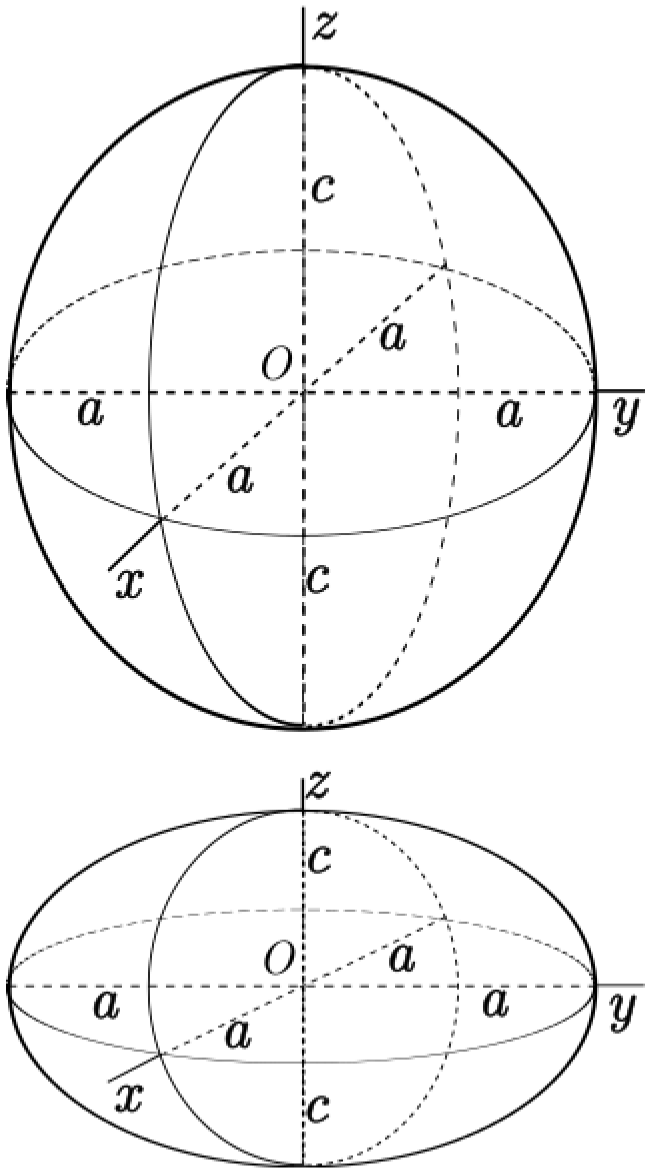
There are four distinct cases of which one is degenerate:

- a>b>c = tri-axial or (rarely) scalene ellipsoid;
- a=b>c = oblate ellipsoid of revolution (oblate spheroid);
- a=b<c = prolate ellipsoid of revolution (prolate spheroid);
- a=b=c = the degenerate case of a sphere;


**Tri-axial ellipsoid with distinct semi-axis lengths**




**Tri-axial ellipsoid with distinct semi-axes a, b and c**



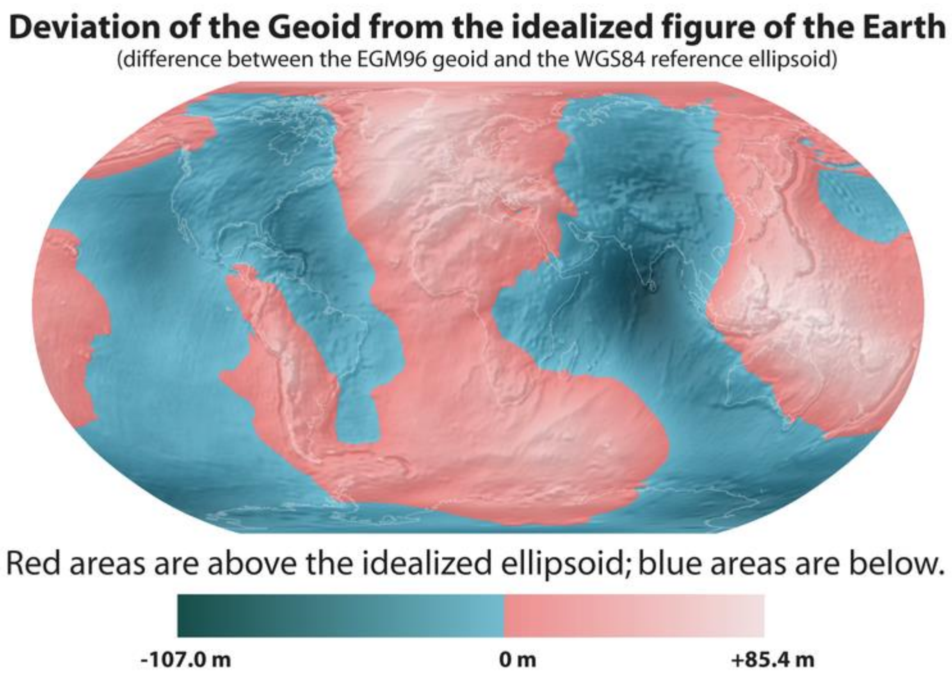

**The ellipsoid is prolate (top) or oblate (bottom) as c is greater than or less than a.**

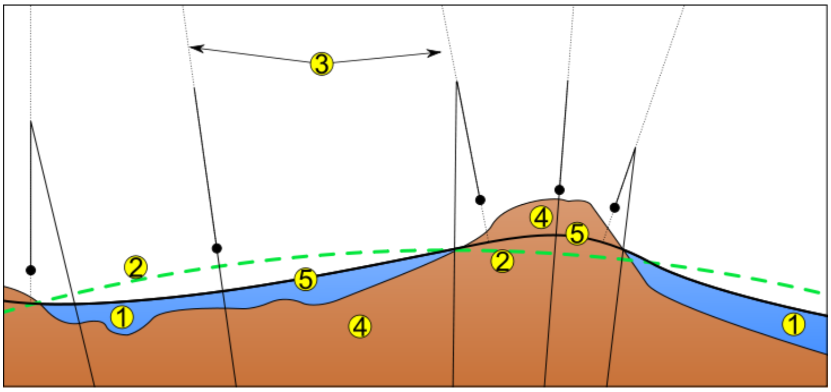
**Geoid**

The geoid is the shape that the surface of the oceans would take under the influence of Earth's gravitation and rotation alone, in the absence of other influences such as winds and tides. This surface is extended through the continents (such as with very narrow hypothetical canals). All points on the geoid have the same gravitational potential. The force of gravity acts everywhere perpendicular to the geoid, meaning that plumb lines point perpendicular and water levels parallel to the geoid.

Specifically, the geoid is the equipotential surface that would coincide with the mean ocean surface of the Earth if the oceans and atmosphere were in equilibrium, at rest relative to the rotating Earth, and extended through the continents (such as with very narrow canals). According to Gauss, who first described it, it is the "mathematical figure of the Earth", a smooth but highly irregular surface whose shape results from the uneven distribution of mass within and on the surface of the Earth. It does not correspond to the actual surface of the Earth's crust, but to a surface which can only be known through extensive gravitational measurements and calculations. Despite being an important concept for almost two hundred years in the history of geodesy and geophysics, it has only been defined to high precision in recent decades. It is often described as the true physical figure of the Earth, in contrast to the idealized geometrical figure of a reference ellipsoid.

The surface of the geoid is higher than the reference ellipsoid wherever there is a positive gravity anomaly (mass excess) and lower than the reference ellipsoid wherever there is a negative gravity anomaly (mass deficit).






1. Ocean
2. Reference ellipsoid
3. Local plumb line
4. Continent
5. Geoid

### 4. Metric properties of maps

Many properties can be measured on the Earth's surface independently of its geography. Some of these properties are:

Area

Shape

Direction

Bearing

Distance

Scale

Map projections can be constructed to preserve at least one of these properties, though only in a limited way for most. Each projection preserves or compromises or approximates basic metric properties in different ways. The purpose of the map determines which projection should form the base for the map. Because many purposes exist for maps, many projections have been created to suit those purposes.

Another consideration in the configuration of a projection is its compatibility with data sets to be used on the map. Data sets are geographic information; their collection depends on the chosen datum (model) of the Earth. Different datums assign slightly different coordinates to the same location, so in large scale maps, such as those from national mapping systems, it is important to match the datum to the projection. The slight differences in coordinate assignation between different datums is not a concern for world maps or other vast territories, where such differences get shrunk to imperceptibility.

### 5. Using globes vs. projecting on a plane

The globe is the only way to represent the earth without distorting one or more of the above-mentioned metric properties. Globes have the advantage of being true to metric properties and able to provide a true picture of spatial relationships on the earth's surface. The disadvantages of the globe are that it is impractical to make large-scale maps with it, it is difficult to measure on a globe, one can't see the whole world at once and it is difficult to handle and transport a globe around (unlike a folding map).

The flat map has the disadvantage of always distorting one or more of the metric properties and it is more difficult to get a true picture of the spatial relationships between objects. Flat maps have numerous advantages however; it is not practical to make large or even medium scale globes, it is easier to measure on a flat map, easy to carry around, and one can see the whole world at once.

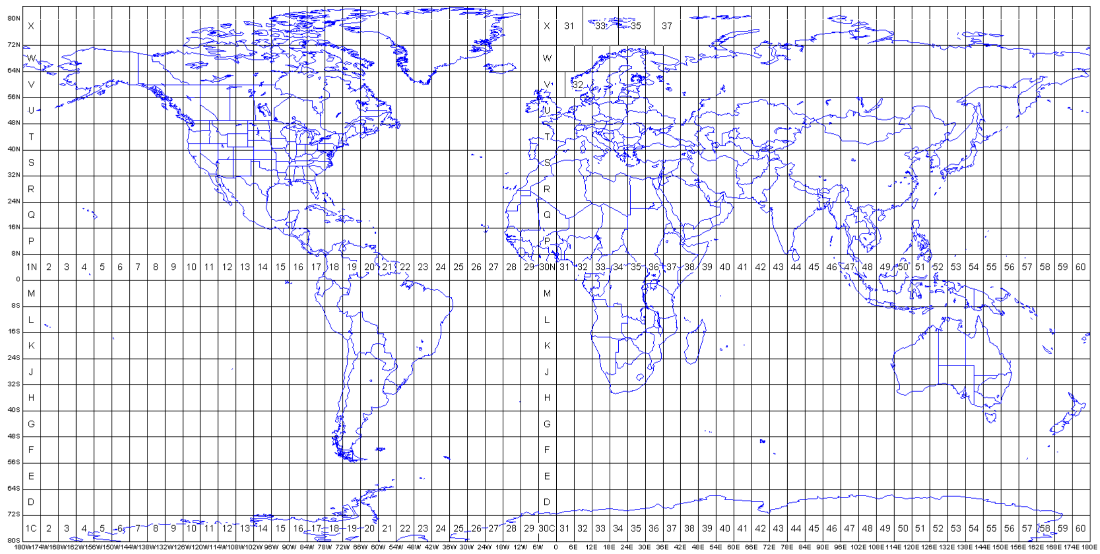
Scale in particular is effected by the choice between using a globe vs. a plane. Only a globe can have a constant scale throughout the entire map surface and the scale for flat maps will vary from point to point and may also vary in different directions from a single point (as in Azimuthal maps). The scale for a flat map can only be true along one or two lines or points (tangent or secant points/lines). The 'scale factor is therefore used to measure the difference between the idealized scale and the actual scale at a particular point on the map.

### *6. Choosing a projection surface*

A surface that can be unfolded or unrolled into a plane or sheet without stretching, tearing or shrinking is called a developable surface. The cylinder, cone and the plane are all **developable surfaces**. The sphere and ellipsoid do not have developable surfaces, so any projection of them onto a plane will have to distort the image. (To compare, one cannot flatten an orange peel without tearing and warping it.)

One way of describing a projection is first to project from the Earth's surface to a developable surface such as a **cylinder** or **cone**, and then to unroll the surface into a **plane**. While the first step inevitably distorts some properties of the globe, the developable surface can then be unfolded without further distortion.

Locations and grid designations in developable surfaces:



### 7. Aspects of the projection

Once a choice is made between projecting onto a cylinder, cone, or plane, the aspect of the shape must be specified. The aspect describes how the developable surface is placed relative to the globe: it may be normal (such that the surface's axis of symmetry coincides with the Earth's axis), transverse (at right angles to the Earth's axis) or oblique (any angle in between). The developable surface may also be either tangent or secant to the sphere or ellipsoid. Tangent means the surface touches but does not slice through the globe; secant means the surface does slice through the globe. Moving the developable surface away from contact with the globe never preserves or optimizes metric properties.

### 8. Scale

A globe is the only way to represent the earth with constant scale throughout the entire map in all directions. A map cannot achieve that property for any area, no matter how small. It can, however, achieve constant scale along specific lines.

Some possible properties are:

- The scale depends on location, but not on direction. This is equivalent to preservation of angles, the defining characteristic of a conformal map.
- Scale is constant along any parallel in the direction of the parallel. This applies for any cylindrical or pseudocylindrical projection in normal aspect.
- Combination of the above: the scale depends on latitude only, not on longitude or direction. This applies for the Mercator projection in normal aspect.
- Scale is constant along all straight lines radiating from a particular geographic location. This is the defining characteristic of an equidistant projection such as the Azimuthal equidistant projection. There are also projections (Maurer, Close) where true distances from two points are preserved.

**Measuring map scale distortion – scale factor & principal (nominal) scale**

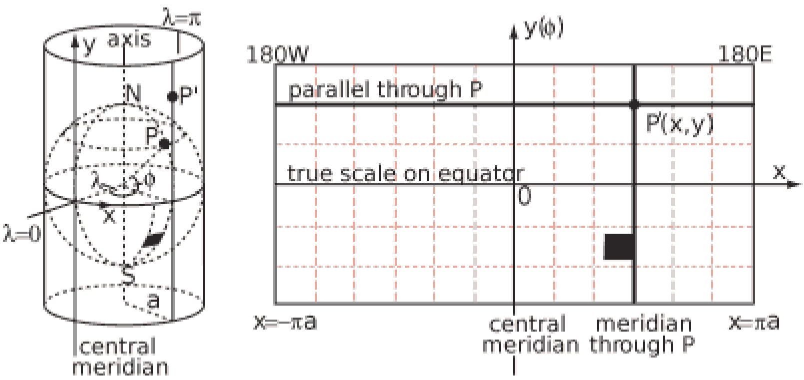
As mentioned above, a reference globe (reference surface of the Earth) is a scaled down model of the Earth. This scale can be measured as the ratio of distance on the globe to the corresponding distance on the Earth. Throughout the globe this scale is constant. For example, a 1:250000 representative fraction scale indicates that 1 unit (e.g. km) on the globe represents 250000 units on Earth. The **principal scale** or **nominal scale** of a flat map (the stated map scale) refers to this scale of its generating globe.

However the projection of the curved surface on the plane and the resulting distortions from the deformation of the surface will result in variation of scale throughout a flat map. In other words the actual map scale is different for different locations on the map plane and it is impossible to have a constant scale throughout the map. This variation of scale can be visualized by *Tissot's indicatrix* explained in detail below. Measure of scale distortion on map plane can also be quantified by the use of **scale factor**.

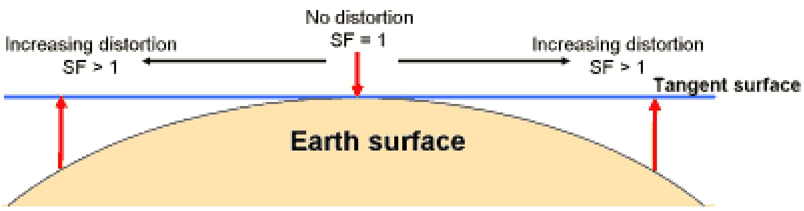
*Scale factor* is the ratio of actual scale at a location on map to the principal (nominal) map scale (SF = actual scale / nominal scale). This can be alternatively stated as ratio of distance on the map to the corresponding distance on the reference globe. A scale factor of 1 indicates actual scale is equal to nominal scale, or no scale distortion at that point on the map. Scale factors of less than or greater than one are indicative of scale distortion. The actual scale at a point on map can be obtained by multiplying the nominal map scale by the scale factor.

As an example, the actual scale at a given point on map with scale factor of 0.99860 at the point and nominal map scale of 1:50000 is equal to (1:50000 x 0.99860) = (0.99860 / 50000) = 1:50070 (which is a smaller scale than the nominal map scale). Scale factor of 2 indicates that the actual map scale is twice the nominal scale; if the nominal scale is 1:4million, then the map scale at the point would be (1:4million x 2) = 1:2million. A scale factor of 0.99950 at a given location on the map indicates that 999.5 meters on the map represents 1000 meters on the reference globe.

As mentioned above, there is no distortion along **standard lines** as evident in following figures. On a tangent surface to the reference globe, there is no scale distortion at the point (or along the line) of tangency and therefore scale factor is 1. Distortion increases with distance from the point (or line) of tangency.

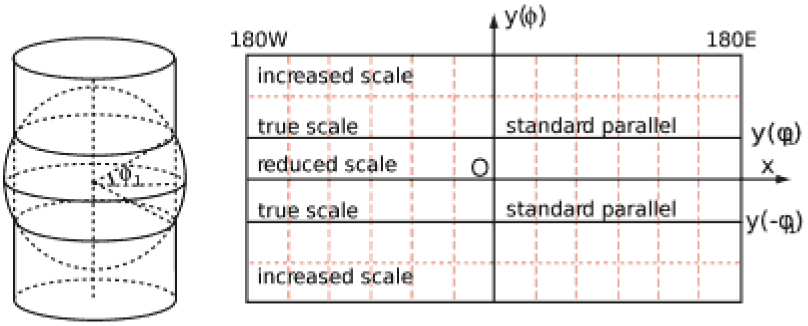



Map scale distortion of a tangent cylindrical projection - SF = 1 along line of tangency

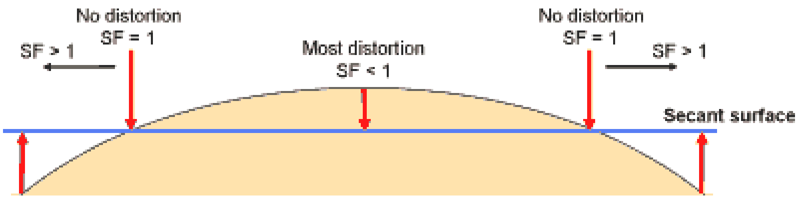



Scale distortion on a tangent surface to the globe

On a secant surface to the reference globe, there is no distortion along the standard lines (lines of intersection) where SF = 1. Between the secant lines where the surface is inside the globe, features appear smaller than in reality and scale factor is less than 1. At places on map where the surface is outside the globe, features appear larger than in reality and scale factor is greater than 1. A map derived from a secant projection surface has less overall distortion than a map from a tangent surface.




Map scale distortion of a secant cylindrical projection - SF = 1 along secant lines




Scale distortion on a secant surface to the globe

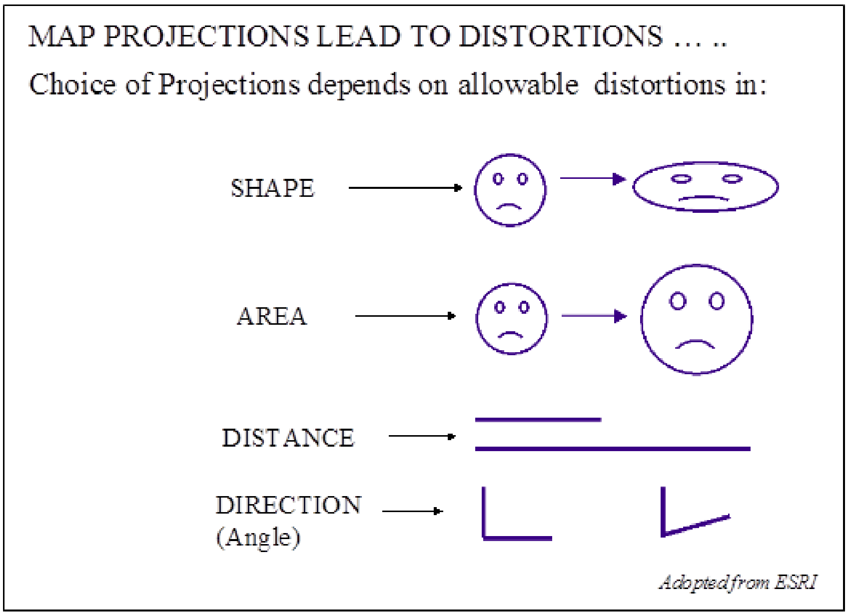
## 9. *Choosing a model for the shape of the body*

Projection construction is also affected by how the shape of the Earth or planetary body is approximated. In the following section on projection categories, the earth is taken as a sphere in order to simplify the discussion. However, the Earth's actual shape is closer to an oblate ellipsoid. Whether spherical or ellipsoidal, the principles discussed hold without loss of generality.

Selecting a model for a shape of the Earth involves choosing between the advantages and disadvantages of a sphere versus an ellipsoid. Spherical models are useful for small-scale maps such as world atlases and globes, since the error at that scale is not usually noticeable or important enough to justify using the more complicated ellipsoid. The ellipsoidal model is commonly used to construct topographic maps and for other large- and medium-scale maps that need to accurately depict the land surface. Auxiliary latitudes are often employed in projecting the ellipsoid.

A third model is the geoid, a more complex and accurate representation of Earth's shape coincident with what mean sea level would be if there were no winds, tides, or land. Compared to the best fitting ellipsoid, a geoidal model would change the characterization of important properties such as distance, conformality and equivalence. Therefore in geoidal projections that preserve such properties, the mapped graticule would deviate from a mapped ellipsoid's graticule. Normally the geoid is not used as an Earth model for projections, however, because Earth's shape is very regular, with the undulation of the geoid amounting to less than 100 m from the ellipsoidal model out of the 6.3 million m Earth radius. For irregular planetary bodies such as asteroids, however, sometimes models analogous to the geoid are used to project maps from.

## *10. Map projections lead to distortations*




### *11. Projection Classification*

A fundamental projection classification is based on the type of projection surface onto which the globe is conceptually projected. The projections are described in terms of placing a gigantic surface in contact with the earth, followed by an implied scaling operation. These surfaces are cylindrical (e.g. Mercator), conic (e.g., Albers), or azimuthal or plane (e.g. stereographic). Many mathematical projections, however, do not neatly fit into any of these three conceptual projection methods. Hence other peer categories have been described in the literature, such as pseudoconic, pseudocylindrical, pseudoazimuthal, retroazimuthal, and polyconic.

Another way to classify projections is according to properties of the model they preserve. Some of the more common categories are:

- Preserving **direction** (azimuthal or zenithal), a trait possible only from one or two points to every other point
- Preserving **shape** locally (conformal or orthomorphic)
- Preserving **area** (equal-area or equiareal or equivalent or authalic)
- Preserving **distance** (equidistant), a trait possible only between one or two points and every other point
- Preserving **shortest route**, a trait preserved only by the gnomonic projection

Because the sphere is not a developable surface, it is impossible to construct a map projection that is both equal-area and conformal.

### *12. Projections by surface*

The three developable surfaces (plane, cylinder, and cone) provide useful models for understanding, describing, and developing map projections. However, these models are limited in two fundamental ways. For one thing, most world projections in actual use do not fall into any of those categories. For another thing, even most projections that do fall into those categories are not naturally attainable through physical projection. As L.P. Lee notes,

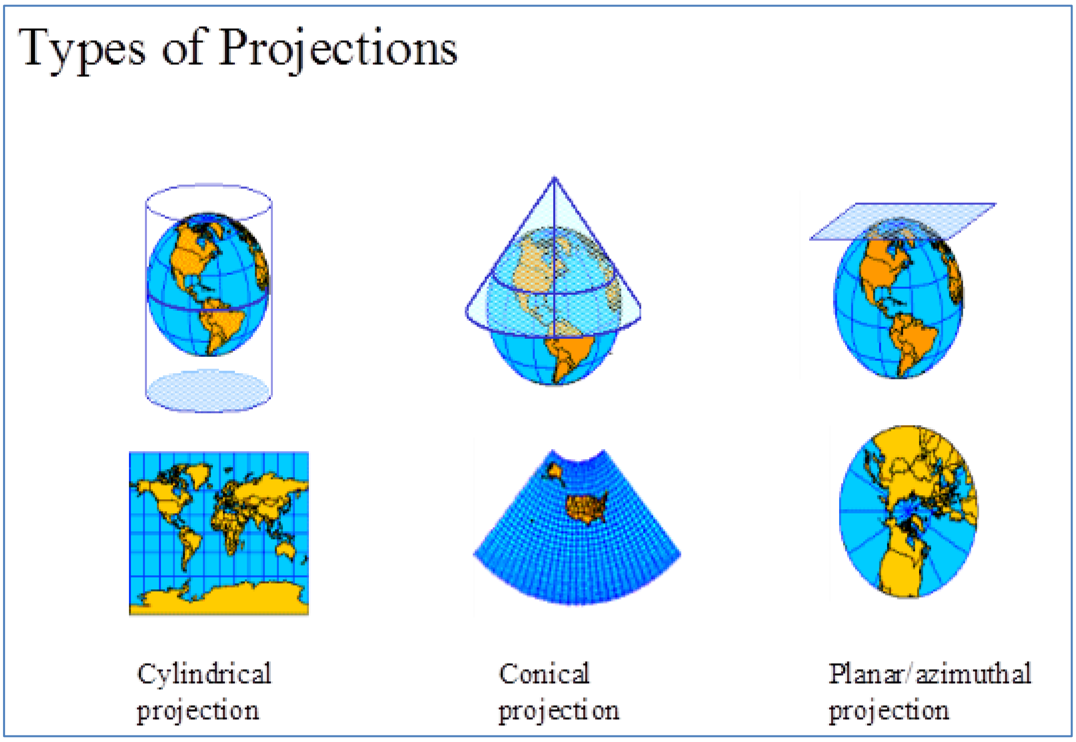
No reference has been made in the above definitions to cylinders, cones or planes. The projections are termed cylindric or conic because they can be regarded as developed on a cylinder or a cone, as the case may be, but it is as well to dispense with picturing cylinders and cones, since they have given rise to much misunderstanding. Particularly is this so with regard to the conic projections with two standard parallels: they may be regarded as developed on cones, but they are cones which bear no simple relationship to the sphere. In reality, cylinders and cones provide us with convenient descriptive terms, but little else.

Lee's objection refers to the way the terms cylindrical, conic, and planar (azimuthal) have been abstracted in the field of map projections. If maps were projected as in light shining through a globe onto a developable surface, then the spacing of parallels would follow a very limited set of possibilities. Such a cylindrical projection (for example) is one which:

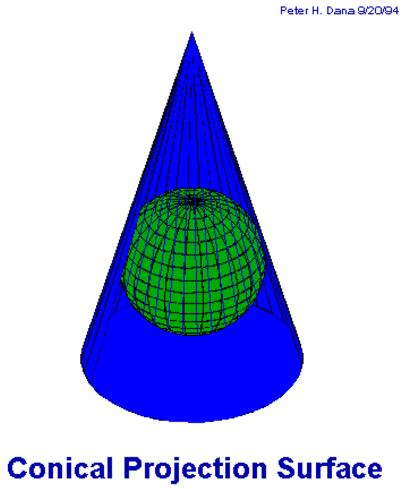
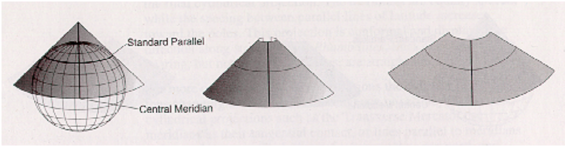
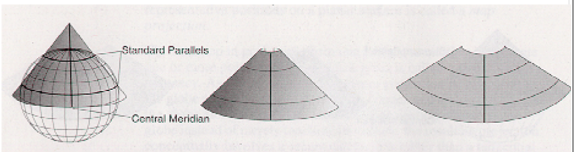
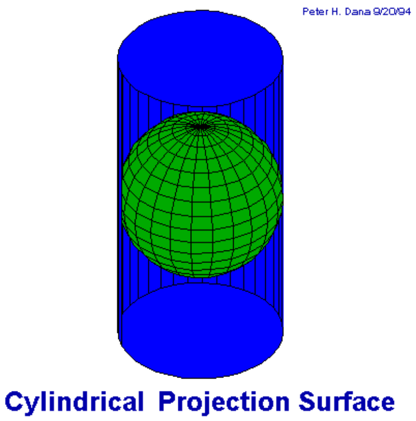
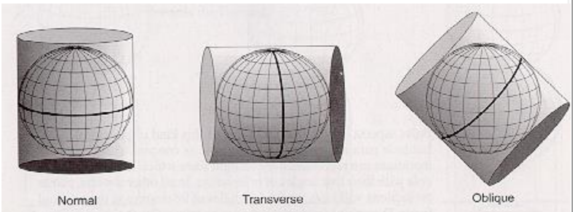
1. Is rectangular;
2. Has straight vertical meridians, spaced evenly;
3. Has straight parallels symmetrically placed about the equator;
4. Has parallels constrained to where they fall when light shines through the globe onto the cylinder, with the light source someplace along the line formed by the intersection of the prime meridian with the equator, and the center of the sphere.

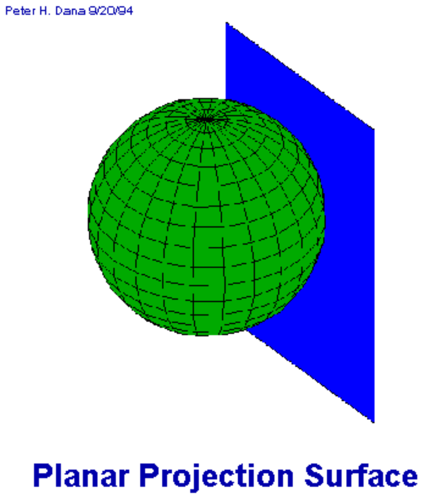
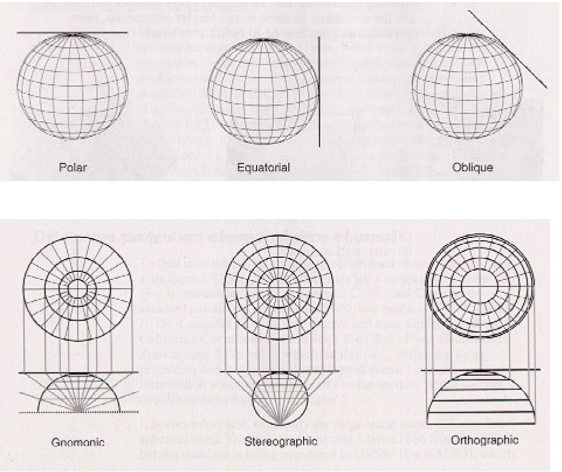
(If you rotate the globe before projecting then the parallels and meridians will not necessarily still be straight lines. Rotations are normally ignored for the purpose of classification.)

Where the light source emanates along the line described in this last constraint is what yields the differences between the various "natural" cylindrical projections. But the term cylindrical as used in the field of map projections relaxes the last constraint entirely. Instead the parallels can be placed according to any algorithm the designer has decided suits the needs of the map. The famous Mercator projection is one in which the placement of parallels does not arise by "projection"; instead parallels are placed how they need to be in order to satisfy the property that a course of constant bearing is always plotted as a straight line.



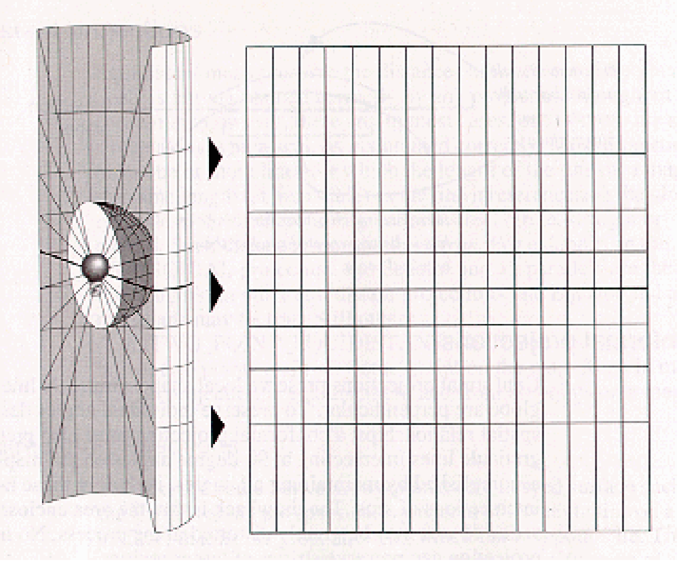
The projection surfaces (i.e., cylinders, cones, and planes) form the basic types of projections:

| | | |
|---|---|---|
| Conic |  | The conic tangent case:<br><br>The conic secant case:<br><br>Standard parallels are where the cone touches or slices through the globe. The central meridian is opposite the edge where the cone is sliced open.<br><br>Conic projections are used frequently for mapping large areas (e.g., states, large countries, or continents). |
| Cylindrical |  | Different cylindrical projection orientations:<br><br>The most common cylindrical projection is the Mercator projection, which is the basis of the UTM (Universal Transverse Mercator) system. |

| | | |
|---|---|---|
| Planar (Orthographic) |  | Different orthographic projection parameters:  |

## 1. Cylindrical

The term "normal cylindrical projection" is used to refer to any projection in which meridians are mapped to equally spaced vertical lines and circles of latitude (parallels) are mapped to horizontal lines.



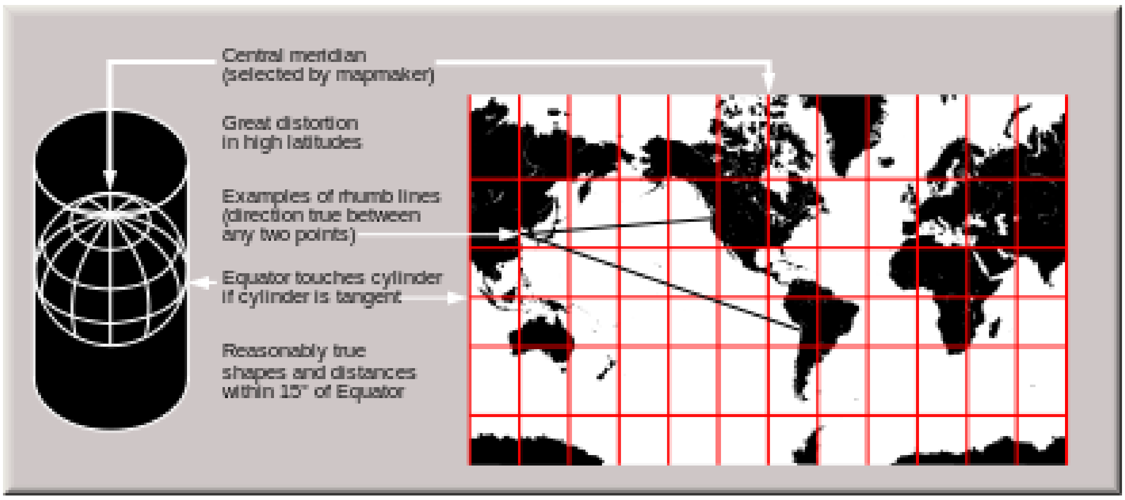
The mapping of meridians to vertical lines can be visualized by imagining a cylinder whose axis coincides with the Earth's axis of rotation. This cylinder is wrapped around the Earth, projected onto, and then unrolled.

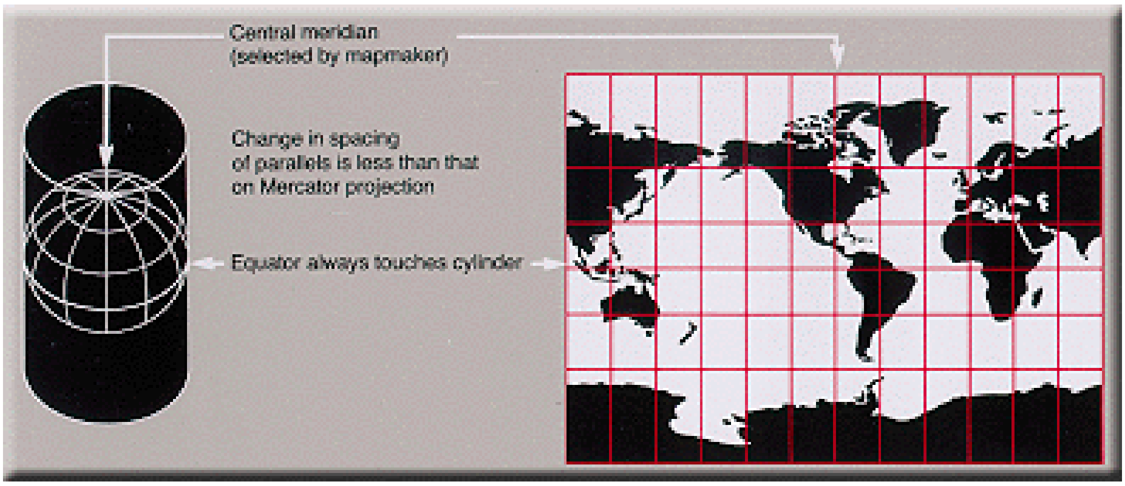
By the geometry of their construction, cylindrical projections stretch distances east-west. The amount of stretch is the same at any chosen latitude on all cylindrical projections, and is given by the secant of the latitude as a multiple of the equator's scale. The various cylindrical projections are distinguished from each other solely by their north-south stretching (where latitude is given by ϕ):

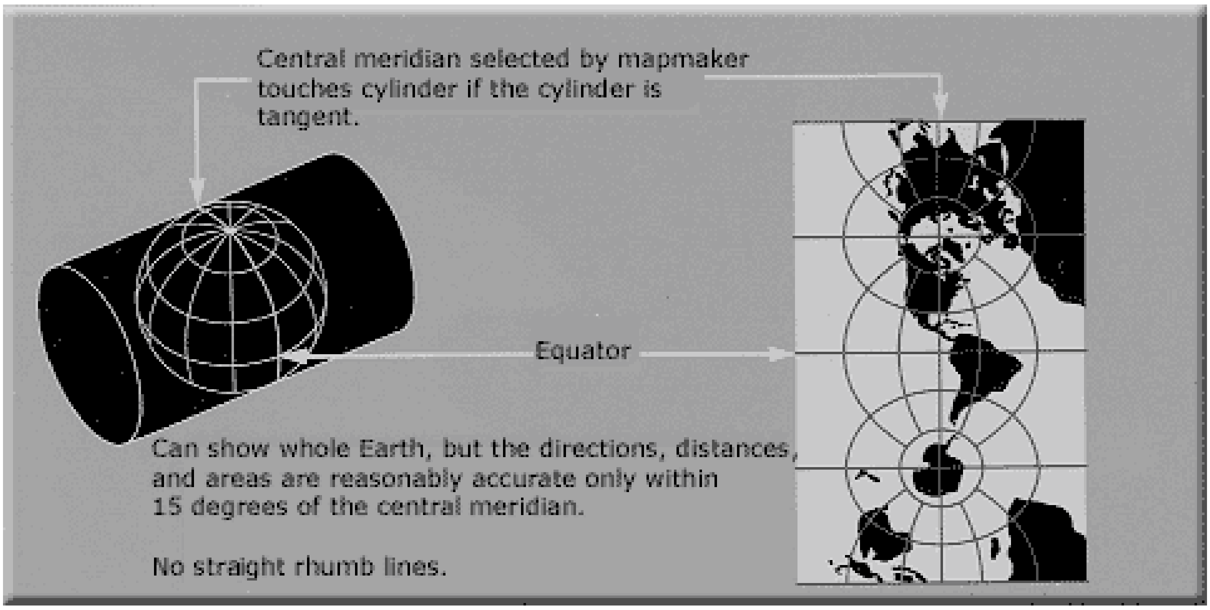
- North-south stretching equals east-west stretching (secant ϕ): The east-west scale matches the north-south scale: conformal cylindrical or Mercator; this distorts areas excessively in high latitudes (see also transverse Mercator).
- North-south stretching grows with latitude faster than east-west stretching ($secant^2$ ϕ): The cylindric perspective (or central cylindrical) projection; unsuitable because distortion is even worse than in the Mercator projection.
- North-south stretching grows with latitude, but less quickly than the east-west stretching: such as the Miller cylindrical projection (secant[4ϕ/5]).
- North-south distances neither stretched nor compressed (1): equirectangular projection or "plate carrée".
- North-south compression equals the cosine of the latitude (the reciprocal of east-west stretching): equal-area cylindrical. This projection has many named specializations differing only in the scaling constant, such as the Gall–Peters or Gall orthographic, Behrmann, and Lambert cylindrical equal-area. Since this projection scales north-south distances by the reciprocal of east-west stretching, it preserves area at the expense of shapes.

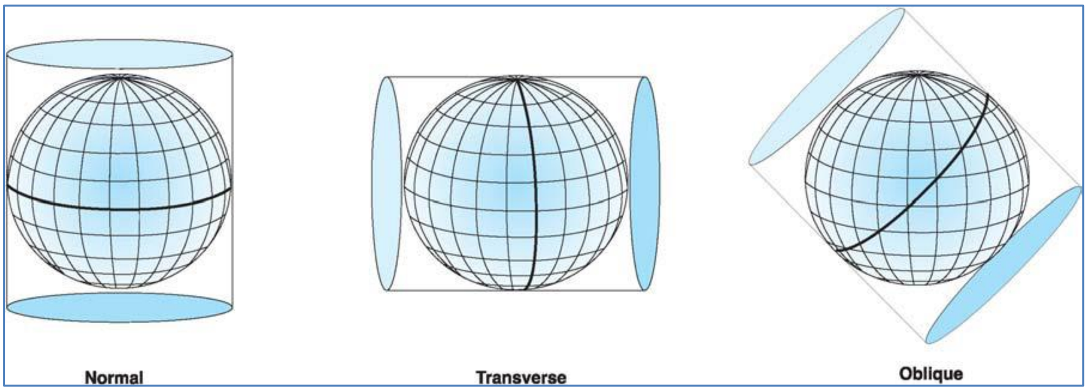
In the first case (Mercator), the east-west scale always equals the north-south scale. In the second case (central cylindrical), the north-south scale exceeds the east-west scale everywhere away from the equator. Each remaining case has a pair of secant lines—a pair of identical latitudes of opposite sign (or else the equator) at which the east-west scale matches the north-south-scale.

Normal cylindrical projections map the whole Earth as a finite rectangle, except in the first two cases, where the rectangle stretches infinitely tall while retaining constant width.

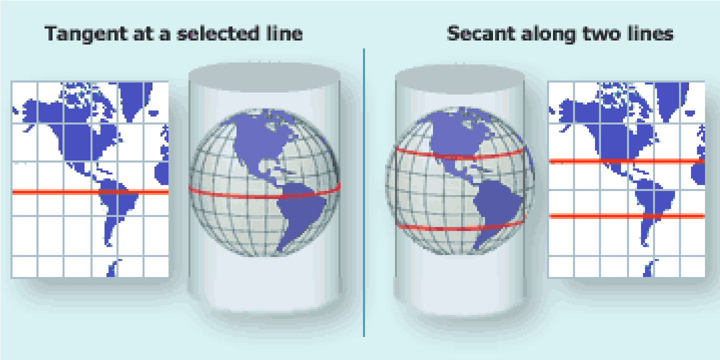





**Tangent vs. secant cylindrical projection**




Cylindrical projection - tangent and secant equatorial aspect

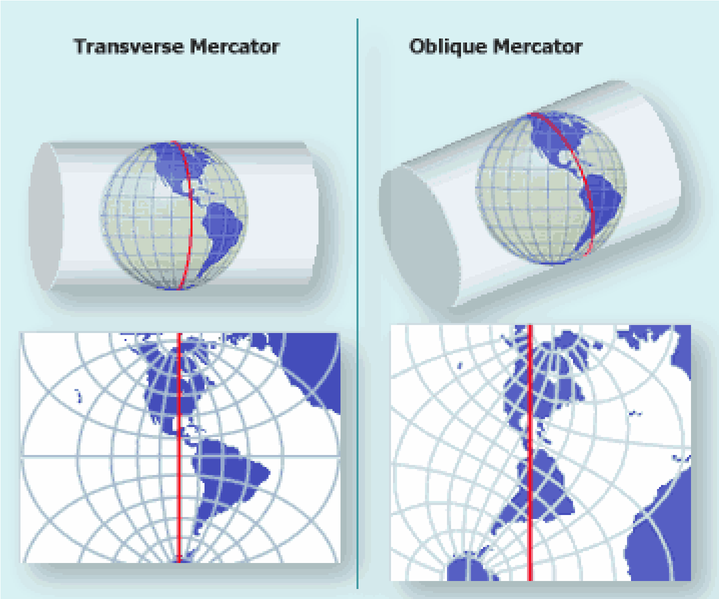
The cylinder may be either tangent or secant to the reference surface of the Earth. In the *tangent* case, the cylinder's circumference touches the reference globe's surface along a great circle (any circle having the same diameter as the sphere and thus dividing it into two equal halves). The diameter of the cylinder is equal to the diameter of the globe. The tangent line is the equator for the equatorial or normal aspect; while in the transverse aspect, the cylinder is tangent along a chosen meridian (i.e. central meridian).

In the *secant* case, the cylinder intersects the globe; that is the diameter of the cylinder is smaller than the globe's. At the place where the cylinder cuts through the globe two secant lines are formed.

The tangent and secant lines are important since scale is constant along these lines (equals that of the globe), and therefore there is no distortion (*scale factor* = 1). Such lines of *true scale* are called *standard lines*. These are lines of equidistance. Distortion increases by moving away from standard lines.

In normal aspect of cylindrical projection, the secant or standard lines are along two parallels of latitude equally spaced from equator, and are called *standard parallels*. In transverse aspect, the two standard lines run north-south parallel to meridians. Secant case provides a more even distribution of distortion throughout the map. Features appear smaller between secant lines (scale < 1) and appear larger outside these lines (scale > 1).

**Cylindrical aspect – equatorial (normal), transverse, oblique**




Cylindrical projection - transverse and oblique aspect © USGS

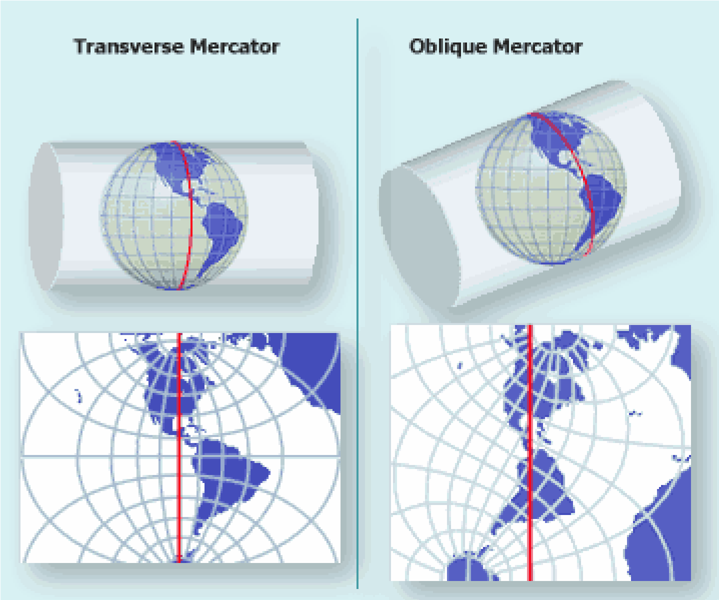
The aspect of the map projection refers to the orientation of the developable surface relative to the reference globe. The graticule layout is affected by the choice of the aspect.

In *normal* or *equatorial aspect*, the cylinder is oriented (lengthwise) parallel to the Earth's polar axis with its center located along the equator (tangent or secant). The meridians are vertical and equally spaced; the parallels of latitude are horizontal straight lines parallel to the equator with their spacing increasing toward the poles. Therefore the distortion increases towards the poles. Meridians and parallels are perpendicular to each other. The meridian that lies along the projection center is called the central meridian.

In *transverse aspect*, the cylinder is oriented perpendicular to the Earth's axis with its center located on a chosen meridian (a line going through the poles). And the *oblique aspect* refers to the cylinder being centered along a great circle between the equator and the meridians with its orientation at an angle greater than 0 and less than 90 degrees relative to the Earth's axis.

Examples of cylindrical projections include *Mercator*, *Transverse Mercator*, *Oblique Mercator*, *Plate Carré*, *Miller Cylindrical*, *Cylindrical equal-area*, *Gall–Peters*, *Hobo–Dyer*, *Behrmann*, and *Lambert Cylindrical Equal-Area* projections.

**Cylindrical aspect – equatorial (normal), transverse, oblique**




Cylindrical projection - transverse and oblique aspect © USGS

The aspect of the map projection refers to the orientation of the developable surface relative to the reference globe. The graticule layout is affected by the choice of the aspect.

In *normal* or *equatorial aspect*, the cylinder is oriented (lengthwise) parallel to the Earth's polar axis with its center located along the equator (tangent or secant). The meridians are vertical and equally spaced; the parallels of latitude are horizontal straight lines parallel to the equator with their spacing increasing toward the poles. Therefore the distortion increases towards the poles. Meridians and parallels are perpendicular to each other. The meridian that lies along the projection center is called the central meridian.

In *transverse aspect*, the cylinder is oriented perpendicular to the Earth's axis with its center located on a chosen meridian (a line going through the poles). And the *oblique aspect* refers to the cylinder being centered along a great circle between the equator and the meridians with its orientation at an angle greater than 0 and less than 90 degrees relative to the Earth's axis.

Examples of cylindrical projections include *Mercator*, *Transverse Mercator*, *Oblique Mercator*, *Plate Carré*, *Miller Cylindrical*, *Cylindrical equal-area*, *Gall–Peters*, *Hobo–Dyer*, *Behrmann*, and *Lambert Cylindrical Equal-Area* projections.

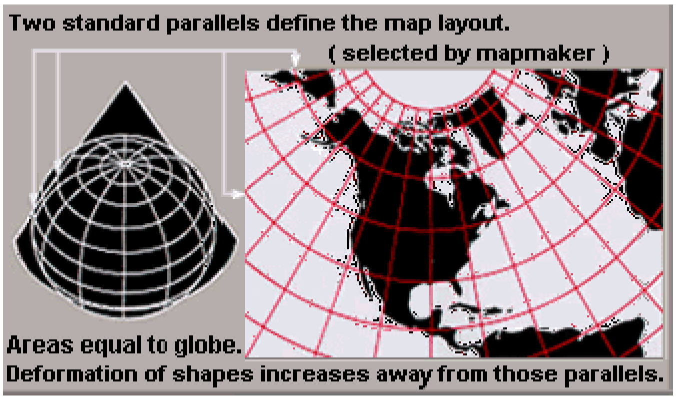
### *2. Conic*

The term "conic projection" is used to refer to any projection in which meridians are mapped to equally spaced lines radiating out from the apex and circles of latitude (parallels) are mapped to circular arcs centered on the apex.

When making a conic map, the map maker arbitrarily picks two standard parallels. Those standard parallels may be visualized as secant lines where the cone intersects the globe—or, if the map maker chooses the same parallel twice, as the tangent line where the cone is tangent to the globe. The resulting conic map has low distortion in scale, shape, and area near those standard parallels. Distances along the parallels to the north of both standard parallels or to the south of both standard parallels are stretched; distances along parallels between the standard parallels are compressed. When a single standard parallel is used, distances along all other parallels are stretched.

The most popular conic maps include:

- Equidistant conic, which keeps parallels evenly spaced along the meridians to preserve a constant distance scale along each meridian, typically the same or similar scale as along the standard parallels.
- Albers conic, which adjusts the north-south distance between non-standard parallels to compensate for the east-west stretching or compression, giving an equal-area map.
- Lambert conformal conic, which adjusts the north-south distance between non-standard parallels to equal the east-west stretching, giving a conformal map.




### 3. *Azimuthal (projections onto a plane)*

In planar (also known as azimuthal or zenithal) projections, the reference spherical surface is projected onto a plane.

Azimuthal projections have the property that directions from a central point are preserved and therefore great circles through the central point are represented by straight lines on the map. Usually these projections also have radial symmetry in the scales and hence in the distortions: map distances from the central point are computed by a function r(d) of the true distance d, independent of the angle; correspondingly, circles with the central point as center are mapped into circles which have as center the central point on the map.

The mapping of radial lines can be visualized by imagining a plane tangent to the Earth, with the central point as tangent point.

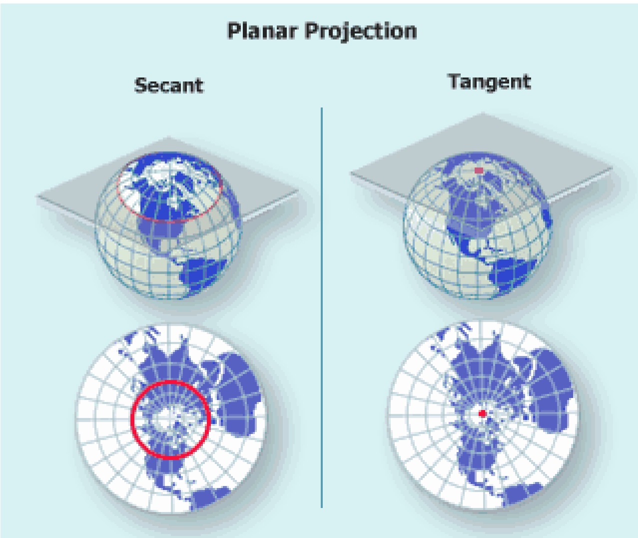
Some azimuthal projections are true perspective projections; that is, they can be constructed mechanically, projecting the surface of the Earth by extending lines from a point of perspective (along an infinite line through the tangent point and the tangent point's antipode) onto the plane:

- The gnomonic projection displays great circles as straight lines. Can be constructed by using a point of perspective at the center of the Earth. r(d) = c tan(d/R); a hemisphere already requires an infinite map.
- The General Perspective projection can be constructed by using a point of perspective outside the earth. Photographs of Earth (such as those from the International Space Station) give this perspective.
- The orthographic projection maps each point on the earth to the closest point on the plane. Can be constructed from a point of perspective an infinite distance from the tangent point; r(d) = c sin(d/R). Can display up to a hemisphere on a finite circle. Photographs of Earth from far enough away, such as the Moon, give this perspective.
- The azimuthal conformal projection, also known as the stereographic projection, can be constructed by using the tangent point's antipode as the point of perspective. r(d) = c tan(d/2R); the scale is $c/(2R\cos^2(d/2R))$. Can display nearly the entire sphere's surface on a finite circle. The sphere's full surface requires an infinite map.

Other azimuthal projections are not true perspective projections:

- Azimuthal equidistant: r(d) = cd; it is used by amateur radio operators to know the direction to point their antennas toward a point and see the distance to it. Distance from the tangent point on the map is proportional to surface distance on the earth for the case where the tangent point is the North Pole, see the flag of the United Nations)
- Lambert azimuthal equal-area. Distance from the tangent point on the map is proportional to straight-line distance through the earth: r(d) = c sin(d/2R)
- Logarithmic azimuthal is constructed so that each point's distance from the center of the map is the logarithm of its distance from the tangent point on the Earth.

**Tangent vs. secant planar projection**




The plane in planar projections may be tangent to the globe at a single point or may be secant. In the secant case the plane intersects the globe along a small circle forming a standard parallel which has true scale. The normal polar aspect yields parallels as concentric circles, and meridians projecting as straight lines from the center of the map. The distortion is minimal around the point of tangency in the tangent case, and close to the standard parallel in the secant case.

## Planar aspect – polar (normal), transverse (equatorial), oblique

The *polar aspect* is the normal aspect of the planar projection. The plane is tangent to North or South Pole at a single point or is secant along a parallel of latitude (standard parallel). The polar aspect yields parallels of latitude as concentric circles around the center of the map, and meridians projecting as straight lines from this center. Azimuthal projections are used often for mapping Polar Regions, the polar aspect of these projections are also referred to as *polar azimuthal* projections.

In transverse aspect of planar projections, the plane is oriented perpendicular to the equatorial plane. And for the oblique aspect, the plane surface has an orientation between polar and transverse aspects.

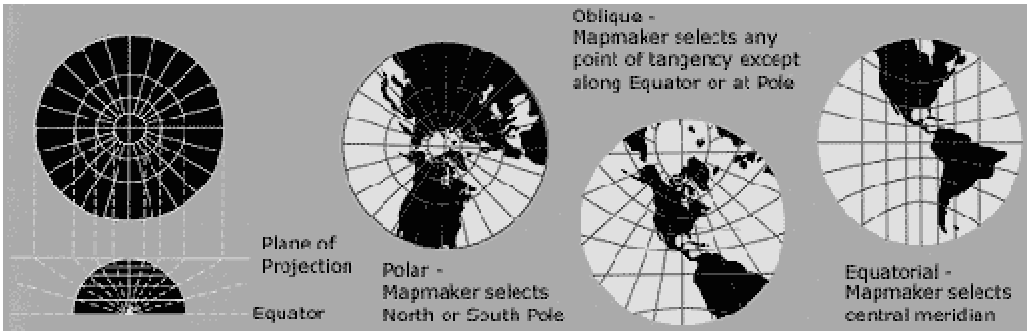
These projections are named *azimuthal* due to the fact that they preserve direction property from the center point of the projection. Great circles passing through the center point are drawn as straight lines.

Examples of azimuthal projections include: *Azimuthal Equidistant*, *Lambert Azimuthal Equal-Area*, *Gnomonic*, *Stereographic*, and *Orthographic* projections.

### Azimuthal Perspective Projections

Some classic azimuthal projections are *perspective projections* and can be produced geometrically. They can be visualized as projection of points on the sphere to the plane by shining rays of light from a light source (or point of perspective). Three projections, namely gnomonic, stereographic and orthographic can be defined based on the location of the perspective point or the light source.

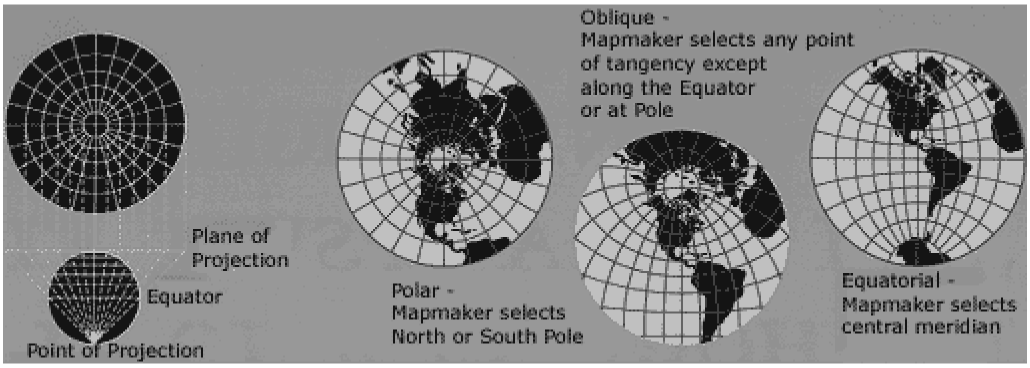
## Gnomonic Projection (also known as Central or Gnomic Projection)




Gnomonic Projection © USGS

The point of perspective or the light source is located at the center of the globe in gnomonic projections. Great circles are the shortest distance between two points on the surface of the sphere (known as great circle route). Gnomonic projections map all great circles as straight lines, and such property makes these projections suitable for use in navigation charts. Distance and shape distortion increase sharply by moving away from the center of the projection.

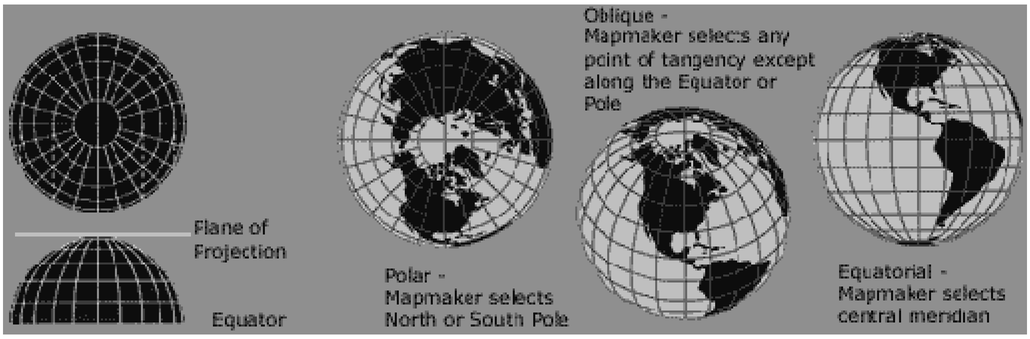
## Stereographic Projection




Stereographic projection © USGS

In stereographic projections, the perspective point is located on the surface of globe directly opposite from the point of tangency of the plane. Points close to center point show great distortion on the map. Stereographic projection is a *conformal projection*, that is over small areas angles and therefore shapes are preserved. It is often used for mapping Polar Regions (with the source located at the opposite pole).

## Orthographic Projection




Orthographic projection © USGS

In orthographic projections, the point of perspective is at infinite distance on the opposite direction from the point of tangency. The light rays travel as parallel lines. The resulting map from this projection looks like a globe (similar to seeing Earth from deep space). There is great distortion towards the borders of the map.

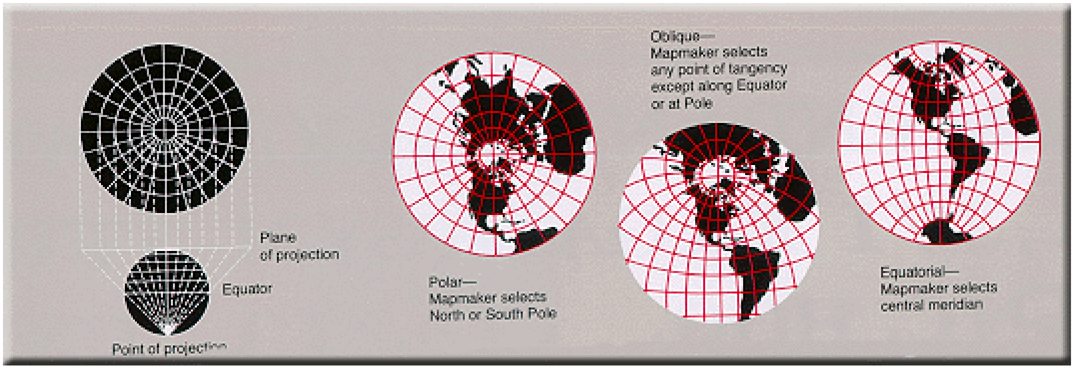
### *13. Projections by preservation of a metric property*

**Conformal**

Conformal, or orthomorphic, map projections preserve angles locally, implying that they map infinitesimal circles of constant size anywhere on the Earth to infinitesimal circles of varying sizes on the map. In contrast, mappings that are not conformal distort most such small circles into ellipses of distortion. An important consequence of conformality is that relative angles at each point of the map are correct, and the local scale (although varying throughout the map) in every direction around any one point is constant. These are some conformal projections:

- Mercator: Rhumb lines are represented by straight segments

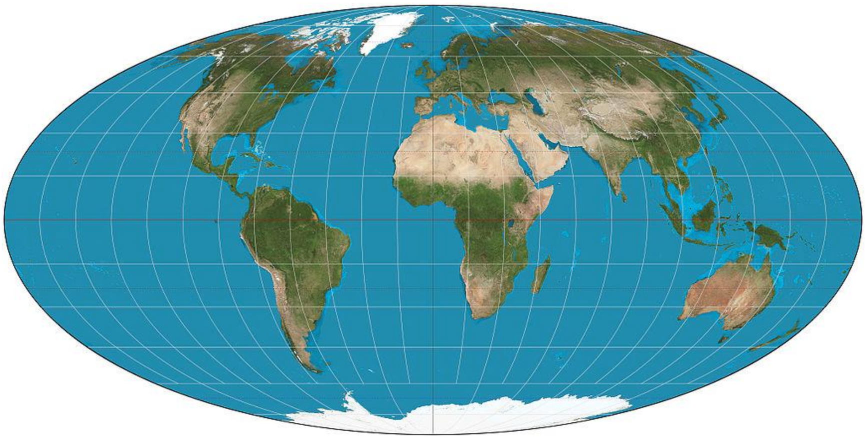
- Transverse Mercator
- Stereographic: Any circle of a sphere, great and small, maps to a circle or straight line.
- Roussilhe
- Lambert conformal conic
- Peirce quincuncial projection
- Adams hemisphere-in-a-square projection
- Guyou hemisphere-in-a-square projection




**Equal-area**

Equal-area maps preserve area measure, generally distorting shapes in order to do that. Equal-area maps are also called equivalent or authalic. These are some projections that preserve area:

- Gall orthographic (also known as Gall–Peters, or Peters, projection)
- Albers conic
- Lambert azimuthal equal-area
- Lambert cylindrical equal-area
- Mollweide
- Hammer
- Briesemeister
- Sinusoidal
- Werner
- Bonne
- Bottomley
- Goode's homolosine
- Hobo–Dyer
- Collignon
- Tobler hyperelliptical
- Snyder's equal-area polyhedral projection, used for geodesic grids.



**Equidistant**

These are some projections that preserve distance from some standard point or line:

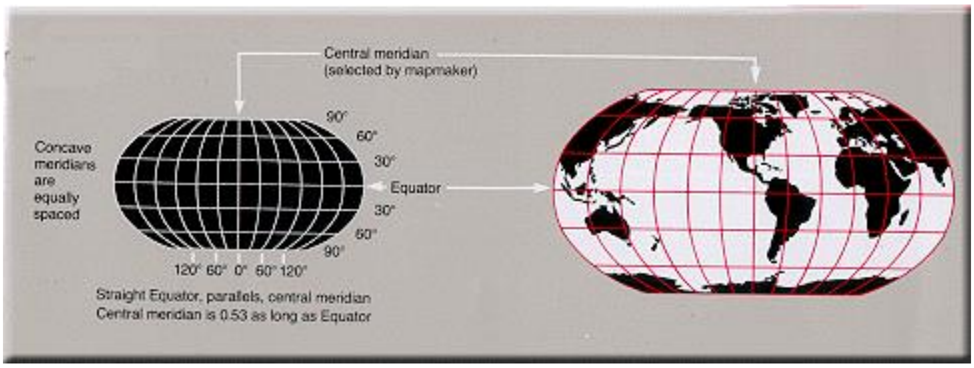
- Equirectangular—distances along meridians are conserved
- Plate carrée—an Equirectangular projection centered at the equator
- Azimuthal equidistant—distances along great circles radiating from centre are conserved
- Equidistant conic
- Sinusoidal—distances along parallels are conserved
- Werner cordiform distances from the North Pole are correct as are the curved distance on parallels
- Soldner

Two-point equidistant: two "control points" are arbitrarily chosen by the map maker. Distance from any point on the map to each control point is proportional to surface distance on the earth.

**Compromise projections**

Compromise projections give up the idea of perfectly preserving metric properties, seeking instead to strike a balance between distortions, or to simply make things "look right". Most of these types of projections distort shape in the polar regions more than at the equator. These are some compromise projections:

- Robinson
- van der Grinten
- Miller cylindrical
- Winkel Tripel
- Buckminster Fuller's Dymaxion
- B.J.S. Cahill's Butterfly Map
- Kavrayskiy VII
- Wagner VI projection
- Chamberlin trimetric
- Oronce Finé's cordiform




### 2.3.2. Geographic coordinate system

A geographic coordinate system is a coordinate system that enables every location on the Earth to be specified by a set of numbers or letters. The coordinates are often chosen such that one of the numbers represents vertical position, and two or three of the numbers represent horizontal position. A common choice of coordinates is latitude, longitude and elevation.

Latitude and Logitude

**Latitude** (*shown as a horizontal line*) is the angular distance, in degrees, minutes, and seconds of a point north or south of the Equator. Lines of latitude are often referred to as parallels.

**Longitude** (*shown as a vertical line*) is the angular distance, in degrees, minutes, and seconds, of a point east or west of the Prime (*Greenwich*) Meridian. Lines of longitude are often referred to as meridians.

**Distance between Lines** If you divide the circumference of the earth (*approximately 25,000 miles*) by 360 degrees, the distance on the earth's surface for each one degree of latitude or longitude is just over 69 miles, or 111 km. **Note:** As you move north or south of the equator, the distance between the lines of longitude gets shorter until they actually meet at the poles. At 45 degrees N or S of the equator, one degree of longitude is about 49 miles.

**Minutes and Seconds** For precision purposes, degrees of longitude and latitude have been divided into minutes (') and seconds ("). There are 60 minutes in each degree. Each minute is divided into 60 seconds. Seconds can be further divided into tenths, hundredths, or even thousandths.

**For example**, our office on Galveston Island, Texas, USA, is located at 29 degrees, 16 minutes, and 22 seconds north of the equator, and 94 degrees, 49 minutes and 46 seconds west of the Prime Meridian.

Relative Locations:

**Relative Location** of a city or destination on the planet is its relationship to another place or nearby landmarks.

As an example, our U.S. office is on **Galveston Island,** located in southeastern Texas in the Gulf of Mexico, about 48 miles southeast of Houston. That's our relative location.

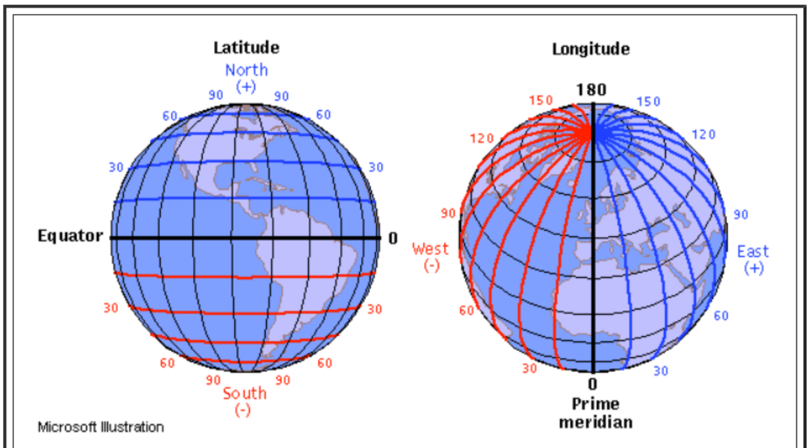
Absolute Locations

**Absolute Location** is the definitive location of a place using a recognized coordinate system. In terms of latitude and longitude, our office in Galveston, Texas, is 29°16' North, 94°49' West.

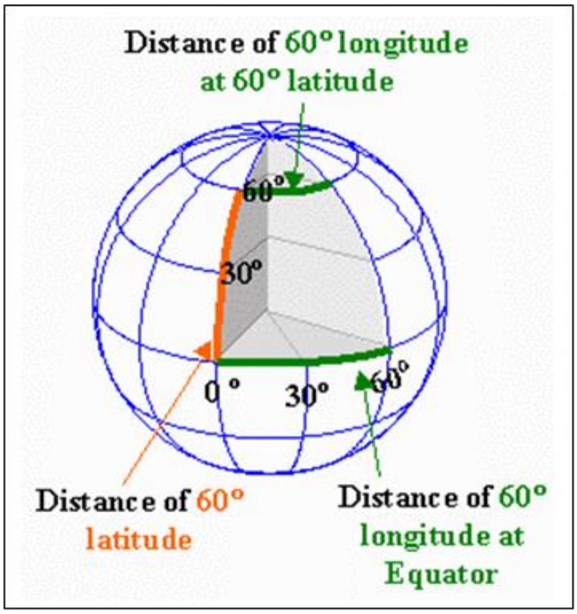
## 1. Latutude and Longitude

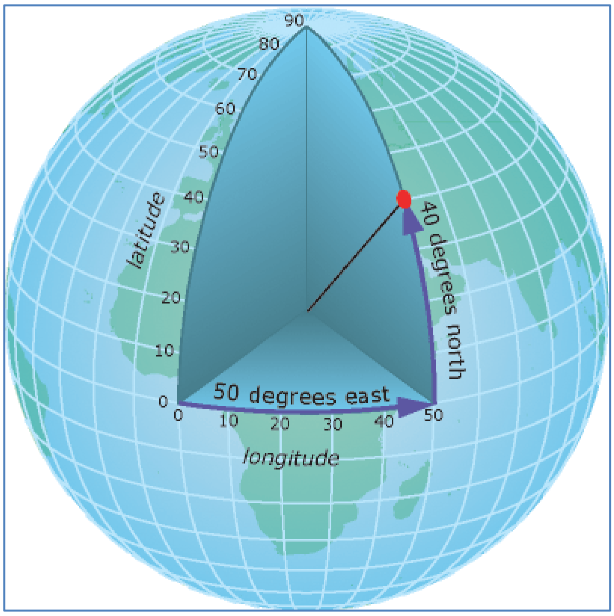
The "**latitude**" (abbreviation: Lat., ϕ, or phi) of a point on the Earth's surface is the angle between the equatorial plane and the straight line that passes through that point and is normal to the surface of a reference ellipsoid which approximates the shape of the Earth. This line passes a few kilometers away from the center of the Earth except at the poles and the equator where it passes through Earth's center. Lines joining points of the same latitude trace circles on the surface of the Earth called parallels, as they are parallel to the equator and to each other. The north pole is 90° N; the south pole is 90° S. The 0° parallel of latitude is designated the equator, the fundamental plane of all geographic coordinate systems. The equator divides the globe into Northern and Southern Hemispheres.

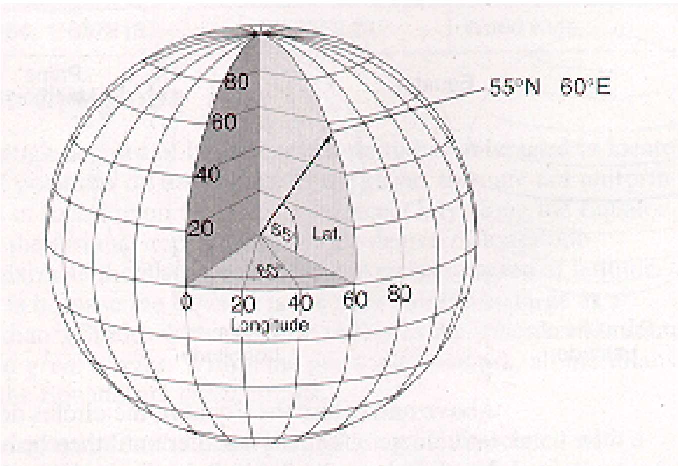
Latitude and Longitude determine the the Global Address of earth features. Every location on earth has a global address. Because the address is in numbers, people can communicate about location no matter what language they might speak. A global address is given as two numbers called coordinates. The two numbers are a location's latitude number and its longitude number ("Lat/Long").




Credit: Illinois State University

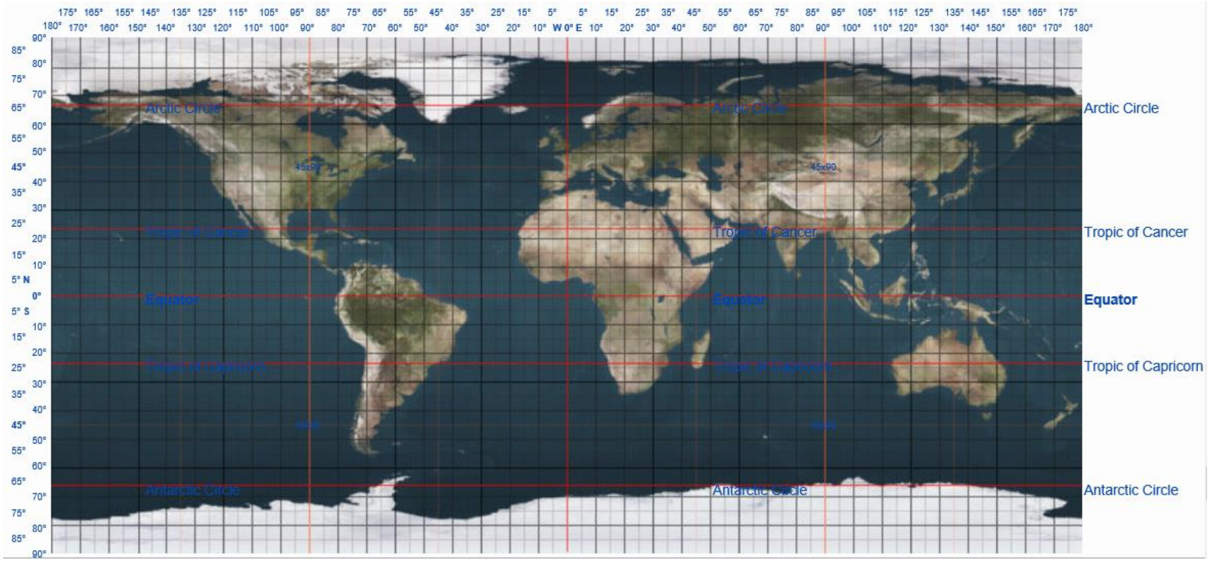



### *Grid Mapping*

Using Lat/Long is different from using a street address. Instead of having a specific street address, Lat/Long works with a numbered grid system, like what you see when you look at graph paper. It has horizontal lines and vertical lines that intersect. A location can be mapped or found on a grid system simply by giving two numbers which are the location's horizontal and vertical coordinates; or, to say it another way, the "intersection" where the place is located).



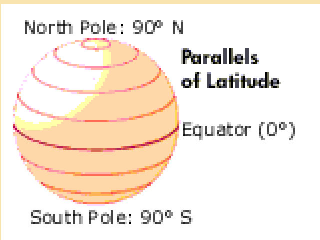

### *Grid Mapping a Globe:*

Latitude and Longitude lines are a grid map system too. But instead of being straight lines on a flat surface, Lat/Long lines encircle the Earth, either as horizontal circles or vertical half circles.

**Latitude**

Horizontal mapping lines on Earth are lines of latitude. They are known as "parallels" of latitude, because they run parallel to the equator. One simple way to visualize this might be to think about having imaginary horizontal "hula hoops" around the earth, with the biggest hoop around the equator, and then progressively smaller ones stacked above and below it to reach the North and South Poles.

**Latitude**

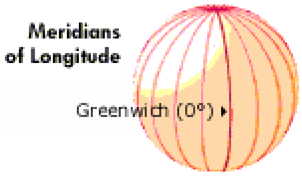



Think about having imaginary horizontal "hula hoops" around the earth, with the biggest hoop around the equator, and then progressively smaller ones stacked above and below it to reach the North and South Poles

Latitude lines are a numerical way to measure how far north or south of the equator a place is located. The equator is the starting point for measuring latitude--that's why it's marked as 0 degrees latitude. The number of latitude degrees will be larger the further away from the equator the place is located, all the way up to 90 degrees latitude at the poles. Latitude locations are given as __ degrees North or __ degrees South.

**Longitude**

Vertical mapping lines on Earth are lines of longitude, known as "meridians". One simple way to visualize this might be to think about having hula hoops cut in half, vertically positioned with one end at the North Pole and the other at the South Pole.

| Longitude | |
|---|---|
|   | Visualize hula hoops cut in half, vertically positioned with one end at the North Pole and the other at the South Pole. |

Longitude lines are a numerical way to show/measure how far a location is east or west of a universal vertical line called the Prime Meridian. This Prime Meridian line runs vertically, north and south, right over the British Royal Observatory in Greenwich England, from the North Pole to the South Pole. As the vertical starting point for longitude, the Prime Meridian is numbered 0 degrees longitude.

To measure longitude east or west of the Prime Meridian, there are 180 vertical longitude lines east of the Prime Meridian and 180 vertical longitude lines west of the Prime Meridian, so longitude locations are given as __ degrees east or __ degrees west. The 180 degree line is a single vertical line called the International Date Line, and it is directly opposite of the Prime Meridian.

The "**longitude**" (abbreviation: Long., λ, or lambda) of a point on the Earth's surface is the angle east or west from a reference meridian to another meridian that passes through that point. All meridians are halves of great ellipses (often improperly called great circles), which converge at the north and south poles.

A line, which was intended to pass through the Royal Observatory, Greenwich (a suburb of London, UK), was chosen as the international zero-longitude reference line, the Prime Meridian. Places to the east are in the eastern hemisphere, and places to the west are in the western hemisphere. The antipodal meridian of Greenwich is both 180°W and 180°E. The zero/zero point is located in the Gulf of Guinea about 625 km south of Tema, Ghana.

**Geodetic height**

To completely specify a location of a topographical feature on, in, or above the Earth, one has to also specify the vertical distance from the centre of the Earth, or from the surface of the Earth. Because of the ambiguity of "surface" and "vertical", it is more commonly expressed relative to a precisely defined vertical datum which holds fixed some known point. Each country has defined its own datum. For example, in the United Kingdom the reference point is Newlyn, while in Canada, Mexico and the United States, the point is near Rimouski, Quebec, Canada. The distance to Earth's centre can be used both for very deep positions and for positions in space

### 2.3.3. Cartesian coordinates

Every point that is expressed in ellipsoidal coordinates can be expressed as an x y z (Cartesian) coordinate. Cartesian coordinates simplify many mathematical calculations. The origin is usually the center of mass of the earth, a point close to the Earth's center of figure.

With the origin at the center of the ellipsoid, the conventional setup is the expected right-hand:

Z-axis along the axis of the ellipsoid, positive northward

X- and Y-axis in the plane of the equator, X-axis positive toward 0 degrees longitude and Y-axis positive toward 90 degrees east longitude.

An example is the NGS data for a brass disk near Donner Summit, in California. Given the dimensions of the ellipsoid, the conversion from lat/lon/height-above-ellipsoid coordinates to X-Y-Z is straightforward—calculate the X-Y-Z for the given lat-lon on the surface of the ellipsoid and add the X-Y-Z vector that is perpendicular to the ellipsoid there and has length equal to the point's height above the ellipsoid. The reverse conversion is harder: given X-Y-Z we can immediately get longitude, but no closed formula for latitude and height exists. See "Geodetic system." Using Bowring's formula in 1976 Survey Review the first iteration gives latitude correct within 10^{-11} degree as long as the point is within 10000 meters above or 5000 meters below the ellipsoid.

### 2.3.4. PROJECTION SYSTEM

#### *Types of map projections based on developable surface*

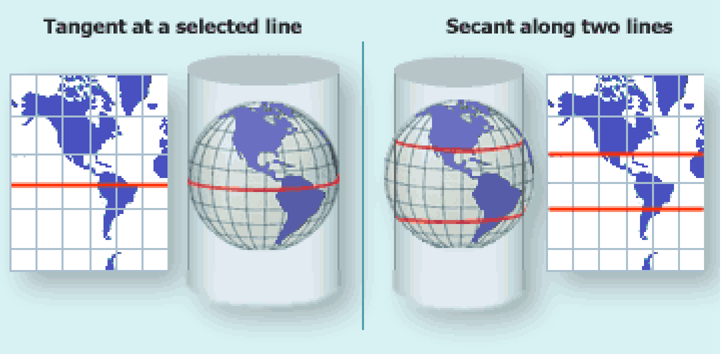
One way of classifying map projections is by the type of the **developable surface** onto which the reference sphere is projected. A developable surface is a geometric shape that can be laid out into a flat surface without stretching or tearing. The three types of developable surfaces are cylinder, cone and plane, and their corresponding projections are called *cylindrical*, *conical* and *planar*. Projections can be further categorized based on their point(s) of contact (tangent or secant) with the reference surface of the Earth and their orientation (aspect).

Keep in mind that while some projections use a geometric process, in reality most projections use mathematical equations to transform the coordinates from a globe to a flat surface. The resulting map plane in most instances can be rolled around the globe in the form of cylinder, cone or placed to the side of the globe in the case of the plane. The developable surface serves as a good illustrative analogy of the process of flattening out a spherical object onto a plane.

#### *Cylindrical projection*

In *cylindrical projections*, the reference spherical surface is projected onto a cylinder wrapped around the globe. The cylinder is then cut lengthwise and unwrapped to form a flat map.

#### *Tangent vs. secant cylindrical projection*




Cylindrical projection - tangent and secant equatorial aspect © USGS

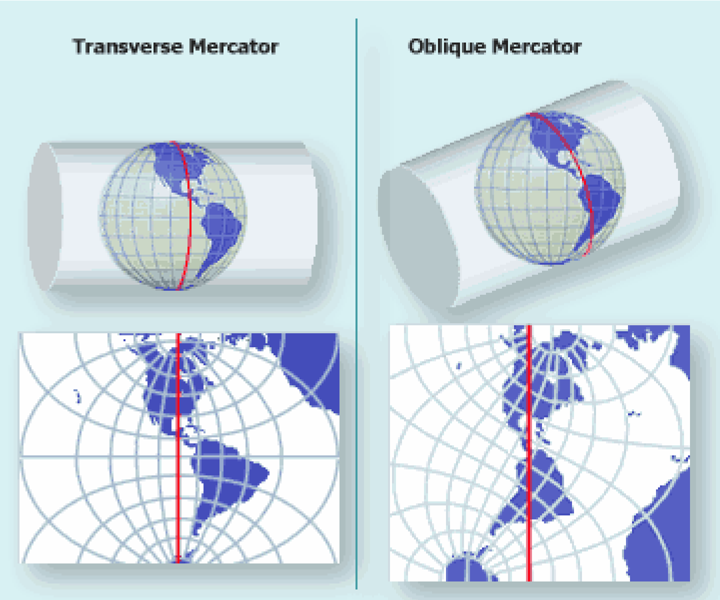
The cylinder may be either tangent or secant to the reference surface of the Earth. In the *tangent* case, the cylinder's circumference touches the reference globe's surface along a great circle (any circle having the same diameter as the sphere and thus dividing it into two equal halves). The diameter of the cylinder is equal to the diameter of the globe. The tangent line is the equator for the equatorial or normal aspect; while in the transverse aspect, the cylinder is tangent along a chosen meridian (i.e. central meridian).

In the *secant* case, the cylinder intersects the globe; that is the diameter of the cylinder is smaller than the globe's. At the place where the cylinder cuts through the globe two secant lines are formed.

The tangent and secant lines are important since scale is constant along these lines (equals that of the globe), and therefore there is no distortion (*scale factor* = 1). Such lines of *true scale* are called *standard lines*. These are lines of equidistance. Distortion increases by moving away from standard lines.

In normal aspect of cylindrical projection, the secant or standard lines are along two parallels of latitude equally spaced from equator, and are called *standard parallels*. In transverse aspect, the two standard lines run north-south parallel to meridians. Secant case provides a more even distribution of distortion throughout the map. Features appear smaller between secant lines (scale < 1) and appear larger outside these lines (scale > 1).

### *Cylindrical aspect – equatorial (normal), transverse, oblique*




Cylindrical projection - transverse and oblique aspect © USGS

The aspect of the map projection refers to the orientation of the developable surface relative to the reference globe. The graticule layout is affected by the choice of the aspect.

In *normal* or *equatorial aspect*, the cylinder is oriented (lengthwise) parallel to the Earth's polar axis with its center located along the equator (tangent or secant). The meridians are vertical and equally spaced; the parallels of latitude are horizontal straight lines parallel to the equator with their spacing increasing toward the poles. Therefore the distortion increases towards the poles. Meridians and parallels are perpendicular to each other. The meridian that lies along the projection center is called the central meridian.

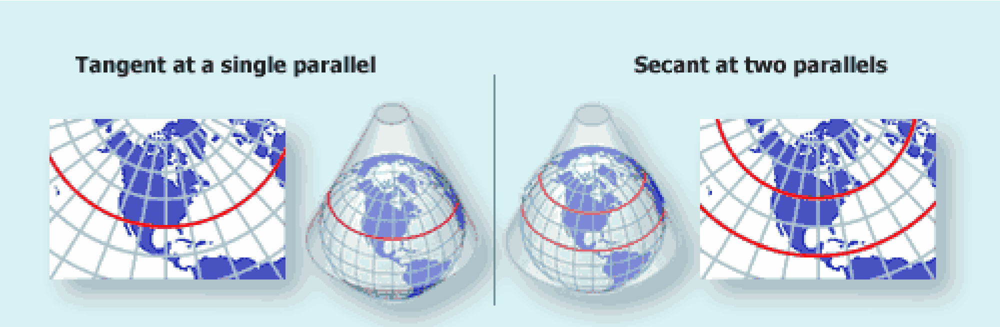
In *transverse aspect*, the cylinder is oriented perpendicular to the Earth's axis with its center located on a chosen meridian (a line going through the poles). And the *oblique aspect* refers to the cylinder being centered along a great circle between the equator and the meridians with its orientation at an angle greater than 0 and less than 90 degrees relative to the Earth's axis.

Examples of cylindrical projections include *Mercator*, *Transverse Mercator*, *Oblique Mercator*, *Plate Carré*, *Miller Cylindrical*, *Cylindrical equal-area*, *Gall–Peters*, *Hobo–Dyer*, *Behrmann*, and *Lambert Cylindrical Equal-Area* projections.

### *Conical (conic) projection*

In *conical* or *conic projections*, the reference spherical surface is projected onto a cone placed over the globe. The cone is cut lengthwise and unwrapped to form a flat map.

## Tangent vs. secant conical projection




Conic projection - tangent and secant © USGS

The cone may be either tangent to the reference surface along a small circle (any circle on the globe with a diameter less than the sphere's diameter) or it may cut through the globe and be secant (intersect) at two small circles.

For the *polar* or *normal aspect*, the cone is tangent along a parallel of latitude or is secant at two parallels. These parallels are called *standard parallels*. This aspect produces a map with meridians radiating out as straight lines from the cone's apex, and parallels drawn as concentric arcs perpendicular to meridians.

Scale is true (scale factor = 1) and there is no distortion along standard parallels. Distortion increases by moving away from standard parallels. Features appear smaller between secant parallels and appear larger outside these parallels. Secant projections lead to less overall map distortion.

### *Conical aspect – equatorial (normal), transverse, oblique*

The *polar aspect* is the *normal aspect* of the conic projection. In this aspect the cone's apex is situated along the polar axis of the Earth, and the cone is tangent along a single parallel of latitude or secant at two parallels. The cone can be situated over the North or South Pole. The polar conic projections are most suitable for maps of mid-latitude (temperate zones) regions with an east-west orientation such as the United States.

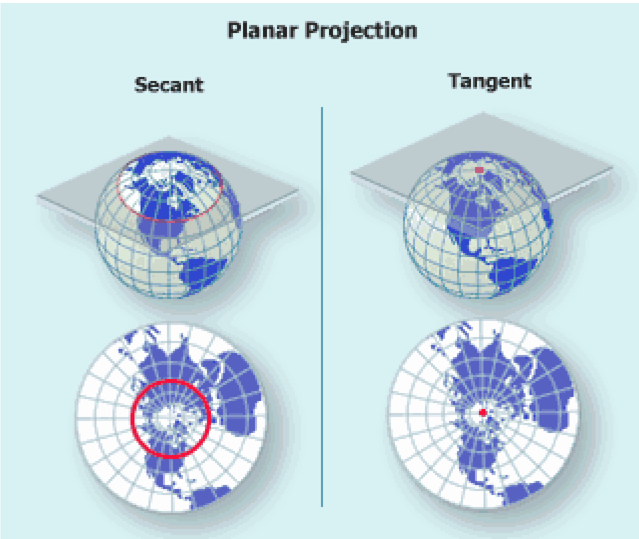
In *transverse aspect* of conical projections, the axis of the cone is along a line through the equatorial plane (perpendicular to Earth's polar axis). *Oblique aspect* has an orientation between transverse and polar aspects. Transverse and oblique aspects are seldom used.

Examples of conic projections include *Lambert Conformal Conic*, *Albers Equal Area Conic*, and *Equidistant Conic* projections.

### *Planar projection – Azimuthal or Zenithal*

In *planar* (also known as *azimuthal* or *zenithal*) projections, the reference spherical surface is projected onto a plane.

## Tangent vs. secant planar projection




Planar (azimuthal) projection - tangent and secant © USGS

The plane in planar projections may be tangent to the globe at a single point or may be secant. In the secant case the plane intersects the globe along a small circle forming a *standard parallel* which has true scale. The normal polar aspect yields parallels as concentric circles, and meridians projecting as straight lines from the center of the map. The distortion is minimal around the point of tangency in the tangent case, and close to the standard parallel in the secant case.

### *Planar aspect – polar (normal), transverse (equatorial), oblique*

The *polar aspect* is the normal aspect of the planar projection. The plane is tangent to North or South Pole at a single point or is secant along a parallel of latitude (standard parallel). The polar aspect yields parallels of latitude as concentric circles around the center of the map, and meridians projecting as straight lines from this center. Azimuthal projections are used often for mapping Polar Regions, the polar aspect of these projections are also referred to as *polar azimuthal* projections.

In transverse aspect of planar projections, the plane is oriented perpendicular to the equatorial plane. And for the oblique aspect, the plane surface has an orientation between polar and transverse aspects.

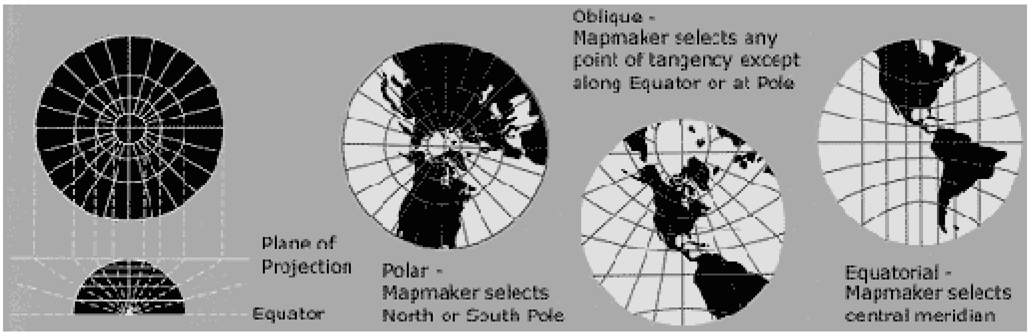
These projections are named *azimuthal* due to the fact that they preserve direction property from the center point of the projection. Great circles passing through the center point are drawn as straight lines.

Examples of azimuthal projections include: *Azimuthal Equidistant*, *Lambert Azimuthal Equal-Area*, *Gnomonic*, *Stereographic*, and *Orthographic* projections.

### *Azimuthal Perspective Projections*

Some classic azimuthal projections are *perspective projections* and can be produced geometrically. They can be visualized as projection of points on the sphere to the plane by shining rays of light from a light source (or point of perspective). Three projections, namely gnomonic, stereographic and orthographic can be defined based on the location of the perspective point or the light source.

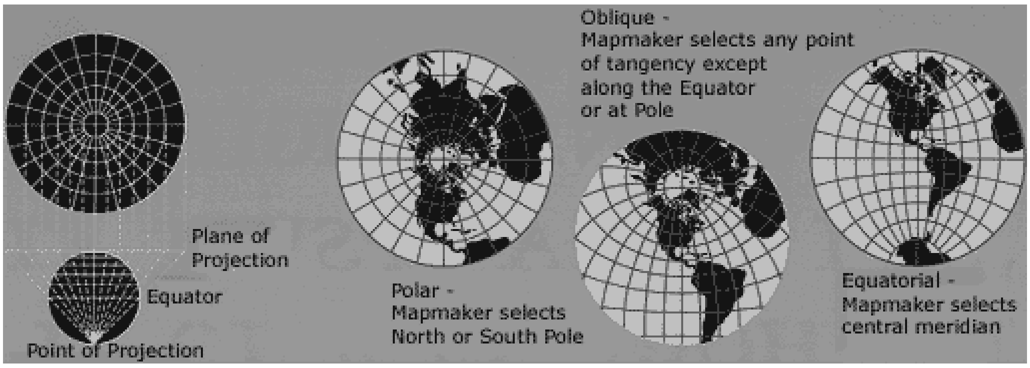
### *Gnomonic Projection (also known as Central or Gnomic Projection)*




Gnomonic Projection © USGS

The point of perspective or the light source is located at the center of the globe in gnomonic projections. Great circles are the shortest distance between two points on the surface of the sphere (known as great circle route). Gnomonic projections map all great circles as straight lines, and such property makes these projections suitable for use in navigation charts. Distance and shape distortion increase sharply by moving away from the center of the projection.

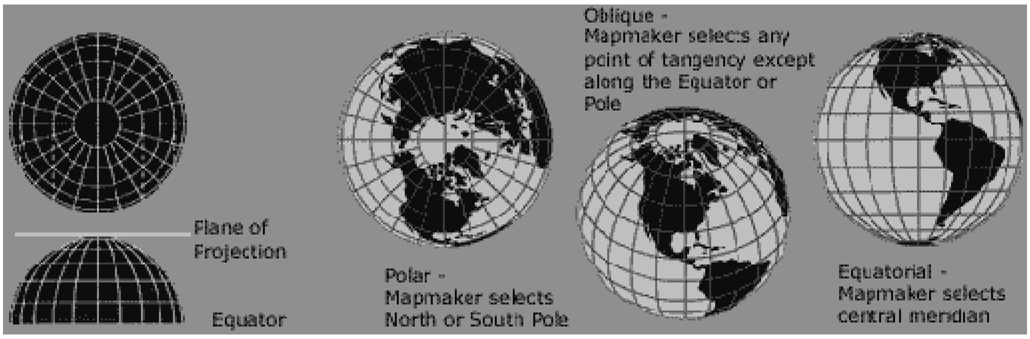
### *Stereographic Projection*




Stereographic projection © USGS

In stereographic projections, the perspective point is located on the surface of globe directly opposite from the point of tangency of the plane. Points close to center point show great distortion on the map. Stereographic projection is a *conformal projection*, that is over small areas angles and therefore shapes are preserved. It is often used for mapping Polar Regions (with the source located at the opposite pole).

### *Orthographic Projection*




Orthographic projection © USGS

In orthographic projections, the point of perspective is at infinite distance on the opposite direction from the point of tangency. The light rays travel as parallel lines. The resulting map from this projection looks like a globe (similar to seeing Earth from deep space). There is great distortion towards the borders of the map.

## 2.3.5. Map projection types based on distortion characteristics

As stated above spherical bodies such as globes can represent size, shape, distance and directions of the Earth features with reasonable accuracy. It is impossible to flatten any spherical surface (e.g. an orange peel) onto a flat surface without some stretching, tearing, or shearing. Similarly, when trying to project a spherical surface of the Earth onto a map plane, the curved surface will get deformed, causing distortions in shape (angle), area, direction or distance of features. All projections cause distortions in varying degrees; there is no one perfect projection preserving all of the above properties, rather each projection is a compromise best suited for a particular purpose.

Different projections are developed for different purposes. Some projections minimize distortion or preserve some properties at the expense of increasing distortion of others. The choice of a projection for a map depends on such factors as the purpose for which the map will be used, the area being mapped, and the map's scale (distortion is more pronounced in small-scale mapping).

### *Measuring map scale distortion – scale factor & principal (nominal) scale*

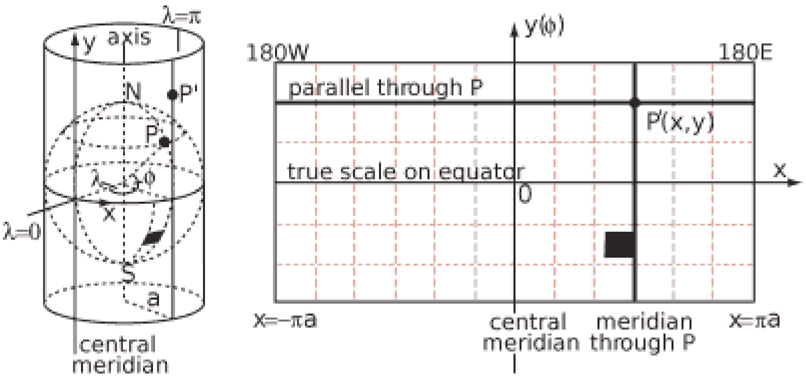
As mentioned above, a reference globe (reference surface of the Earth) is a scaled down model of the Earth. This scale can be measured as the ratio of distance on the globe to the corresponding distance on the Earth. Throughout the globe this scale is constant. For example, a 1:250000 representative fraction scale indicates that 1 unit (e.g. km) on the globe represents 250000 units on Earth. The **principal scale** or **nominal scale** of a flat map (the stated map scale) refers to this scale of its generating globe.

However the projection of the curved surface on the plane and the resulting distortions from the deformation of the surface will result in variation of scale throughout a flat map. In other words the actual map scale is different for different locations on the map plane and it is impossible to have a constant scale throughout the map. This variation of scale can be visualized by *Tissot's indicatrix* explained in detail below. Measure of scale distortion on map plane can also be quantified by the use of **scale factor**.

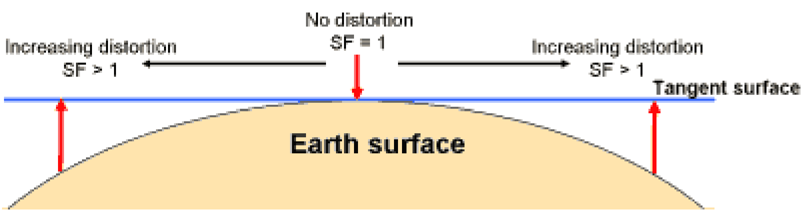
*Scale factor* is the ratio of actual scale at a location on map to the principal (nominal) map scale (SF = actual scale / nominal scale). This can be alternatively stated as ratio of distance on the map to the corresponding distance on the reference globe. A scale factor of 1 indicates actual scale is equal to nominal scale, or no scale distortion at that point on the map. Scale factors of less than or greater than one are indicative of scale distortion. The actual scale at a point on map can be obtained by multiplying the nominal map scale by the scale factor.

As an example, the actual scale at a given point on map with scale factor of 0.99860 at the point and nominal map scale of 1:50000 is equal to (1:50000 x 0.99860) = (0.99860 / 50000) = 1:50070 (which is a smaller scale than the nominal map scale). Scale factor of 2 indicates that the actual map scale is twice the nominal scale; if the nominal scale is 1:4million, then the map scale at the point would be (1:4million x 2) = 1:2million. A scale factor of 0.99950 at a given location on the map indicates that 999.5 meters on the map represents 1000 meters on the reference globe.

As mentioned above, there is no distortion along **standard lines** as evident in following figures. On a tangent surface to the reference globe, there is no scale distortion at the point (or along the line) of tangency and therefore scale factor is 1. Distortion increases with distance from the point (or line) of tangency.

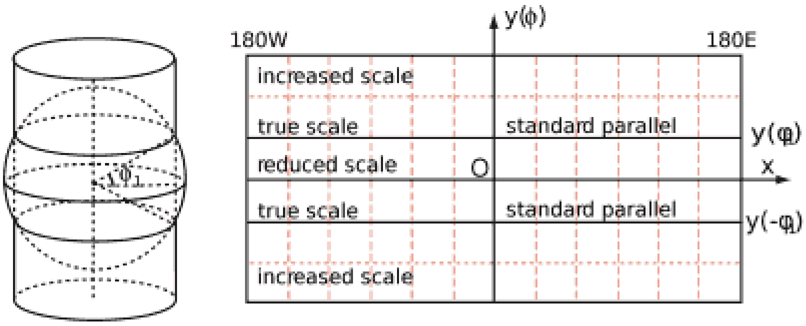



Map scale distortion of a tangent cylindrical projection - SF = 1 along line of tangency

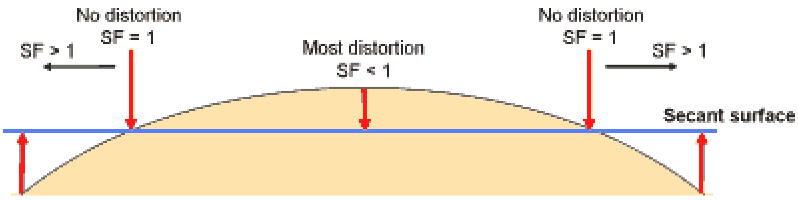



Scale distortion on a tangent surface to the globe

On a secant surface to the reference globe, there is no distortion along the standard lines (lines of intersection) where SF = 1. Between the secant lines where the surface is inside the globe, features appear smaller than in reality and scale factor is less than 1. At places on map where the surface is outside the globe, features appear larger than in reality and scale factor is greater than 1. A map derived from a secant projection surface has less overall distortion than a map from a tangent surface.




Map scale distortion of a secant cylindrical projection - SF = 1 along secant lines

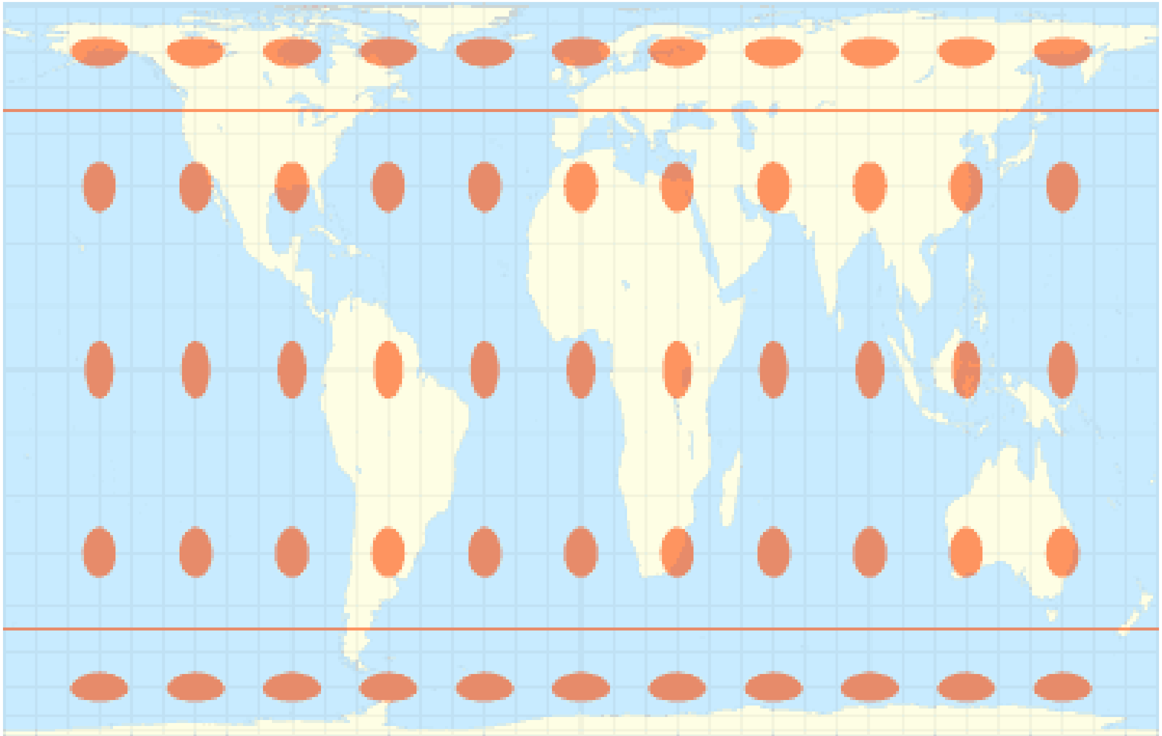



Scale distortion on a secant surface to the globe

### *Tissot's indicatrix – visualizing map distortion pattern*

A common method of classification of map projections is according to distortion characteristics - identifying properties that are preserved or distorted by a projection. The distortion pattern of a projection can be visualized by **distortion ellipses**, which are known as Tissot's indicatrices. Each indicatrix (ellipse) represents the distortion at the point it is centered on. The two axes of the ellipse indicate the directions along which the scale is maximal and minimal at that point on the map. Since scale distortion varies across the map, distortion ellipses are drawn on the projected map in an array of regular intervals to show the spatial distortion pattern across the map. The ellipses are usually centered at the intersection of meridians and parallels. Their shape represents the distortion of an imaginary circle on the spherical surface after being projected on the map plane. The size, shape and orientation of the ellipses are changed as the result of projection. Circular shapes of the same size indicate preservation of properties with no distortion occurring.

### *Equal Area Projection – Equivalent or Authalic*



Gall-Peters cylindrical equal-area projection Tissot's indicatrix

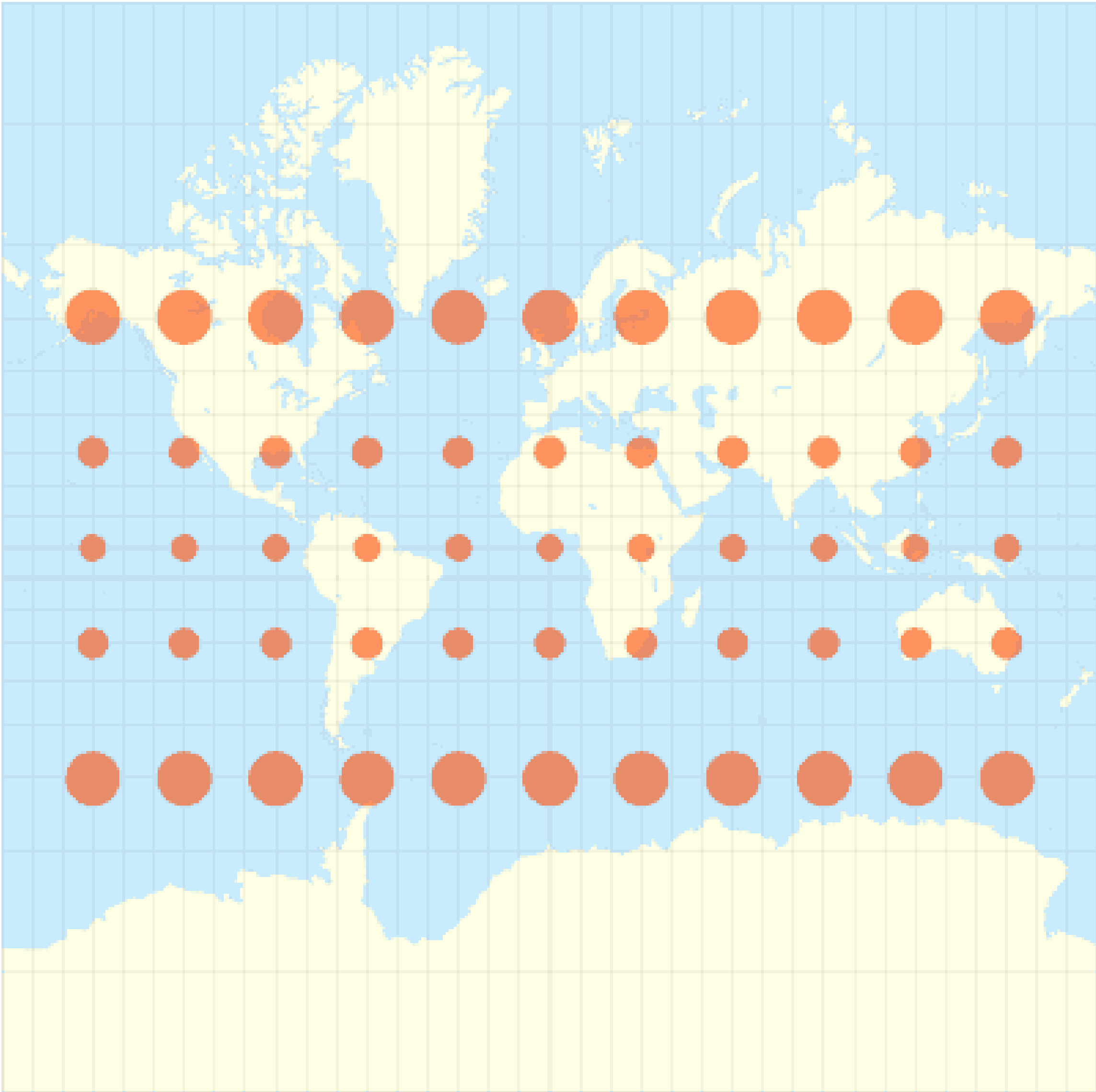
© Eric Gaba – Wikimedia Commons user: Sting

*Equal area map projections* (also known as *equivalent or authalic projection*) represent areas correctly on the map. The areas of features on the map are proportional to their areas on the reference surface of Earth. Maintaining relative areas of features causes distortion in their shapes, which is more pronounced in small-scale maps.

The shapes of the Tissot's ellipses in this world map Gall-Peters cylindrical equal-area projection are distorted; however each of them occupies the same amount of area. Along the standard parallel lines in this map (45° N and 45°S), there is no scale distortion and therefore the ellipses would be circular.

Equal area projections are useful where relative size and area accuracy of map features is important (such as displaying countries / continents in world maps), as well as for showing spatial distributions and general thematic mapping such as population, soil and geological maps. Some examples are Albers Equal-Area Conic, Cylindrical Equal Area, Sinusoidal Equal Area, and Lambert Azimuthal Equal Area projections.

### *Conformal Projection – Orthomorphic or Autogonal*



Mercator - conformal projection Tissot's indicatrix



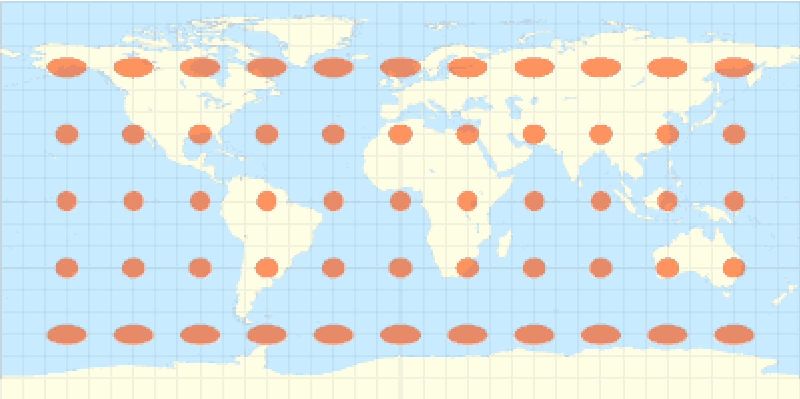
In *conformal map projections* (also known as *orthomorphic or autogonal projection*) local angles are preserved; that is angles about every point on the projected map are the same as the angles around the point on the curved reference surface. Similarly constant local scale is maintained in every direction around a point. Therefore shapes are represented accurately and without distortion for small areas. However shapes of large areas do get distorted. Meridians and parallels intersect at right angles. As a result of preserving angles and shapes, area or size of features are distorted in these maps. No map can be both conformal and equal area.

Tissot's indicatrices are all circular (shape preserved) in this world map **Mercator projection**, however they vary in size (area distorted). Here the area distortion is more pronounced as we move towards the poles. A classic example of area exaggeration is the comparison of land masses on the map, where for example Greenland appears bigger than South America and comparable in size to Africa, while in reality it is about one-eight the size of S. America and one-fourteenth the size of Africa.

A feature that has made Mercator projection especially suited for nautical maps and navigation is the representation of rhumb line or loxodrome (line that crosses meridians at the same angle) as a straight line on the map. A straight line drawn on the Mercator map represents an accurate compass bearing.

Preservation of angles makes conformal map projections suitable for navigation charts, weather maps, topographic mapping, and large scale surveying. Examples of common conformal projections include Lambert Conformal Conic, Mercator, Transverse Mercator, and Stereographic projection.

### *Equidistant Projection*



Equirectangular (equidistant cylindrical) projection Tissot's indicatrix
© Eric Gaba – Wikimedia Commons user: Sting

In *equidistant map projections*, accurate distances (constant scale) are maintained only between one or two points to every other point on the map. Also in most projections there are one or more standard lines along which scale remains constant (true scale). Distances measured along these lines are proportional to the same distance measurement on the curved reference surface. Similarly if a projection is centered on a point, distances to every other point from the center point remain accurate. Equidistant projections are neither conformal nor equal-area, but rather a compromise between them.

In this world map *equidistant cylindrical projection* (also known as *plate carrée*), Tissot's ellipses are distorted in size and shape. However while there are changes in the ellipses, their north-south axis has remained equal in length. This indicates that any line joining north and south poles (meridian) is true to scale and therefore distances are accurate along these lines. Plate carrée is a case of *equirectangular projection* with Equator being a standard parallel.

Equidistant projections are used in air and sea navigation charts, as well as radio and seismic mapping. They are also used in atlases and thematic mapping. Examples of equidistant projections are azimuthal equidistant, equidistant conic, and equirectangular projections.

### *True-Direction Projection – Azimuthal or Zenithal*

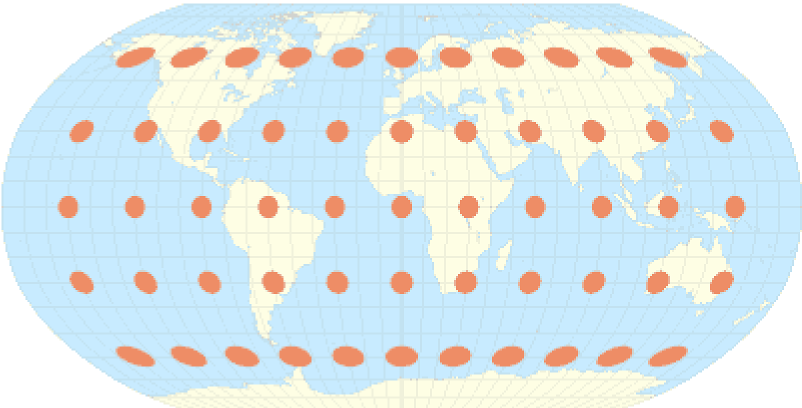
Gnomonic projection © Wikimedia Commons

Directions from a central point to all other points are maintained accurately in *azimuthal projections* (also known as *zenithal or true-direction projections*). These projections can also be equal area, conformal or equidistant.

The *gnomonic map projection* in the image is centered on the North Pole with meridians radiating out as straight lines. In gnomonic maps great circles are displayed as straight lines. Directions are true from the center point (North Pole).

True-direction projections are used in applications where maintaining directional relationships are important, such as aeronautical and sea navigation charts. Examples include Lambert Azimuthal Equal-Area, Gnomonic, and azimuthal equidistant projections.

### *Compromise Projections*



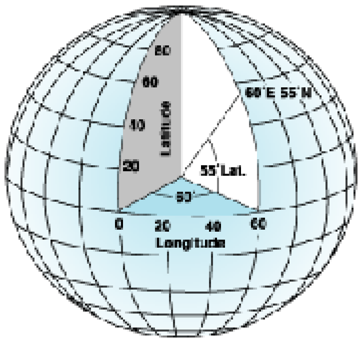
Robinson projection © Eric Gaba – Wikimedia Commons user: Sting

Some projections do not preserve any of the properties of the reference surface of the Earth; however they try to balance out distortions in area, shape, distant, and direction (thus the name compromise), so that no property is grossly distorted throughout the map and the overall view is improved. They are used in thematic mapping. Examples include Robinson projection and Winkel Tripel projection.

### *About geographic coordinate systems*

A geographic coordinate system (GCS) uses a three-dimensional spherical surface to define locations on the earth. A GCS is often incorrectly called a datum, but a datum is only one part of a GCS. A GCS includes an angular unit of measure, a prime meridian, and a datum (based on a spheroid).

A point is referenced by its longitude and latitude values. Longitude and latitude are angles measured from the earth's center to a point on the earth's surface. The angles often are measured in degrees (or in grads). The following illustration shows the world as a globe with longitude and latitude values.

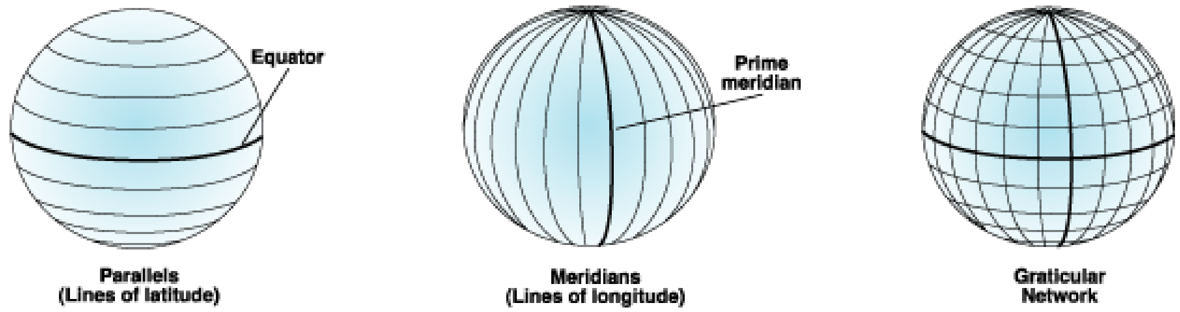



In the spherical system, horizontal lines, or east–west lines, are lines of equal latitude, or parallels. Vertical lines, or north–south lines, are lines of equal longitude, or meridians. These lines encompass the globe and form a gridded network called a graticule.

The line of latitude midway between the poles is called the equator. It defines the line of zero latitude. The line of zero longitude is called the prime meridian. For most geographic coordinate systems, the prime meridian is the longitude that passes through Greenwich, England. Other countries use longitude lines that pass through Bern, Bogota, and Paris as prime meridians. The origin of the graticule (0,0) is defined by where the equator and prime meridian intersect. The globe is then divided into four geographical quadrants that are based on compass bearings from the origin. North and south are above and below the equator, and west and east are to the left and right of the prime meridian.

View an illustration of the spherical system

This illustration shows the parallels and meridians that form a graticule.




Latitude and longitude values are traditionally measured either in decimal degrees or in degrees, minutes, and seconds (DMS). Latitude values are measured relative to the equator and range from -90° at the South Pole to +90° at the North Pole. Longitude values are measured relative to the prime meridian. They range from -180° when traveling west to 180° when traveling east. If the prime meridian is at Greenwich, then Australia, which is south of the equator and east of Greenwich, has positive longitude values and negative latitude values.

It may be helpful to equate longitude values with X and latitude values with Y. Data defined on a geographic coordinate system is displayed as if a degree is a linear unit of measure. This method is basically the same as the Plate Carrée projection.

Learn more about the Plate Carrée projection

Although longitude and latitude can locate exact positions on the surface of the globe, they are not uniform units of measure. Only along the equator does the distance represented by one degree of longitude approximate the distance represented by one degree of latitude. This is because the equator is the only parallel as large as a meridian. (Circles with the same radius as the spherical earth are called great circles. The equator and all meridians are great circles.)

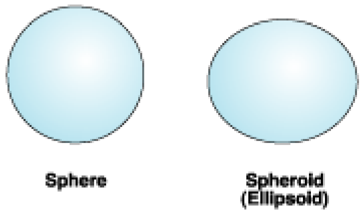
Above and below the equator, the circles defining the parallels of latitude get gradually smaller until they become a single point at the North and South Poles where the meridians converge. As the meridians converge toward the poles, the distance represented by one degree of longitude decreases to zero. On the Clarke 1866 spheroid, one degree of longitude at the equator equals 111.321 km, while at 60° latitude it is only 55.802 km. Because degrees of latitude and longitude don't have a standard length, you can’t measure distances or areas accurately or display the data easily on a flat map or computer screen.

### *Spheroids and spheres*

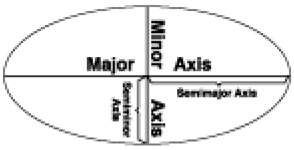
The shape and size of a geographic coordinate system's surface is defined by a sphere or spheroid. Although the earth is best represented by a spheroid, the earth is sometimes treated as a sphere to make mathematical calculations easier. The assumption that the earth is a sphere is possible for small-scale maps (smaller than 1:5,000,000). At this scale, the difference between a sphere and a spheroid is not detectable on a map. However, to maintain accuracy for larger-scale maps (scales of 1:1,000,000 or larger), a spheroid is necessary to represent the shape of the earth. Between those scales, choosing to use a sphere or spheroid will depend on the map's purpose and the accuracy of the data.

## Definition of a spheroid

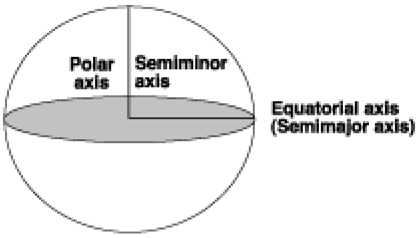
A sphere is based on a circle, while a spheroid (or ellipsoid) is based on an ellipse.




The shape of an ellipse is defined by two radii. The longer radius is called the semimajor axis, and the shorter radius is called the semiminor axis.




Rotating the ellipse around the semiminor axis creates a spheroid. A spheroid is also known as an oblate ellipsoid of revolution. The following graphic shows the semimajor and semiminor axes of a spheroid.




A spheroid is defined by either the semimajor axis, a, and the semiminor axis, b, or by a and the flattening. The flattening is the difference in length between the two axes expressed as a fraction or a decimal. The flattening, f, is:

$$f = (a - b) / a$$

The flattening is a small value, so usually the quantity 1/f is used instead. The spheroid parameters for the World Geodetic System of 1984 (WGS 1984 or WGS84) are:

a = 6378137.0 meters
b = 6356752.31424 meters
1/f = 298.257223563

The flattening ranges from zero to one. A flattening value of zero means the two axes are equal, resulting in a sphere. The flattening of the earth is approximately 0.003353. Another quantity that, like the flattening, describes the shape of a spheroid is the square of the eccentricity, $e^2$. It is represented by:

$$e^2 = \frac{a^2 - b^2}{a^2}$$

Defining different spheroids for accurate mapping

The earth has been surveyed many times to better understand its surface features and their peculiar irregularities. The surveys have resulted in many spheroids that represent the earth. Generally, a spheroid is chosen to fit one country or a particular area. A spheroid that best fits one region is not necessarily the same one that fits another region. Until recently, North American data used a spheroid determined by Clarke in 1866. The semimajor axis of the Clarke 1866 spheroid is 6,378,206.4 meters, and the semiminor axis is 6,356,583.8 meters.

Because of gravitational and surface feature variations, the earth is neither a perfect sphere nor a perfect spheroid. Satellite technology has revealed several elliptical deviations; for example, the South Pole is closer to the equator than the North Pole. Satellite-determined spheroids are replacing the older ground-measured spheroids. For example, the new standard spheroid for North America is the Geodetic Reference System of 1980 (GRS 1980), whose radii are 6,378,137.0 and 6,356,752.31414 meters. The GRS 1980 spheroid parameters were set by the International Union for Geodesy and Geophysics in 1979.

Because changing a coordinate system's spheroid will change all feature coordinate values, many organizations haven't switched to newer (and more accurate) spheroids.

### *Datums*

While a spheroid approximates the shape of the earth, a datum defines the position of the spheroid relative to the center of the earth. A datum provides a frame of reference for measuring locations on the surface of the earth. It defines the origin and orientation of latitude and longitude lines. Learn more about spheroids and spheres

Whenever you change the datum, or more correctly, the geographic coordinate system, the coordinate values of your data will change. Here are the coordinates in DMS of a control point in Redlands, California, on the North American Datum of 1983 (NAD 1983 or NAD83).

-117 12 57.75961
34 01 43.77884

Here's the same point on the North American Datum of 1927 (NAD 1927 or NAD27).

-117 12 54.61539

34 01 43.72995

The longitude value differs by approximately three seconds, while the latitude value differs by about 0.05 seconds.

NAD 1983 and the World Geodetic System of 1984 (WGS 1984) are identical for most applications. Here are the coordinates for the same control point based upon WGS 1984.

-117 12 57.75961

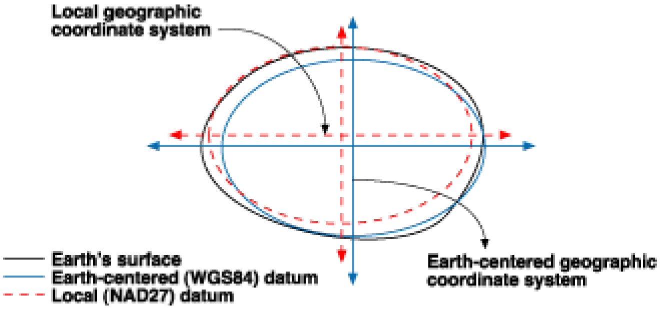
34 01 43.778837

### Geocentric datums

In the last 15 years, satellite data has provided geodesists with new measurements to define the best earth-fitting spheroid, which relates coordinates to the earth's center of mass. An earth-centered, or geocentric, datum uses the earth's center of mass as the origin. The most recently developed and widely used datum is WGS 1984. It serves as the framework for locational measurement worldwide.

### Local datums

A local datum aligns its spheroid to closely fit the earth's surface in a particular area. A point on the surface of the spheroid is matched to a particular position on the surface of the earth. This point is known as the origin point of the datum. The coordinates of the origin point are fixed, and all other points are calculated from it.




The coordinate system origin of a local datum is not at the center of the earth. The center of the spheroid of a local datum is offset from the earth's center. NAD 1927 and the European Datum of 1950 (ED 1950) are local datums. NAD 1927 is designed to fit North America reasonably well, while ED 1950 was created for use in Europe. Because a local datum aligns its spheroid so closely to a particular area on the earth's surface, it's not suitable for use outside the area for which it was designed.

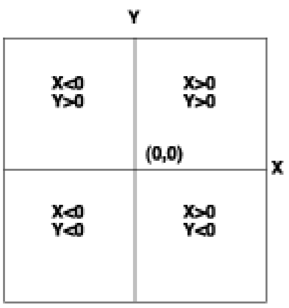
### *About projected coordinate systems*

A projected coordinate system is defined on a flat, two-dimensional surface. Unlike a geographic coordinate system, a projected coordinate system has constant lengths, angles, and areas across the two dimensions. A projected coordinate system is always based on a geographic coordinate system that is based on a sphere or spheroid.

In a projected coordinate system, locations are identified by x,y coordinates on a grid, with the origin at the center of the grid. Each position has two values that reference it to that central location. One specifies its horizontal position and the other its vertical position. The two values are called the x-coordinate and y-coordinate. Using this notation, the coordinates at the origin are x = 0 and y = 0.

On a gridded network of equally spaced horizontal and vertical lines, the horizontal line in the center is called the x-axis and the central vertical line is called the y-axis. Units are consistent and equally spaced across the full range of x and y. Horizontal lines above the origin and vertical lines to the right of the origin have positive values; those below or to the left have negative values. The four quadrants represent the four possible combinations of positive and negative X and Y coordinates.

When working with data in a geographic coordinate system, it is sometimes useful to equate the longitude values with the X axis and the latitude values with the Y axis.



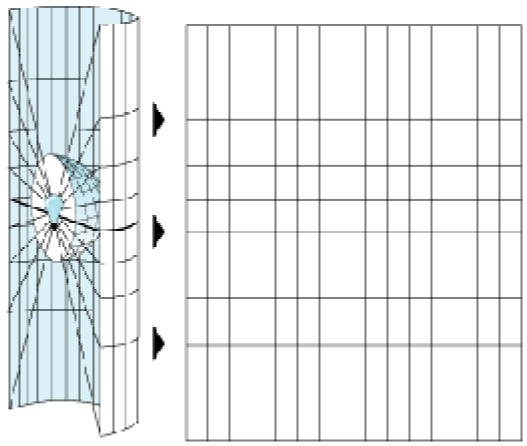
### *About map projections*

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. One easy way to understand how map projections alter spatial properties is to visualize shining a light through the earth onto a surface, called the projection surface. Imagine the earth's surface is clear with the graticule drawn on it. Wrap a piece of paper around the earth. A light at the center of the earth will cast the shadows of the graticule onto the piece of paper. You can now unwrap the paper and lay it flat. The shape of the graticule on the flat paper is different from that on the earth. The map projection has distorted the graticule.

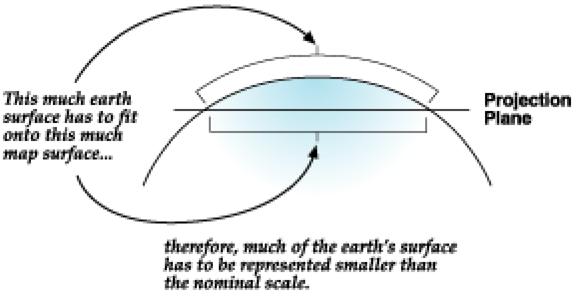
A spheroid can't be flattened to a plane any more easily than a piece of orange peel can be flattened—it will rip. Representing the earth's surface in two dimensions causes distortion in the shape, area, distance, or direction of the data.

A map projection uses mathematical formulas to relate spherical coordinates on the globe to flat, planar coordinates.

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. In the graphic below, data near the poles is stretched.



The following diagram shows how three-dimensional features are compressed to fit onto a flat surface.




Map projections are designed for specific purposes. One map projection might be used for large-scale data in a limited area, while another is used for a small-scale map of the world. Map projections designed for small-scale data are usually based on spherical rather than spheroidal geographic coordinate systems.

### *Conformal projections*

Conformal projections preserve local shape. To preserve individual angles describing the spatial relationships, a Conformal projection must show the perpendicular graticule lines intersecting at 90-degree angles on the map. A map projection accomplishes this by maintaining all angles. The drawback is that the area enclosed by a series of arcs may be greatly distorted in the process. No map projection can preserve shapes of larger regions.

### *Equal area projections*

Equal area projections preserve the area of displayed features. To do this, the other properties—shape, angle, and scale—are distorted. In Equal area projections, the meridians and parallels may not intersect at right angles. In some instances, especially maps of smaller regions, shapes are not obviously distorted, and distinguishing an Equal area projection from a Conformal projection is difficult unless documented or measured.

### *Equidistant projections*

Equidistant maps preserve the distances between certain points. Scale is not maintained correctly by any projection throughout an entire map. However, there are in most cases, one or more lines on a map along which scale is maintained correctly. Most Equidistant projections have one or more lines in which the length of the line on a map is the same length (at map scale) as the same line on the globe, regardless of whether it is a great or small circle, or straight or curved. Such distances are said to be true. For example, in the Sinusoidal projection, the equator and all parallels are their true lengths. In other Equidistant projections, the equator and all meridians are true. Still others (for example, Two-point Equidistant) show true scale between one or two points and every other point on the map. Keep in mind that no projection is equidistant to and from all points on a map.

True-direction projections

The shortest route between two points on a curved surface such as the earth is along the spherical equivalent of a straight line on a flat surface. That is the great circle on which the two points lie. True-direction, or Azimuthal projections maintain some of the great circle arcs, giving the directions or azimuths of all points on the map correctly with respect to the center. Some True-direction projections are also conformal, equal area, or equidistant.

### Projection types

Because maps are flat, some of the simplest projections are made onto geometric shapes that can be flattened without stretching their surfaces. These are called developable surfaces. Some common examples are cones, cylinders, and planes. A map projection systematically projects locations from the surface of a spheroid to representative positions on a flat surface using mathematical algorithms.

The first step in projecting from one surface to another is creating one or more points of contact. Each contact is called a point (or line) of tangency. A Planar projection is tangential to the globe at one point. Tangential cones and cylinders touch the globe along a line. If the projection surface intersects the globe instead of merely touching its surface, the resulting projection is a secant rather than a tangent case. Whether the contact is tangent or secant, the contact points or lines are significant because they define locations of zero distortion. Lines of true scale include the central meridian and standard parallels and are sometimes called standard lines. In general, distortion increases with the distance from the point of contact.

Many common map projections are classified according to the projection surface used: conic, cylindrical, or planar.

### Planar projections

Planar projections project map data onto a flat surface touching the globe. A Planar projection is also known as an Azimuthal projection or a Zenithal projection.

This type of projection is usually tangent to the globe at one point but may be secant, also. The point of contact may be the North Pole, the South Pole, a point on the equator, or any point in between. This point specifies the aspect and is the focus of the projection. The focus is identified by a central longitude and a central latitude. Possible aspects are polar, equatorial, and oblique.

Polar aspects are the simplest form. Parallels of latitude are concentric circles centered on the pole, and meridians are straight lines that intersect with their true angles of orientation at the pole. In other aspects, Planar projections will have graticular angles of 90 degrees at the focus. Directions from the focus are accurate.

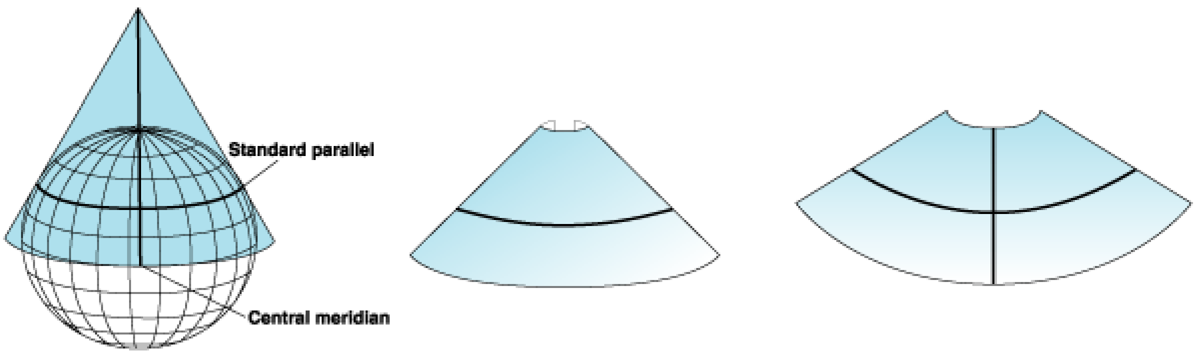
Great circles passing through the focus are represented by straight lines; thus the shortest distance from the center to any other point on the map is a straight line. Patterns of area and shape distortion are circular about the focus. For this reason, Azimuthal projections accommodate circular regions better than rectangular regions. Planar projections are used most often to map polar regions.

Some Planar projections view surface data from a specific point in space. The point of view determines how the spherical data is projected onto the flat surface. The perspective from which all locations are viewed varies between the different Azimuthal projections. The perspective point may be the center of the earth, a surface point directly opposite from the focus, or a point external to the globe, as if seen from a satellite or another planet.

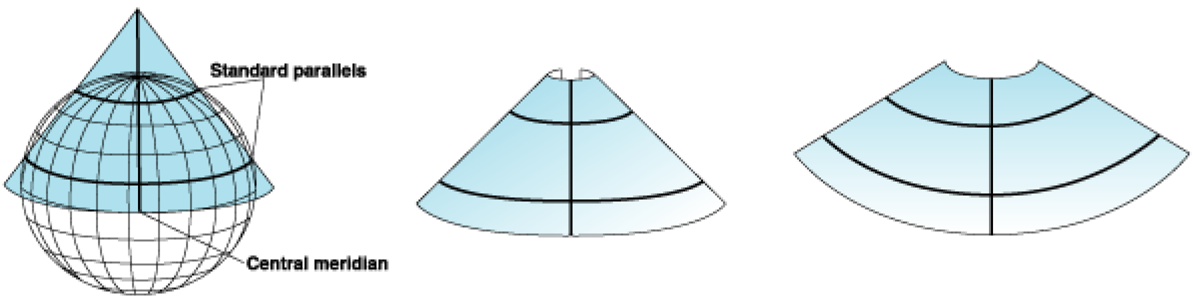
Azimuthal projections are classified in part by the focus and, if applicable, by the perspective point. The Gnomonic projection views the surface data from the center of the earth, whereas the Stereographic projection views it from pole to pole. The Orthographic projection views the earth from an infinite point, as if from deep space. Note how the differences in perspective determine the amount of distortion toward the equator.

Each of the main projection types—Conic, Cylindrical, and Planar are illustrated below.

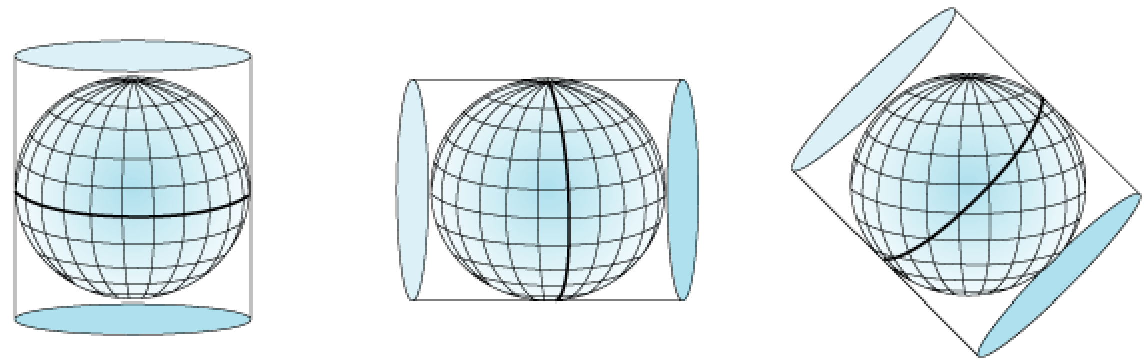
## Conic (tangent)




## Conic (secant)

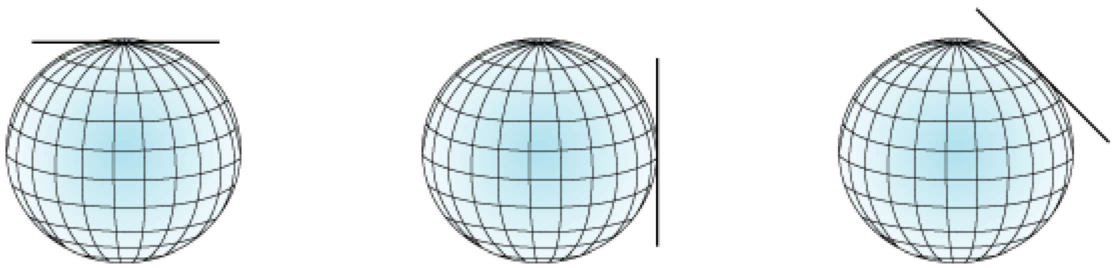



## Cylindrical Aspects



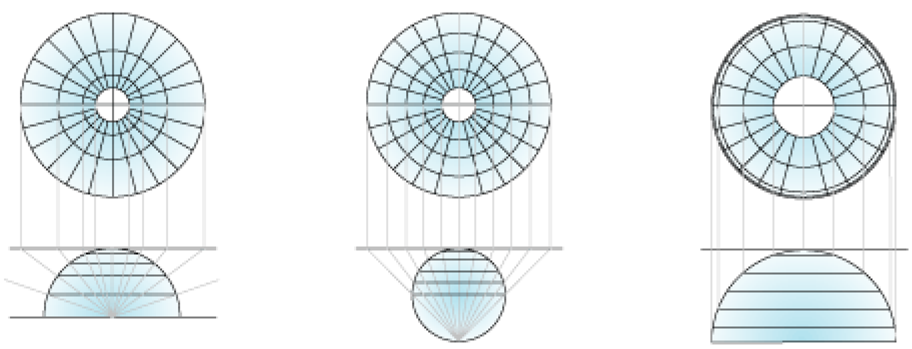
Normal Transverse Oblique

## Planar Aspects



Polar Equatorial Oblique

## Polar Aspect (different perspectives)



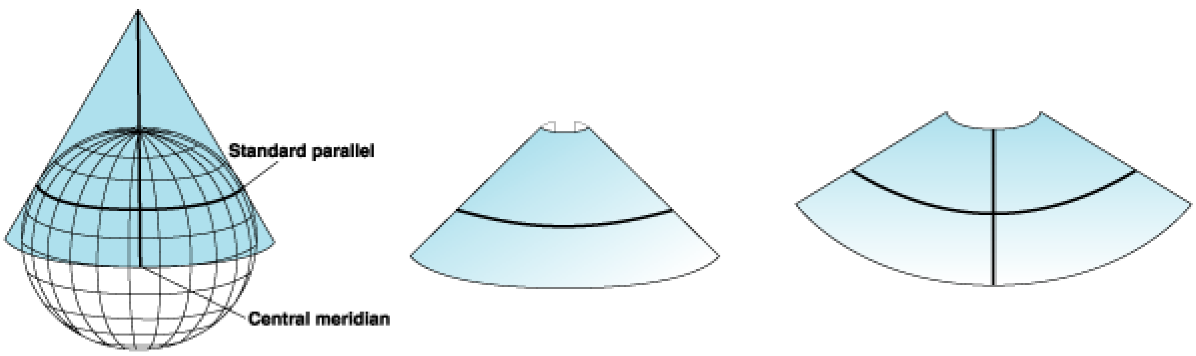
Gnomonic Stereographic Orthographic

### *Conic projections*

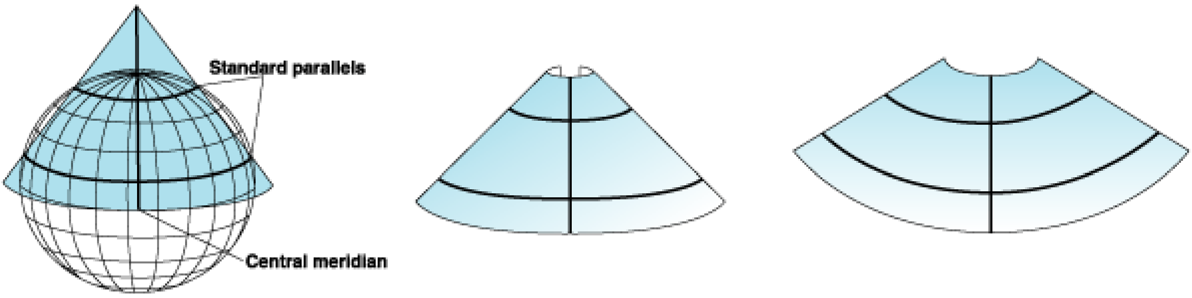
The most simple Conic projection is tangent to the globe along a line of latitude. This line is called the standard parallel. The meridians are projected onto the conical surface, meeting at the apex, or point, of the cone. Parallel lines of latitude are projected onto the cone as rings. The cone is then "cut" along any meridian to produce the final conic projection, which has straight converging lines for meridians and concentric circular arcs for parallels. The meridian opposite the cut line becomes the central meridian.

In general, the further you get from the standard parallel, the more distortion increases. Thus, cutting off the top of the cone produces a more accurate projection. You can accomplish this by not using the polar region of the projected data. Conic projections are used for midlatitude zones that have an east–west orientation.
Somewhat more complex Conic projections contact the global surface at two locations. These projections are called Secant projections and are defined by two standard parallels. It is also possible to define a Secant projection by one standard parallel and a scale factor. The distortion pattern for Secant projections is different between the standard parallels than beyond them. Generally, a Secant projection has less overall distortion than a Tangent projection. On still more complex Conic projections, the axis of the cone does not line up with the polar axis of the globe. These types of projections are called oblique.

## Conic (tangent)




## Conic (secant)




The representation of geographic features depends on the spacing of the parallels. When equally spaced, the projection is equidistant north–south but neither conformal nor equal area. An example of this type of projection is the Equidistant Conic projection. For small areas, the overall distortion is minimal. On the Lambert Conic Conformal projection, the central parallels are spaced more closely than the parallels near the border, and small geographic shapes are maintained for both small-scale and large-scale maps. On the Albers Equal Area Conic projection, the parallels near the northern and southern edges are closer together than the central parallels, and the projection displays equivalent areas.

### *Cylindrical projections*

Like Conic projections, Cylindrical projections can also have tangent or secant cases. The Mercator projection is one of the most common cylindrical projections, and the equator is usually its line of tangency. Meridians are geometrically projected onto the cylindrical surface, and parallels are mathematically projected. This produces graticular angles of 90 degrees. The cylinder is "cut" along any meridian to produce the final cylindrical projection. The meridians are equally spaced, while the spacing between parallel lines of latitude increases toward the poles. This projection is conformal and displays true direction along straight lines. On a Mercator projection, rhumb lines, lines of constant bearing, are straight lines, but most great circles are not.

Learn more about conic projections

Learn more about the Mercator projection

For more complex Cylindrical projections the cylinder is rotated, thus changing the tangent or secant lines. Transverse Cylindrical projections, such as the Transverse Mercator, use a meridian as the tangential contact or lines parallel to meridians as lines of secancy. The standard lines then run north–south, along which the scale is true. Oblique cylinders are rotated around a great circle line located anywhere between the equator and the meridians. In these more complex projections, most meridians and lines of latitude are no longer straight.

Learn more about the Transverse Mercator projection

In all Cylindrical projections, the line of tangency or lines of secancy have no distortion and thus are lines of equidistance. Other geographical properties vary according to the specific projection.

### *Projection parameters*

A map projection by itself isn't enough to define a projected coordinate system. 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 can't 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 geographic coordinate system units, while linear parameters use the projected coordinate system 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 center line of a projection. The rotation angle measures east from north. 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 don't 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. 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.

For other conic cases, the y coordinate origin is defined by the latitude of origin parameter.

- Longitude of first point
- Latitude of first point
- Longitude of second point
- Latitude of second point

The four parameters above are used with the Two-Point Equidistant and Hotine Oblique Mercator projections. They specify two geographic points that define the center axis of a projection.

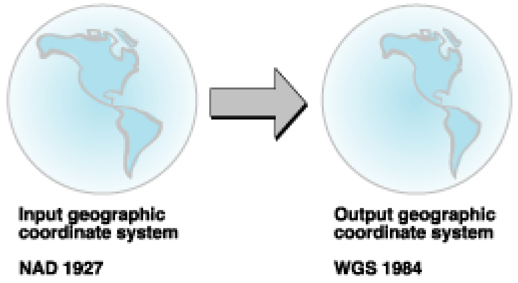
- Pseudo standard parallel 1 is used in the Krovak projection to define the oblique cone's standard parallel.
- x, y plane rotation defines the orientation of the Krovak projection along with the x scale and y scale parameters.

Unitless parameters

- Scale factor is a unitless value applied to the center point or line 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.
- Option is used in the Cube and Fuller projections. In the Cube projection, option defines the location of the polar facets. An option of 0 in the Fuller projection will display all 20 facets. Specifying an option value between 1–20 will display a single facet.

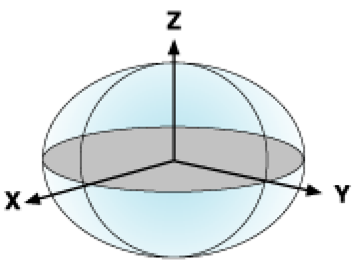
### *Geographic transformation methods*

Moving your data between coordinate systems sometimes includes transforming between the geographic coordinate systems.




Because the geographic coordinate systems contain datums that are based on spheroids, a geographic transformation also changes the underlying spheroid. There are several methods, which have different levels of accuracy and ranges, for transforming between datums. The accuracy of a particular transformation can range from centimeters to meters depending on the method and the quality and number of control points available to define the transformation parameters.

A geographic transformation is always defined in a particular direction. The picture above illustrates a transformation that converts from NAD 1927 to WGS 1984. When working with geographic transformations, if no mention is made of the direction, an application or tool like ArcMap will handle the directionality automatically. For example, if converting data from WGS 1984 to NAD 1927, you can pick a transformation called NAD_1927_to_WGS_1984_3 and the software will apply it correctly.
(ArcMap automatically loads one geographic transformation. It's designed for the lower 48 states of the United States and converts between NAD 1927 and NAD 1983.)
A geographic transformation always converts geographic (longitude–latitude) coordinates. Some methods convert the geographic coordinates to geocentric (X,Y,Z) coordinates, transform the X,Y,Z coordinates, and convert the new values back to geographic coordinates.




These include the Geocentric Translation, Molodensky, and Coordinate Frame methods.
Learn about Equation-based transformation methods

Other methods, such as NADCON and NTv2 use a grid of differences and convert the longitude–latitude values directly.

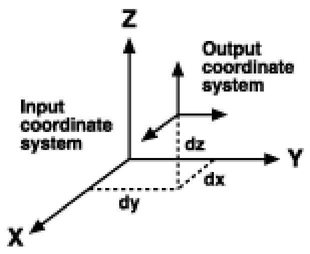
Learn about Grid-based transformation methods

### *Equation-based methods*

Equation-based transformation methods can be classified into the following four method types.

## Three-parameter methods

- The simplest datum transformation method is a geocentric, or three-parameter, transformation. The geocentric transformation models the differences between two datums in the X,Y,Z coordinate system. One datum is defined with its center at 0,0,0. The center of the other datum is defined at some distance (DX,DY,DZ) in meters away.




Usually the transformation parameters are defined as going "from" a local datum "to" WGS 1984 or another geocentric datum.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} + \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{original}$$

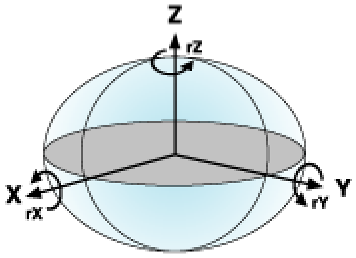
The three parameters are linear shifts and are always in meters.

## Seven-parameter methods

- A more complex and accurate datum transformation is possible by adding four more parameters to a geocentric transformation. The seven parameters are three linear shifts (DX,DY,DZ), three angular rotations around each axis (rx,ry,rz), and a scale factor.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} + (1+s) \cdot \begin{bmatrix} 1 & r_z & -r_y \\ -r_z & 1 & r_x \\ r_y & -r_x & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{original}$$

The rotation values are given in decimal seconds, while the scale factor is in parts per million (ppm). The rotation values are defined in two different ways. It's possible to define the rotation angles as positive either clockwise or counterclockwise as you look toward the origin of the X,Y,Z systems.




The previous equation is how the United States and Australia define the equations and is called the Coordinate Frame Rotation transformation. The rotations are positive counterclockwise. Europe uses a different convention called the Position Vector transformation. Both methods are sometimes referred to as the Bursa–Wolf method. In the Projection Engine, the Coordinate Frame and Bursa–Wolf methods are the same. Both Coordinate Frame and Position Vector methods are supported, and it is easy to convert transformation values from one method to the other simply by changing the signs of the three rotation values. For example, the parameters to convert from the WGS 1972 datum to the WGS 1984 datum with the Coordinate Frame method are (in the order DX,DY,DZ,rx,ry,rz,s):

(0.0, 0.0, 4.5, 0.0, 0.0, -0.554, 0.227)

To use the same parameters with the Position Vector method, change the sign of the rotation so the new parameters are:

(0.0, 0.0, 4.5, 0.0, 0.0, +0.554, 0.227)

It's impossible to tell from the parameters alone which convention is being used. If you use the wrong method, your results can return inaccurate coordinates. The only way to determine how the parameters are defined is by checking a control point whose coordinates are known in the two systems.

The Molodensky–Badekas method is a variation of the seven-parameter methods. It has an additional three parameters that define the XYZ origin of rotation. Sometimes this point is known as the origin of the datum, or geographic coordinate system. Given the XYZ origin of rotation point, it is possible to calculate an equivalent Coordinate Frame transformation. The DX, DY, and DZ values will change but the rotation and scale values will remain the same.

Molodensky method

- The Molodensky method converts directly between two geographic coordinate systems without actually converting to an X,Y,Z system. The Molodensky method requires three shifts (DX,DY,DZ) and the differences between the semimajor axes (Da) and the flattenings (Df) of the two spheroids. The Projection Engine automatically calculates the spheroid differences according to the datums involved.

$$(M+h)\Delta\varphi = -\sin\varphi\cos\lambda\Delta X - \sin\varphi\sin\lambda\Delta Y$$
$$+\cos\varphi\Delta Z + \frac{e^2\sin\varphi\cos\varphi}{(1-e^2\sin^2\varphi)^{1/2}}\Delta a$$
$$+\sin\varphi\cos\varphi(M\frac{a}{b}+N\frac{b}{a})\Delta f$$

$$(N+h)\cos\varphi\Delta\lambda = -\sin\lambda\Delta X + \cos\lambda\Delta Y$$

$$\Delta h = \cos\varphi\cos\lambda\,\Delta X + \cos\varphi\sin\lambda\,\Delta Y$$
$$+\sin\varphi\,\Delta Z - (1-e^2\sin^2\varphi)^{1/2}\Delta a$$
$$+\frac{a(1-f)}{(1-e^2\sin^2\varphi)^{1/2}}\sin^2\varphi\,\Delta f$$

- h ellipsoid height (meters)
- j latitude
- l longitude
- a semimajor axis of the spheroid (meters)
- b semiminor axis of the spheroid (meters)
- f flattening of the spheroid
- e eccentricity of the spheroid

M and N are the meridional and prime vertical radii of curvature, respectively, at a given latitude. The equations for M and N are:

$$M = \frac{a(1-e^2)}{(1-e^2\sin^2\varphi)^{3/2}}$$

$$N = \frac{a}{(1-e^2\sin^2\varphi)^{1/2}}$$

You solve for Δλand ΔΦ. The amounts are added automatically by the Projection Engine.

Abridged Molodensky method

- The Abridged Molodensky method is a simplified version of the Molodensky method. The equations are:

$$M\Delta\varphi = -\sin\varphi\cos\lambda\Delta X - \sin\varphi\sin\lambda\Delta Y$$
$$+\cos\varphi\Delta Z + (a\Delta f + f\Delta a)\cdot 2\sin\varphi\cos\varphi$$

$$N\cos\varphi\Delta\lambda = -\sin\lambda\Delta X + \cos\lambda\Delta Y$$

$$\Delta h = \cos\varphi\cos\lambda\Delta X + \cos\varphi\sin\lambda\Delta Y$$
$$+\sin\varphi\Delta Z + (a\Delta f + f\Delta a)\sin^2\varphi - \Delta a$$

### *Grid-based methods*

- Grid-based transformation methods include the following:

NADCON and HARN methods

- The United States uses a grid-based method to convert between geographic coordinate systems. Grid-based methods allow you to model the differences between the systems and are potentially the most accurate method. The area of interest is divided into cells. The National Geodetic Survey (NGS) publishes grids to convert between NAD 1927 and other older geographic coordinate systems and NAD 1983. These transformations are grouped into the NADCON method. The main NADCON grid, CONUS, converts the contiguous 48 states. The other NADCON grids convert older geographic coordinate systems to NAD 1983 for:

  - Alaska
  - Hawaiian islands
  - Puerto Rico and Virgin Islands
  - St. George, St. Lawrence, and St. Paul Islands in Alaska

The accuracy is approximately 0.15 meters for the contiguous states, 0.50 for Alaska and its islands, 0.20 for Hawaii, and 0.05 for Puerto Rico and the Virgin Islands. Accuracies can vary depending on how good the geodetic data in the area was when the grids were computed (NADCON, 1999).

The Hawaiian islands were never on NAD 1927. They were mapped using several datums that are collectively known as the Old Hawaiian datums.

New surveying and satellite measuring techniques have allowed NGS and the states to update the geodetic control point networks. As each state is finished, the NGS publishes a grid that converts between NAD 1983 and the more accurate control point coordinates. Originally, this effort was called the High Precision Geodetic Network (HPGN). It is now called the High Accuracy Reference Network (HARN). Four territories and 46 states have published HARN grids as of January 2004. HARN transformations have an accuracy approximately 0.05 meters (NADCON, 2000).

The difference values in decimal seconds are stored in two files: one for longitude and the other for latitude. A bilinear interpolation is used to calculate the exact difference between the two geographic coordinate systems at a point. The grids are binary files, but a program, NADGRD, from the NGS allows you to convert the grids to American Standard Code for Information Interchange (ASCII) format. Shown at the bottom of the page is the header and first "row" of the CSHPGN.LOA file. This is the longitude grid for Southern California. The format of the first row of numbers is, in order, the number of columns, number of rows, number of z–values (always one), minimum longitude, cell size, minimum latitude, cell size, and not used.

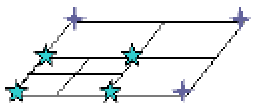
The next 37 values in this case are the longitude shifts from -122° to -113° at 32° N in 0.25°, or 15 minute, intervals in longitude.

```
NADCON EXTRACTED REGION          NADGRD

37 21 1 -122.00000 .25  32.00000 .25  .00000

.007383 .004806 .002222 -.000347 -.002868
-.005296 -.007570 -.009609 -.011305 -.012517
-.013093 -.012901 -.011867 -.009986 -.007359
-.004301 -.001389 .001164 .003282 .004814
.005503 .005361 .004420 .002580 .000053
-.002869 -.006091 -.009842 -.014240 -.019217
-.025104 -.035027 -.050254 -.072636 -.087238
-.099279 -.110968
```

## National Transformation version 2

Like the United States, Canada uses a grid-based method to convert between NAD 1927 and NAD 1983. The National Transformation version 2 (NTv2) method is quite similar to NADCON. A set of binary files contains the differences between the two geographic coordinate systems. A bilinear interpolation is used to calculate the exact values for a point.

Unlike NADCON, which can only use one grid at a time, NTv2 is designed to check multiple grids for the most accurate shift information. A set of low-density base grids exists for Canada. Certain areas such as cities have high-density local subgrids that overlay portions of the base, or parent, grids. If a point is within one of the high-density grids, NTv2 will use the high-density grid; otherwise, the point "falls through" to the low-density grid.



If a point falls in the lower-left part of the above picture between the stars, the shifts are calculated with the high-density subgrid. A point whose coordinates are anywhere else will have its shifts calculated with the low-density base grid. The software automatically calculates which base or subgrid to use.

The parent grids for Canada have spacings ranging from five to 20 minutes. The high-density grids usually have a cell size of 30 seconds, or 0.08333333°.

Unlike NADCON grids, NTv2 grids list the accuracy of each point. Accuracy values can range from a few centimeters to around a meter. The high-density grids usually have subcentimeter accuracy.

Australia and New Zealand adopted the NTv2 format to convert between geographic coordinate systems as well. Australia has released several state-based grids that convert between either Australian Geodetic Datum of 1966 (AGD 1966) or AGD 1984 and Geocentric Datum of Australia of 1994 (GDA 1994). The state grids have been merged into countrywide grids. New Zealand has released a countrywide grid to convert between New Zealand Geodetic Datum of 1949 (NZGD 1949) and NZGD 2000.

## National Transformation version 1

Like NADCON, the National Transformation version 1 (NTv1) uses a single grid to model the differences between NAD 1927 and NAD 1983 in Canada. This version is also known as CNT in ArcInfo Workstation. The accuracy is within 0.01 m of the actual difference for 74 percent of the points and within 0.5 m for 93 percent of the cases.

### *List of supported map projections*

| Map projection | Description |
|---|---|
| Aitoff | This compromise projection was developed in 1889 and used for world maps. |
| Alaska Grid | This projection was developed to provide a conformal map of Alaska with less scale distortion than other conformal projections. |
| Alaska Series E | This was developed in 1972 by the United States Geological Survey (USGS) to publish a map of Alaska at 1:2,500,000 scale. |
| Albers Equal Area Conic | This conic projection uses two standard parallels to reduce some of the distortion of a projection with one standard parallel. Shape and linear scale distortion are minimized between the standard parallels. |
| Azimuthal Equidistant | The most significant characteristic of this projection is that both distance and direction are accurate from the central point. |
| Behrmann Equal Area Cylindrical | This projection is an equal-area cylindrical projection suitable for world mapping. |
| Bipolar Oblique Conformal Conic | This projection was developed specifically for mapping North and South America and maintains conformality. |
| Bonne | This equal-area projection has true scale along the central meridian and all |

| | |
|---|---|
| | parallels. |
| Cassini-Soldner | This transverse cylindrical projection maintains scale along the central meridian and all lines parallel to it. This projection is neither equal area nor conformal. |
| Chamberlin Trimetric | This projection was developed and used by the National Geographic Society for continental mapping. The distance from three input points to any other point is approximately correct. |
| Craster Parabolic | This pseudo cylindrical equal-area projection is primarily used for thematic maps of the world. |
| Cube | This is a faceted projection that is used for ArcGlobe. |
| Cylindrical Equal Area | Lambert first described this equal-area projection in 1772. It is used infrequently. |
| Double Stereographic | This azimuthal projection is conformal. |
| Eckert I | This pseudo cylindrical projection is used primarily as a novelty map. |
| Eckert II | This is a pseudo cylindrical equal-area projection. |
| Eckert III | This pseudo cylindrical projection is used primarily for world maps. |
| Eckert IV | This equal-area projection is used primarily for world maps. |
| Eckert V | This pseudo cylindrical projection is used primarily for world maps. |
| Eckert VI | This equal-area projection is used primarily for world maps. |
| Equidistant Conic | This conic projection can be based on one or two standard parallels. As the name implies, all circular parallels are spaced evenly along the meridians. |
| Equidistant Cylindrical | This is one of the easiest projections to construct because it forms a grid of equal rectangles. |
| Equirectangular | This projection is very simple to construct because it forms a grid of equal rectangles. |
| Fuller | The final version of this interrupted projection was described by Buckminster Fuller in 1954. |
| Gall's Stereographic | The Gall's Stereographic projection is a cylindrical projection designed around 1855 with two standard parallels at latitudes 45° N and 45° S. |

| | |
|---|---|
| Gauss-Krüger | This projection is similar to the Mercator except that the cylinder is tangent along a meridian instead of the equator. The result is a conformal projection that does not maintain true directions. |
| Geocentric Coordinate System | The geocentric coordinate system is not a map projection. The earth is modeled as a sphere or spheroid in a right-handed x,y,z system. |
| Geographic Coordinate System | The geographic coordinate system is not a map projection. The earth is modeled as a sphere or spheroid. |
| Gnomonic | This azimuthal projection uses the center of the earth as its perspective point. |
| Goode's Homolosine | This interrupted equal area pseudocylindrical projection is used for world raster data. |
| Great Britain National Grid | This coordinate system uses a Transverse Mercator projected on the Airy spheroid. The central meridian is scaled to 0.9996. The origin is 49° N and 2° W. |
| Hammer-Aitoff | The Hammer–Aitoff projection is a modification of the Lambert Azimuthal Equal Area projection. |
| Hotine Oblique Mercator | This is an oblique rotation of the Mercator projection developed for conformal mapping of areas that do not follow a north–south or east–west orientation but are obliquely oriented. |
| Krovak | The Krovak projection is an oblique Lambert Conformal Conic projection designed for the former Czechoslovakia. |
| Lambert Azimuthal Equal Area | This projection preserves the area of individual polygons while simultaneously maintaining true directions from the center. |
| Lambert Conformal Conic | This projection is one of the best for middle latitudes. It is similar to the Albers Conic Equal Area projection except that the Lambert Conformal Conic projection portrays shape more accurately. |
| Local Cartesian Projection | This is a specialized map projection that does not take into account the curvature of the earth. |
| Loximuthal | This projection shows loxodromes, or rhumb lines, as straight lines with the correct azimuth and scale from the intersection of the central meridian and the central parallel. |
| McBryde-Thomas Flat-Polar Quartic | This equal-area projection is primarily used for world maps. |

| | |
|---|---|
| Mercator | Originally created to display accurate compass bearings for sea travel. An additional feature of this projection is that all local shapes are accurate and clearly defined. |
| Miller Cylindrical | This projection is similar to the Mercator projection except that the polar regions are not as areally distorted. |
| Mollweide | Carl B. Mollweide created this pseudo cylindrical projection in 1805. It is an equal-area projection designed for small-scale maps. |
| New Zealand National Grid | This is the standard projection for large-scale maps of New Zealand. |
| Orthographic | This perspective projection views the globe from an infinite distance. This gives the illusion of a three-dimensional globe. |
| Perspective | This projection is similar to the Orthographic projection in that its perspective is from space. In this projection, the perspective point is not an infinite distance away; instead, you can specify the distance. |
| Plate Carrée | This projection is very simple to construct because it forms a grid of equal rectangles. |
| Polar Stereographic | The projection is equivalent to the polar aspect of the Stereographic projection on a spheroid. The central point is either the North Pole or the South Pole. |
| Polyconic | The name of this projection translates into "many cones" and refers to the projection methodology. |
| Quartic Authalic | This pseudo cylindrical equal-area projection is primarily used for thematic maps of the world. |
| Rectified Skewed Orthomorphic | This oblique cylindrical projection is provided with two options for the national coordinate systems of Malaysia and Brunei. |
| Robinson | This is a compromise projection used for world maps. |
| Simple Conic | This conic projection can be based on one or two standard parallels. |
| Sinusoidal | As a world map, this projection maintains equal area despite conformal distortion. |
| Space Oblique Mercator | This projection is nearly conformal and has little scale distortion within the sensing range of an orbiting mapping satellite, such as Landsat. |
| State Plane Coordinate | The State Plane Coordinate System is not a projection. It is a coordinate system that divides the 50 states of the United States, Puerto Rico, and |

| | |
|---|---|
| System (SPCS) | the U.S. Virgin Islands into more than 120 numbered sections, referred to as zones. |
| Stereographic | This azimuthal projection is conformal. |
| Times | The Times projection was developed by Moir in 1965 for Bartholomew Ltd., a British mapmaking company. It is a modified Gall's Stereographic, but the Times has curved meridians. |
| Transverse Mercator | This is similar to the Mercator except that the cylinder is tangent along a meridian instead of the equator. The result is a conformal projection that does not maintain true directions. |
| Two-Point Equidistant | This modified planar projection shows the true distance from either of two chosen points to any other point on a map. |
| Universal Polar Stereographic | This form of the Polar Stereographic maps areas north of 84° N and south of 80° S that are not included in the UTM Coordinate System. The projection is equivalent to the polar aspect of the Stereographic projection of the spheroid with specific parameters. |
| Universal Transverse Mercator | The Universal Transverse Mercator coordinate system is a specialized application of the Transverse Mercator projection. The globe is divided into 60 zones, each spanning six degrees of longitude. |
| Van der Grinten I | This projection is similar to the Mercator projection except that it portrays the world as a circle with a curved graticule. |
| Vertical Near-Side Perspective | Unlike the Orthographic projection, this perspective projection views the globe from a finite distance. This perspective gives the overall effect of the view from a satellite. |
| Winkel I | This is a pseudo cylindrical projection used for world maps that averages the coordinates from the Equirectangular (Equidistant Cylindrical) and Sinusoidal projections. |
| Winkel II | This is a pseudo cylindrical projection that averages the coordinates from the Equirectangular and Mollweide projections. |
| Winkel Tripel | This is a compromise projection used for world maps that averages the coordinates from the Equirectangular (Equidistant Cylindrical) and Aitoff projections. |

# 3. SPATIAL DATA MODELING AND DATABASE DESIGN

## 3.1. INTRODUCTION TO GEOGRAPHIC PHENOMENA AND DATA MODELING

### 3.1.1. Overview:

The real world is too complex for our immediate and direct understanding. We create "models" of reality that are intended to have some similarity with selected aspects of the real world. Databases are created from these "models" as a fundamental step in coming to know the nature and status of that reality

- A spatial database is a collection of spatially referenced data that acts as a model of reality.
- A database is a model of reality in the sense that the database represents a selected set or approximation of phenomena.
- These selected phenomena are deemed important enough to represent in digital form.
- The digital representation might be for some past, present or future time period (or contain some combination of several time periods in an organized fashion).

Many of the definitions in this Unit have been standardized by the proposed US National Digital Cartographic Standard (DCDSTF, 1988). These standards have been developed to provide a nationally uniform means for portraying and exchanging digital cartographic data. These cartographic standards will form part of a larger standard being developed for the digital representation of all earth science information.

***What is a Spatial Database System?***

There is a need to manage geometric, geographic, or spatial data, which means data related to space. The space of interest can be, for example, the two-dimensional abstraction of (parts of) the surface of the earth – that is, geographic space. At least since the advent of relational database systems there have been attempts to manage such data in database systems. Characteristic for the technology emerging to address these needs is the capability to deal with large collections of relatively simple geometric objects, for example, a set of 100 000 polygons. This is somewhat different from areas like CAD databases (solid modeling etc.) where geometric entities are composed hierarchically into complex structures, although the issues are certainly related.

Several terms have been used for database systems offering such support like pictorial, image, geometric, geographic, or spatial database system. The terms “pictorial” and “image” database system arise from the fact that the data to be managed are often initially captured in the form of digital raster images (e.g. remote sensing by satellites, or computer tomography in medical applications). The term “spatial database system” has become popular during the last few years and is associated with a view of a database as containing sets of objects in space rather than images or pictures of a space. Indeed, the requirements and techniques for dealing with objects in space that have identity and well-defined extents, locations, and relationships are rather different from those for dealing with raster images. It has therefore been suggested to clearly distinguish two classes of systems called spatial database systems and image database systems, respectively. Image database systems may include analysis techniques to extract objects in space from images, and offer some spatial database functionality, but are also prepared to store, manipulate and retrieve raster images as discrete entities.

The spatial database deals with followings:

- A spatial database system is a database system.
- It offers spatial data types (SDTs) in its data model and query language.
- It supports spatial data types in its implementation, providing at least spatial indexing and efficient algorithms for spatial join.

Hence a spatial database system is a full-fledged database system with additionalcapabilities for handling spatial data. (2) Spatial data types, e.g. POINT, LINE, REGION, provide a fundamental abstraction for modeling the structure of geometric entities in space as well as their relationships (l intersects r), properties (area(r) > 1000), and operations (intersection(l, r) – the part of l lying within r).

A system must at least be able to retrieve from a large collection of objects in some space those lying within a particular area without scanning the whole set. Therefore spatial indexing is mandatory. It should also support connecting objects from different classes through some spatial relationship in a better way than by filtering the Cartesian product (at least for those relationships that are important for the application).

### *Modeling*

What needs to be represented?

There are two important alternative views of what needs to be represented:

- Objects in space: We are interested in distinct entities arranged in space each of which has its own geometric description.
- Space: We wish to describe space itself, that is, say something about every point in space.

The first view allows one to model, for example, cities, forests, or rivers. The second view is the one of thematic maps describing e.g. land use or the partition of a country into districts. Since raster images say something about every point in space, they are also closely related to the second view. We can reconcile both views to some extent by offering concepts for modeling (i) single objects, and (ii) spatially related collections of objects.

For modeling single objects, the fundamental abstractions are point, line, and region.

**A point** represents an object for which only its location in space, but not its extent, is relevant. For example, a city may be modeled as a point in a model describing a large geographic area (a large scale map).

**A line** (in this context always to be understood as meaning a curve in space, usually represented by a polyline, a sequence of line segments) is the basic abstraction for facilities for moving through space, or connections in space (roads, rivers, cables for phone, electricity, etc.).

**A region** is the abstraction for something having an extent in 2d-space, e.g. a country, a lake, or a national park. A region may have holes and may also consist of several disjoint pieces. Figure shows the three basic abstractions for single objects.

Figure: The three basic abstractions point, line, and region

The two most important instances of spatially related collections of objects are partitions (of the plane) and networks. **A partition** can be viewed as a set of region objects that are required to be disjoint. The adjacency relationship is of particular interest, that is, there exist often pairs of region objects with a common boundary. Partitions can be used to represent thematic maps. A network can be viewed as a graph embedded into the plane, consisting of a set of point objects, forming its nodes, and a set of line objects describing the geometry of the edges. **Networks** are ubiquitous in geography, for example, highways, rivers, public transport, or power supply lines.

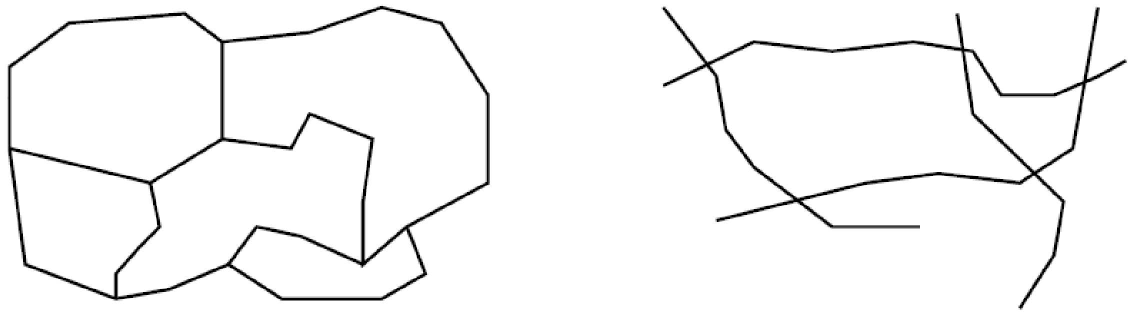


Figure: Partitions and networks

### *Spatial Data Types*

Systems of spatial data types, or spatial algebras, can capture the fundamental abstractions for point, line and region described above together with relationships between them and operations for composition (e.g. forming the intersection of regions). We have stated in Section 1 that they are a mandatory part of the data model for a spatial DBMS, so that indeed, all proposals for models and query languages as well as prototype systems (see Section 5) offer them in some form.

### *Spatial Data Model:*

Data Model - An abstraction of the real world which incorporates only those properties thought to be relevant to the application at hand, define specific groups of entities, and their attributes and the relationships between these entities. A data model is independent of a computer system [Association for Geographic Information].Any time you wish to deal with geographic data, you must choose a geographic data model by which to do it. The choice of data model will yield benefits in terms of simplifying real-world features enough to deal with them easily, but will also incur costs in terms of oversimplifying or misrepresenting different aspects. FOLDOC definition of data model

A paper map is an example of an analog data model -- it is a formalized framework that cartographers use to capture and represent information on a sheet of paper. The same sort of thing is also needed to capture and represent geographic information when the medium is digital rather than ink-and-paper. In a GIS, abstractions of real-world features must therefore be formalized into a data model that defines how the computer will represent and manage the geographic information (geometry and attributes).

Bernhardsen (1999) diagrams the data model formalization process along these lines:

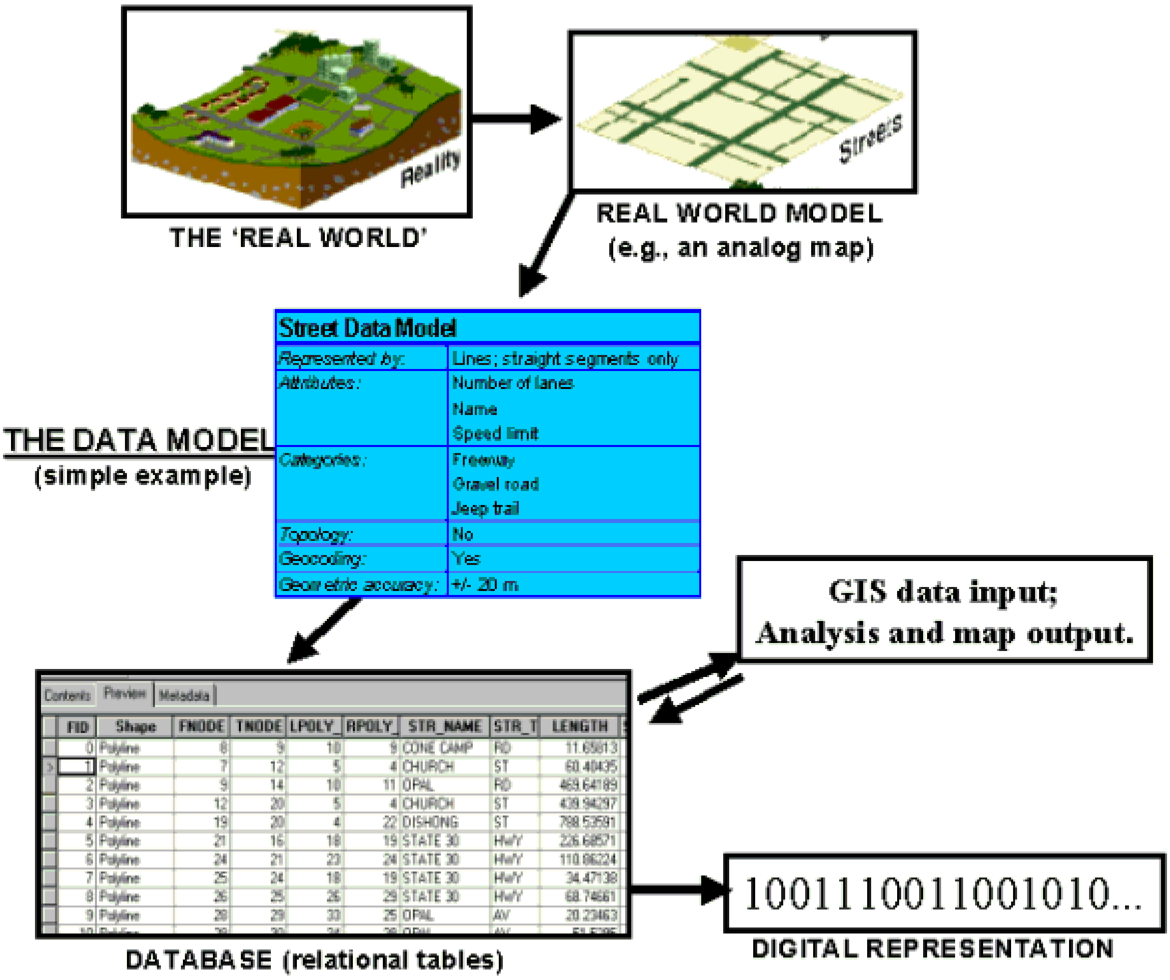



**Figure 1:** The modeling process.(after Bernhardsen 1999, p.39. Map graphics from [www.gis.com](www.gis.com))

Most of the confusion about data models arises from their diversity. Some data models are more abstract/theoretical while others are made with specific database types in mind. For example, the vector data model and the raster data model are very general, whereas the georelational data model and geodatabase data model are made to fit specific categories of database software. Furthermore, a given data model may belong to more than one category: a coverage is both a vector data model (general) and a georelational data model (database specific).

The many types of data models are easier to think about if one pictures of them as being part of a general hierarchy. Below is a figure showing the hierarchy of ArcGIS's data models:

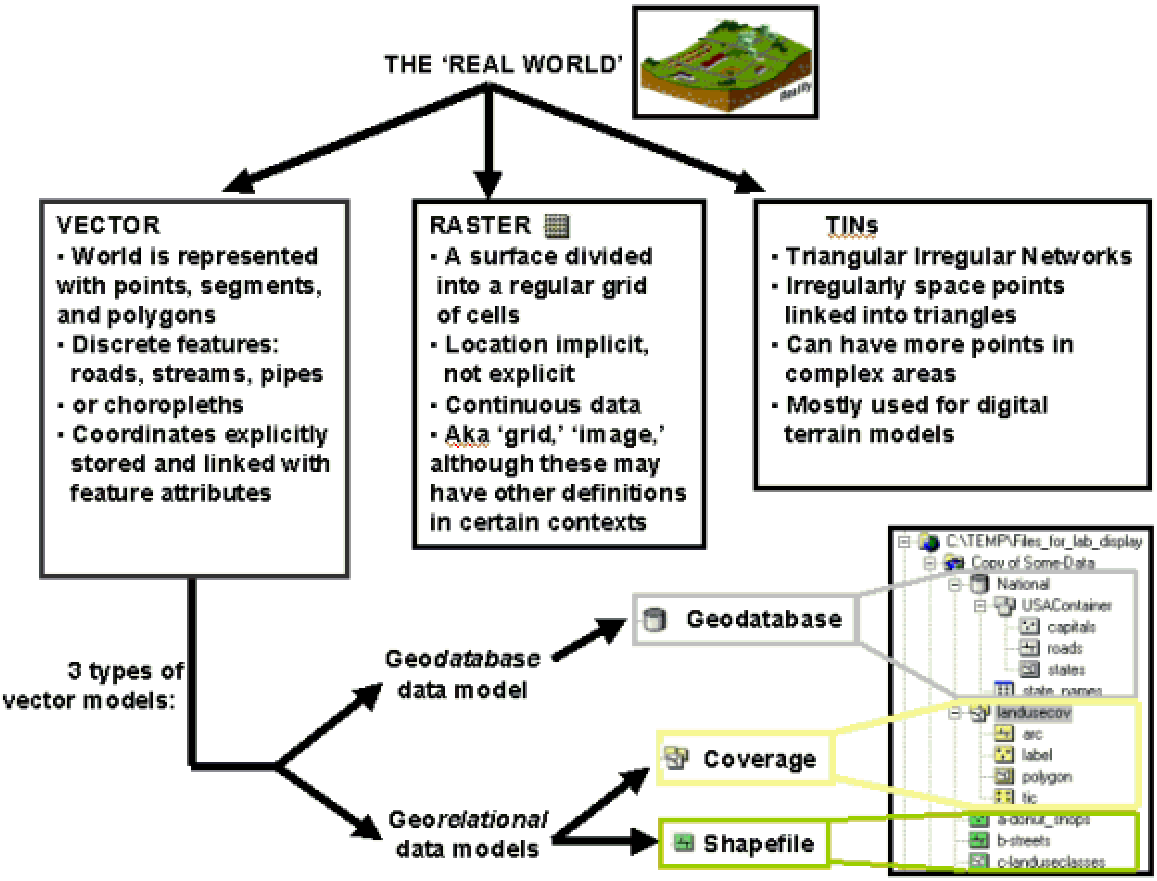



**Figure 2**: Hierarchy of ESRI's ArcGIS data models.

The data models go from most general at the top level (vector, raster, TIN) to most specific at the bottom level (shapefile, coverage, geodatabase). It is important to note that a geodatabase can handle all three general models, not just the vector model. Geographic data models have evolved under the influences of technology (e.g., increasing storage space and processing power, networking, or software evolution) and even history (e.g., ESRI introduced the "coverage" data model in 1980).

Every GIS software package will be capable of supporting a number of data models, but will also have it's own proprietary format (that none of the others will read). The capabilities of the data models may change with new versions of the software, and compatibility issues may arise between different GIS software, and even between different versions of the same software. Certain functions will be accessible using data in the form of one data model but not another.

### 2.1 - Data Structures vs. Data Models

The specific format with which the data are stored on the computer is known as the **data structure**. To illustrate, consider a basic vector data model. The vector model represents features as consisting of lines which individually link together a start node, vertices in between, and an end node. To draw and analyze features represented this way, the computer needs information on the locations of each node and vertex of the lines. This could be provided in the form of a table listing the coordinates of these points, and indicating which line(s) go through them. This table would be the basic **data structure**. Coverages and shapefiles use this type of structure.

In Figure 1 above, the lower left box titled "DATABASE (relational tables)" represents the data structure. In it you can see numbered rows and columns with labels, this is the 'structure' of the data. Some columns have only numbers, some have only text and some have both.

Several different types of data structures can potentially be used to represent the same data model. For example, you could represent a vector data model using coverages, shapefiles, or geodatabases. Although these all take the same basic approach in representing the model, there are still significant differences between them; 1) data models do not necessarily imply any particular data structures; and 2) data structures can represent the same data model while still being very different from one another.

### *Spatial Relationships*

Among the operations offered by spatial algebras, spatial relationships are the most important ones. For example, they make it possible to ask for all objects in a given relationship with a query object, e.g. all objects within a window.

- *Topological relationships*, such as adjacent, inside, disjoint, are invariant under topological transformations like translation, scaling, and rotation.
- *Direction relationships*, for example, above, below, or north_of, southwest_of, etc.
- *Metric relationships*, e.g. "distance < 100".

### Integrating Geometry into the DBMS Data Model

The central idea for integrating geometric modeling into a DBMS data model is to represent "spatial objects" (in the sense of application objects such as river, country, city, etc.) by objects (in the sense of the DBMS data model) with at least one attribute of a spatial data type. Hence the DBMS data model must be extended by SDTs at the level of atomic data types (such as integer, string, etc.), or better be generally open for user-defined types.

### Querying

From one point of view, the problem of querying is to connect the operations of a spatial algebra (including predicates to express spatial relationships) to the facilities of a DBMS query language. But there are also other aspects that have mainly to do with the fact that spatial data require a graphical presentation of results as well as graphical input of queries or at least SDT values used in queries. The fundamental operations needed at the level of manipulating sets of database objects, graphical input and output, and techniques and requirements for extending query languages.

### *Graphical Input and Output*

Traditional database systems deal with alphanumeric data types whose values can easily be entered through a keybord and represented textually within a query result (e.g. a table). For a spatial database system, at least when it is to be used interactively, graphical presentation of SDT values in query results is essential, and entering SDT values to be used as "constants" in queries via a graphical input device is also important. Besides graphical representation of SDT values, another distinctive characteristic of querying a spatial database is that the goal of querying is in general to obtain a "tailored" picture of the space represented in the database, which means that the information to be retrieved is often not the result of a single query but rather a combination of several queries. For example, for GIS applications, the user wants to see a map built by overlaying graphically the results of several queries.

Requirements for spatial querying is given in the following list:

(1) Spatial data types.
(2) Graphical display of query results.

(3) Graphical combination (overlay) of several query results. It should be possible to start a new picture, to add a layer, or to remove a layer from the current display. (Some systems also allow to change the order of layers.
(4) Display of context. To interpret the result of a query, e.g. a point describing the location of a city, it is necessary to show some background, such as the boundary of a state containing it. A raster image of the area can also nicely serve as a background.
(5) A facility for checking the content of a display. When a picture (a map) has been composed by several queries, one should be able to check which queries have built it.
(6) Extended dialog. It should be possible to use pointing devices to select objects within a picture or subareas (zooming in), e.g. by dragging a rectangle over the picture.
(7) Varying graphical representations. It should be possible to assign different graphical representations (colors, patterns, intensity, symbols) to different object classes in a picture, or even to distinguish objects within one class (e.g. use different symbols to distinguish cities by population).
(8) A legend should explain the assignment of graphical representations to object classes.
(9) Label placement. It should be possible to select object attributes to be used as labels within a graphical representation.
(10) Scale selection. At least for GIS applications, selecting subareas should be based on commonly used map scales. The scale determines not only the size of the graphical representation, but possibly also what kind of symbol is used or whether an object is shown at all (cartographic generalization).
(11) Subarea for queries. It should be possible to restrict attention to a particular area of the space for several following queries.

These requirements can in general be fulfilled by offering textual commands in the query language or within the design of a graphical user interface (GUI). A GUI will probably have at least three subwindows:

(1) A text window for displaying the textual representation of a collection of objects, containing for each object its alphanumeric attributes,
(2) A graphics window containing the overlay of the graphical representations of spatial attributes of several object classes or query results, and
(3) A text window for entering queries and perhaps displaying system messages.

### *Tools for Spatial DBMS Implementation: Data Structures and Algorithms*

We now consider system implementation bottom-up. In this section we first describe data structures and algorithms that can be used as tools or building blocks within different system architectures. System architectures themselves are discussed in the next section. The general problem to be solved is implementation of a spatial algebra in such a way that it can be integrated into a database system's query processing. This means, first of all, that we have to provide representations for the algebra's types as well as algorithms/procedures for its operations. However, it does not suffice just to implement atomic operations efficiently such as a test whether two regions intersect. It is also necessary to consider the use of such predicates within set-oriented query processing, that is, when they occur within a spatial selection or a spatial join. Here spatial access methods and spatial join algorithms come into play. Last not least, other set operations of a spatial algebra need their special implementations. In the following subsections we discuss representation of spatial data types and implementation of atomic operations, spatial indexing to support spatial selection, and support of spatial join.

### Spatial Indexing – Supporting Spatial Selection

The main purpose of spatial indexing is to support spatial selection, that is, to retrieve from a large set of spatial objects (objects with an SDT attribute) those in some particular relationship with a query SDT value. A spatial indexing method organizes space and the objects in it in some way so that only parts of the space and a subset of the objects need to be considered to answer such a query. There are two ways to provide spatial indexing: (i) dedicated external spatial data structures are added to the system, offering for spatial attributes what e.g. a B-tree does for standard attributes, and (ii) spatial objects are mapped into a one-dimensional space so that they can be stored within a standard onedimensional index such as a B-tree. Apart from spatial selection, spatial indexing supports also other operations such as spatial join, finding the object closest to a query value, etc.

### System Architecture

At the level of system architecture, the problem is to integrate the tools described in Section 4 for the support of spatial data types – and even more than that. In principle, the following extensions to a standard architecture need to be accomodated:

- representations for the data types of a spatial algebra,
- procedures for the atomic operations,
- spatial index structures,
- access operations for spatial indices,
- filter and refine techniques,
- spatial join algorithms,
- cost functions for all these operations,
- statistics for estimating selectivity of spatial selection and spatial join,
- extensions of the optimizer to map queries into the specialized query processing methods,
- spatial data types and operations within data definition and query language,
- user interface extensions to handle graphical representation and input of SDT values.

In our view, the only clean way to accomodate these extensions is an integrated architecture based on the use of an extensible DBMS. Nevertheless, GIS have been constructed before extensible DBMS technology was available, and we shall first review previous approaches to GIS architecture.

### How maps convey geographic information

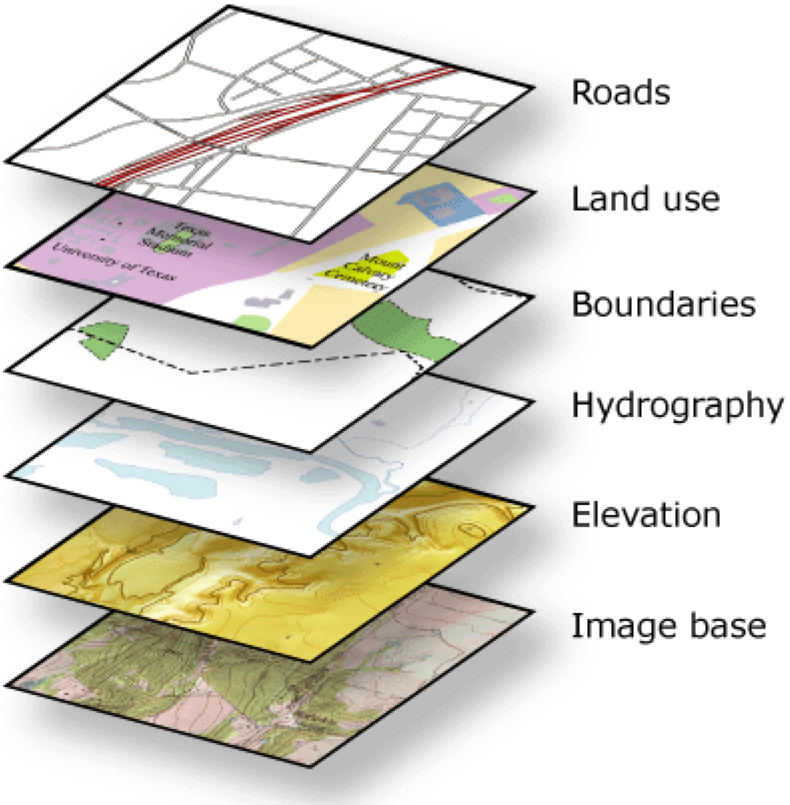
Fundamental GIS concepts are closely linked to maps and their contents. In fact, map concepts form the basis for understanding GIS more fully. This topic explores some fundamental map concepts and describes how they are applied and used within GIS.

### Maps

A map is a collection of map elements laid out and organized on a page. Common map elements include the map frame with map layers, a scale bar, north arrow, title, descriptive text, and a symbol legend.

The primary map element is the map frame, and it provides the principal display of geographic information. Within the map frame, geographical entities are presented as a series of map layers that cover a given map extent—for example, map layers such as roads, rivers, place names, buildings, political boundaries, surface elevation, and satellite imagery.

The following graphic illustrates how geographical elements are portrayed in maps through a series of map layers. Map symbols and text are used to describe the individual geographic elements.



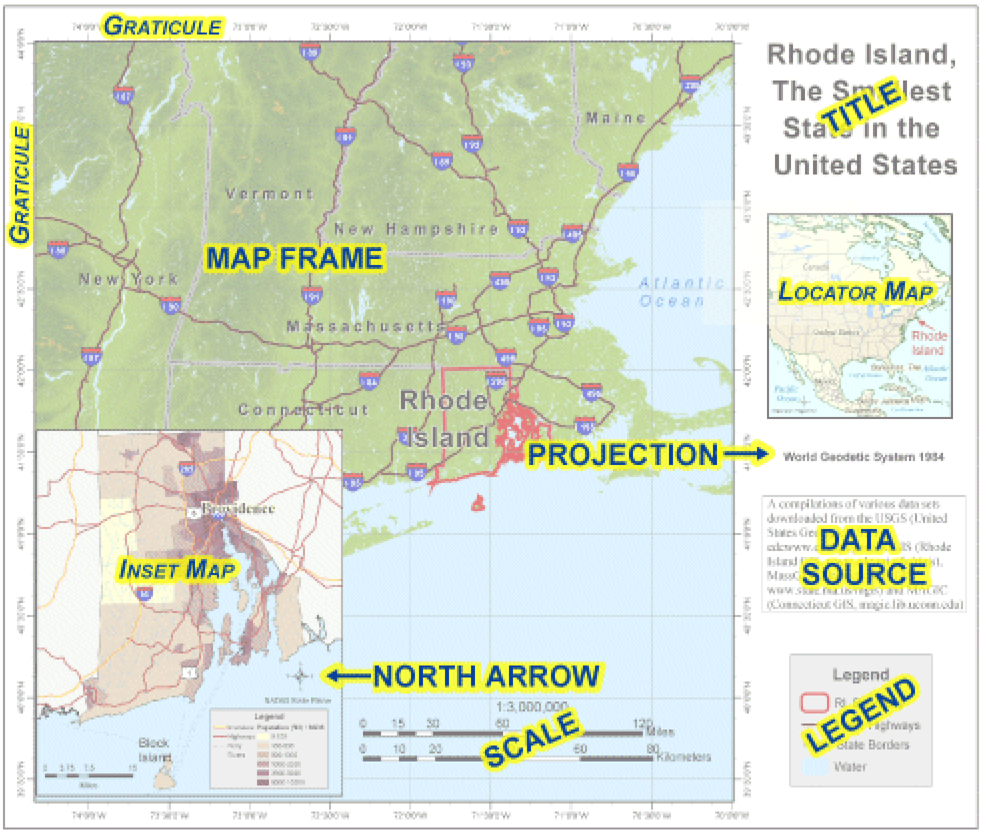

Map layers are thematic representations of geographic information, such as transportation, water, and elevation. Map layers help convey information through:

- Discrete features such as collections of points, lines, and polygons
- Map symbols, colors, and labels that help to describe the objects in the map
- Aerial photography or satellite imagery that covers the map extent
- Continuous surfaces such as elevation which can be represented in a number of ways—for example, as a collection of contour lines and elevation points or as shaded relief

### *Map Layout and composition*

- Along with the map frame, a map presents an integrated series of map elements laid out and arranged on a page. Common map elements include a north arrow, a scale bar, a symbol legend, and other graphical elements. These elements aid in map reading and interpretation.

The map layout below illustrates how map elements are arranged on a page.




Often, maps include additional elements such as graphs, charts, pictures, and text that help to communicate additional critical information.

### *Spatial relationships in a map*

Maps help convey geographic relationships that can be interpreted and analyzed by map readers. Relationships that are based on location are referred to as spatial relationships. Here are some examples.

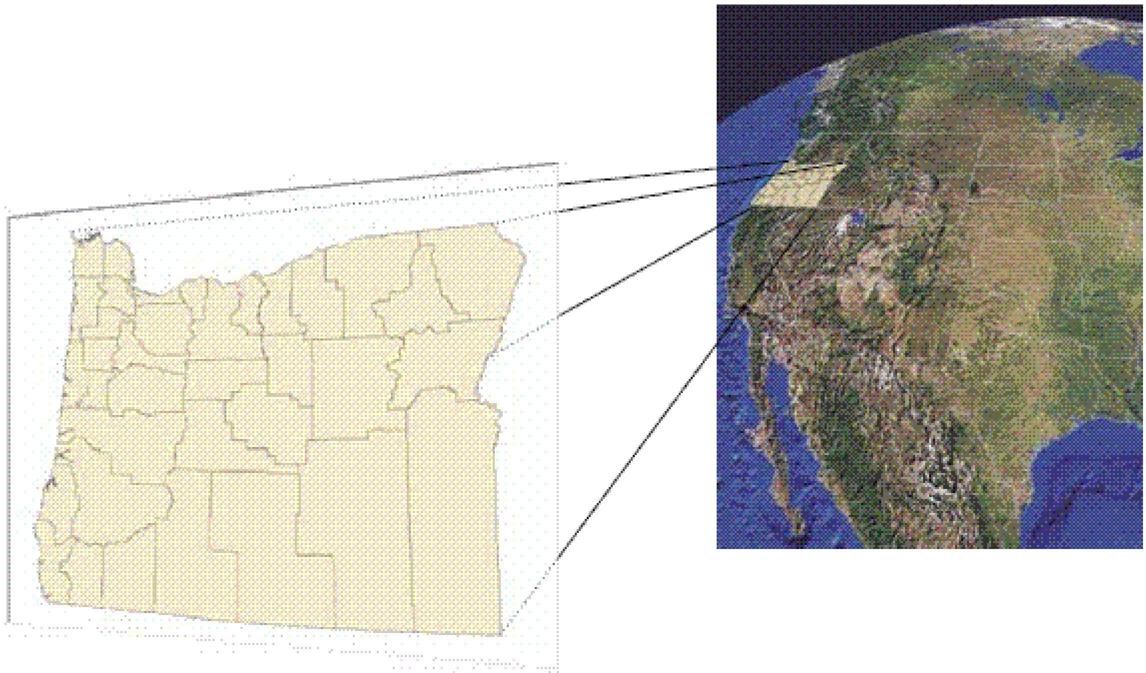
- Which geographic features *connect* to others (for example, Water Street connects with 18th Ave.)
- Which geographic features are ***adjacent*** (contiguous) to others (for example, The city park is adjacent to the university.)
- Which geographic features are ***contained within*** an area (for example, The building footprints are contained within the parcel boundary.)
- Which geographic features ***overlap*** (for example, The railway crosses the freeway.)
- Which geographic features are *near* others (proximity) (for example, The Courthouse is near the State Capitol.)
- The feature geometry *is equal to* another feature (for example, The city park is equal to the historic site polygon).
- The *difference* in elevation of geographic features (for example, The State Capitol is uphill from the water.)
- The feature is *along* another feature (for example, The bus route follows along the street network.).

Within a map, such relationships are not explicitly represented. Instead, as the map reader, you interpret relationships and derive information from the relative position and shape of the map elements, such as the streets, contours, buildings, lakes, railways, and other features. In a GIS, such relationships can be modeled by applying rich data types and behaviors (for example, topologies and networks) and by applying a comprehensive set of spatial operators to the geographic objects (such as buffer and polygon overlay).

### Georeferencing and coordinate systems

Georeferencing: Assigning map coordinates and spatial location

All the elements in a map layer have a specific geographic location and extent that enables them to be located on or near the earth's surface. The ability to accurately describe geographic locations is critical in both mapping and GIS. This process is called georeferencing.



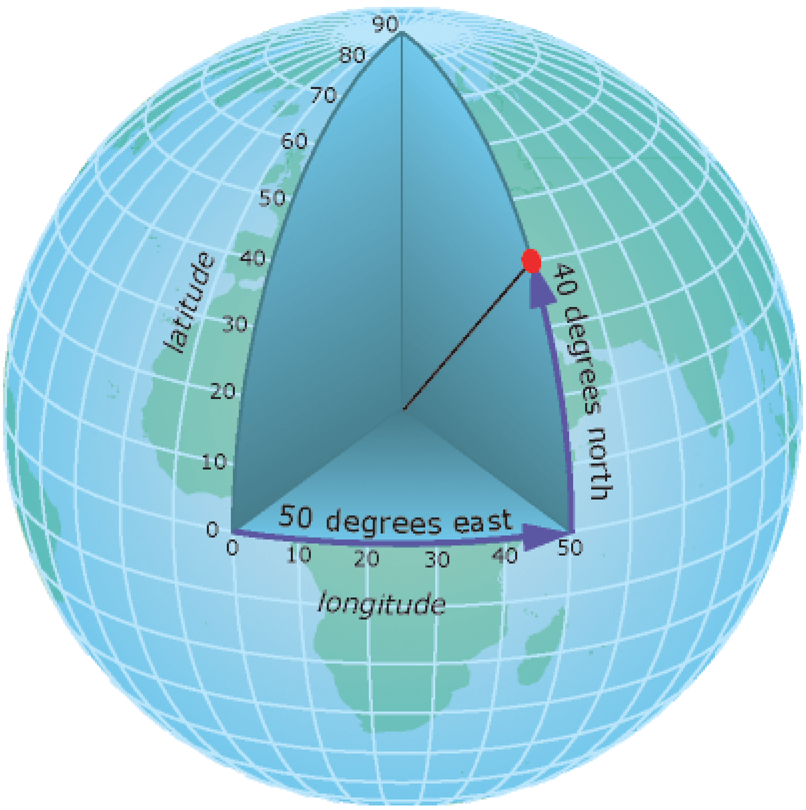
Describing the correct location and shape of features requires a framework for defining real-world locations. A geographic coordinate system is used to assign geographic locations to objects. A global coordinate system of latitude-longitude is one such framework. Another is a planar or Cartesian coordinate system derived from the global framework.

Maps represent locations on the earth's surface using grids, graticules, and tic marks labeled with various ground locations (both in measures of latitude-longitude and in projected coordinate systems (such as UTM meters). The geographic elements contained in various map layers are drawn in a specific order (on top of one another) for the given map extent.

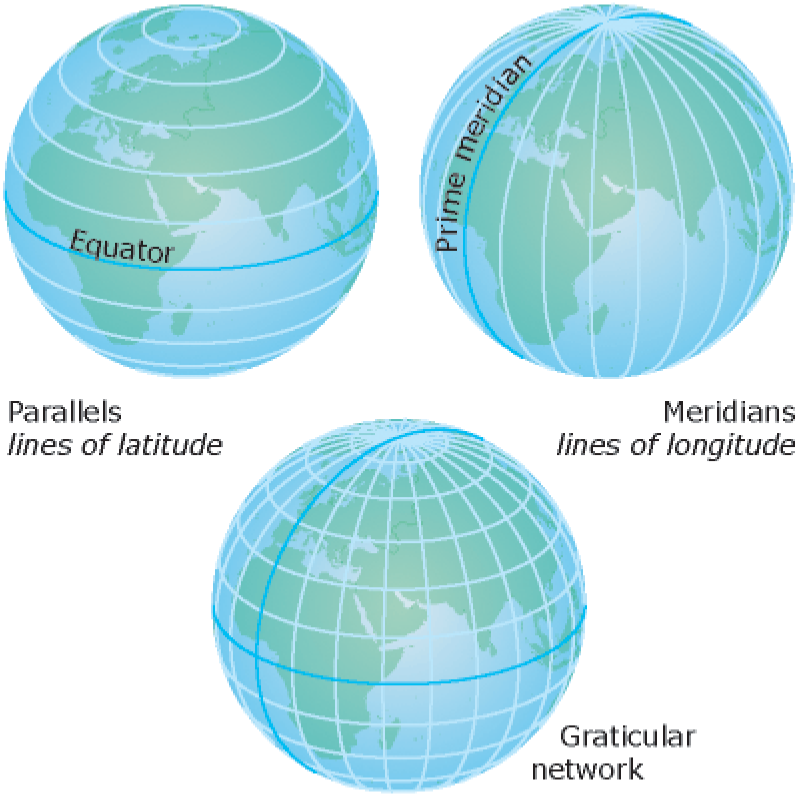
GIS datasets contain coordinate locations within a global or Cartesian coordinate system to record geographic locations and shapes.

### *Latitude and longitude*

One method for describing the position of a geographic location on the earth's surface is using spherical measures of latitude and longitude. They are measures of the angles (in degrees) from the center of the earth to a point on the earth's surface. This reference system is often referred to as a geographic coordinate system.




Latitude angles are measured in a north-south direction. The equator is at an angle of 0. Often, the northern hemisphere has positive measures of latitude and the southern hemisphere has negative measures of latitude. Longitude measures angles in an east-west direction. Longitude measures are traditionally based on the Prime Meridian, which is an imaginary line running from the North Pole through Greenwich, England to the South Pole. This angle is Longitude 0. West of the Prime Meridian is often recorded as negative Longitude and east is recorded as positive. For example, the location of Los Angeles, California is roughly Latitude "plus 33 degrees, 56 minutes" and Longitude "minus 118 degrees, 24 minutes."




Although longitude and latitude can locate exact positions on the surface of the globe, they are not uniform units of measure. Only along the equator does the distance represented by one degree of longitude approximate the distance represented by one degree of latitude. This is because the equator is the only parallel as large as a meridian. (Circles with the same radius as the spherical earth are called great circles. The equator and all meridians are great circles.)

Above and below the equator, the circles defining the parallels of latitude get gradually smaller until they become a single point at the North and South Poles where the meridians converge. As the meridians converge toward the poles, the distance represented by one degree of longitude decreases to zero. On the Clarke 1866 spheroid, one degree of longitude at the equator equals 111.321 km, while at 60° latitude, it is only 55.802 km. Since degrees of latitude and longitude don't have a standard length, you can't measure distances or areas accurately or display the data easily on a flat map or computer screen. Performing GIS analysis and mapping applications requires a more stable coordinate framework, which is provided by projected coordinate systems.

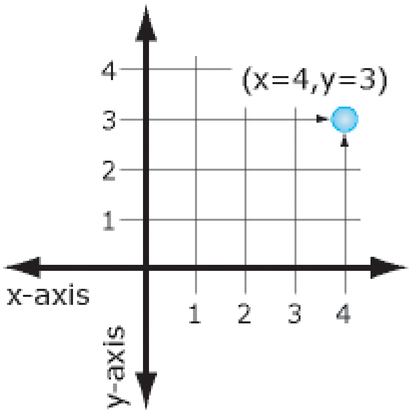
### *Map projections using Cartesian coordinates*

Projected coordinate systems are any coordinate system designed for a flat surface, such as a printed map or a computer screen.

2D and 3D Cartesian coordinate systems provide the mechanism for describing the geographic location and shape of features using x and y values (and, as you will read later, by using columns and rows in rasters).

The Cartesian coordinate system uses two axes: one horizontal (x), representing east-west, and one vertical (y), representing north-south. The point at which the axes intersect is called the origin. Locations of geographic objects are defined relative to the origin, using the notation (x,y), where x refers to the distance along the horizontal axis, and y refers to the distance along the vertical axis. The origin is defined as (0,0).

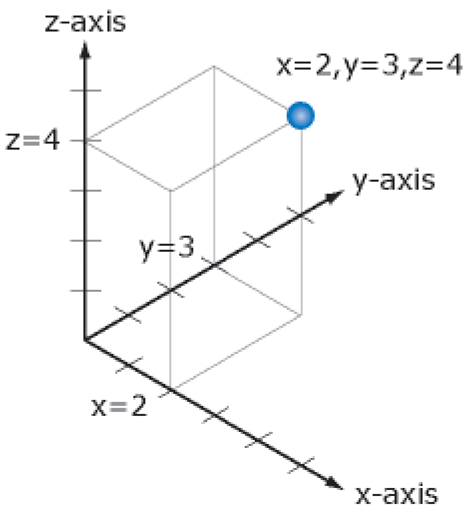
In the illustration below, the notation (4, 3) records a point that is four units over in x and three units up in y from the origin.




### 3D coordinate systems

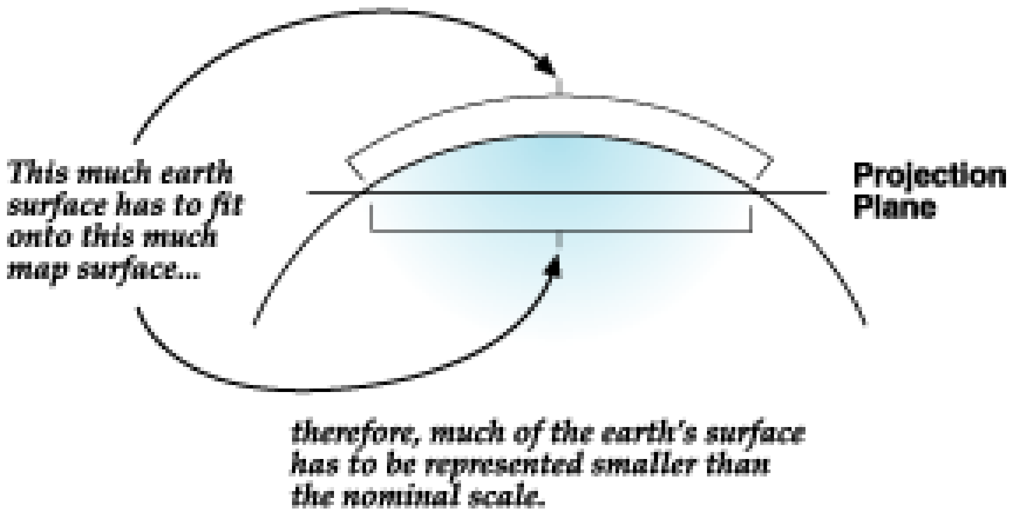
Increasingly, projected coordinate systems also use a Z value to measure elevation above or below mean sea level.

In the illustration below, the notation (2, 3, 4) records a point that is two units over in x and three units in y from the origin and whose elevation is 4 units above the earth's surface (such as 4 meters above mean sea level).



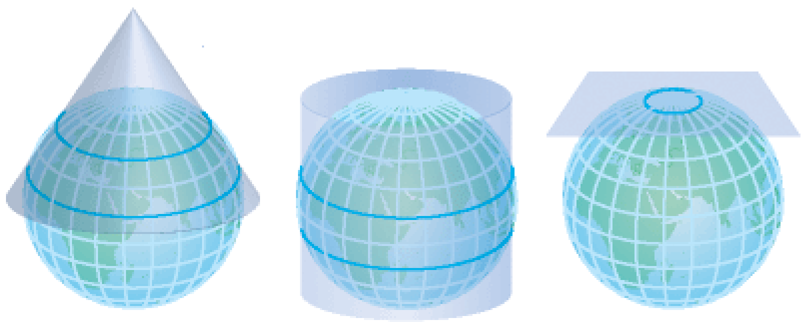

### Properties and distortion in map projections

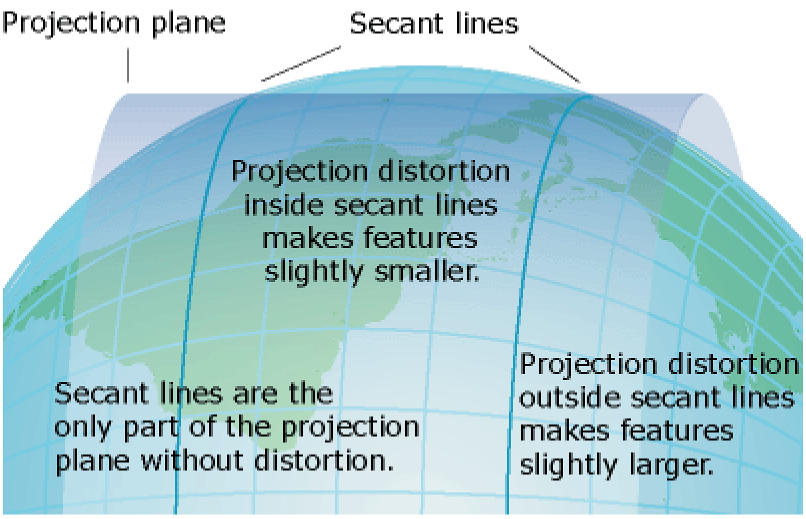
Since the earth is spherical, a challenge faced by cartographers and GIS professionals is how to represent the real world using a flat or planar coordinate system. To understand their dilemma, consider how you would flatten half of a basketball; it can't be done without distorting its shape or creating areas of discontinuity. The process of flattening the earth is called projection, hence the term map projection.




A projected coordinate system is defined on a flat, two dimensional surface. Projected coordinates can be defined for both 2D (x,y) and 3D (x,y,z) in which the x,y measurements represent the location on the earth's surface and z would represent height above or below mean sea level.

Below are some examples of various methods for deriving planar map projections.






Unlike a geographic coordinate system, a projected coordinate system has constant lengths, angles, and areas across the two dimensions. However, all map projections representing the earth's surface as a flat map, create distortions in some aspect of distance, area, shape, or direction.

Users cope with these limitations by using map projections that fit their intended uses, geographic location, and extent. GIS software also can transform information between coordinate systems to support integration and critical workflows.

Many map projections are designed for specific purposes. One map projection might be used for preserving shape while another might be used for preserving the area (conformal versus equal area).

These properties—the map projection (along with Spheroid and Datum), become important parameters in the definition of the coordinate system for each GIS dataset and each map. By recording detailed descriptions of these properties for each GIS dataset, computers can re-project and transform the geographic locations of dataset elements on the fly into any appropriate coordinate system. As a result, it's possible to integrate and combine information from multiple GIS layers. This is a fundamental GIS capability. Accurate location forms the basis for almost all GIS operations.

### *Elements of geographic information*

There are some universal principles that provide the foundation for how GIS systems represent, operate on, manage, and share geographic information. The purpose of this topic is to provide you with a solid foundation for understanding these key concepts and how ArcGIS employs them.

Like a map, a GIS is layer-based. And like the layers in a map, GIS datasets represent collections of individual features with their geographic locations and shapes as well as with descriptive information stored as attributes.

There are four fundamental types of geographic representations:

- Features (collections or points, lines, and polygons)
- Attributes
- Imagery
- Continuous surfaces (such as elevation)

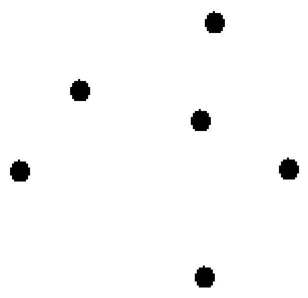
All of the rich GIS behavior for representing and managing geographic information is based on these fundamental types.

**Features - Points, lines, and polygons**

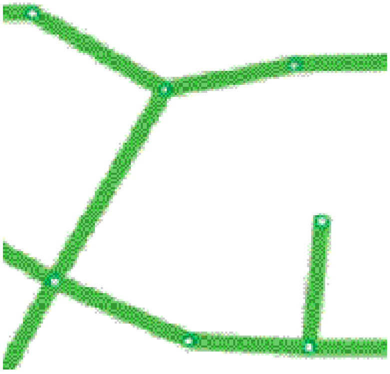
Geographic features are representations of things located on or near the surface of the earth. Geographic features can occur naturally (such as rivers and vegetation), can be constructions (such as roads, pipelines, wells, and buildings), and can be subdivisions of land (such as counties, political divisions, and land parcels).

Although there are a number of additional types, geographic features are most commonly represented as points, lines, and polygons.

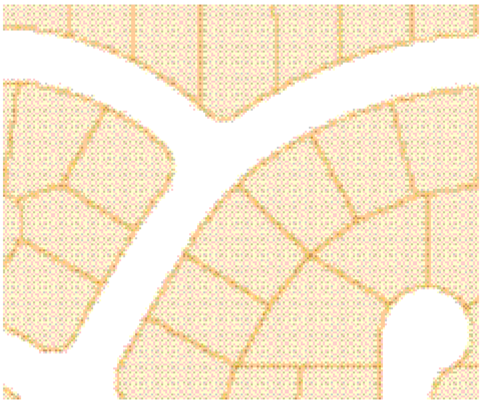
Points define discrete locations of geographic features too small to be depicted as lines or areas, such as well locations, telephone poles, and stream gauges. Points can also represent locations such as address locations, GPS coordinates, or mountain peaks.



Lines represent the shape and location of geographic objects too narrow to depict as areas (such as street centerlines and streams). Lines are also used to represent features that have length but no area such as contour lines and administrative boundaries. (Contours are interesting, as you'll read later on, because they provide one of a number of alternatives for representing continuous surfaces.)



Polygons are enclosed areas (many-sided figures) that represent the shape and location of homogeneous features such as states, counties, parcels, soil types, and land use zones. In the example below, the polygons represent Parcels.



## Attributes

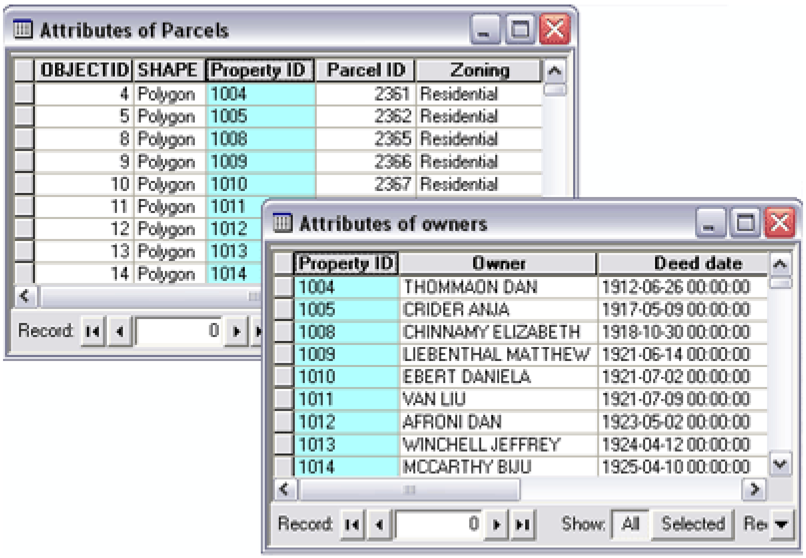
Maps convey descriptive information through map symbols, colors, and labels. Here are some typical examples:

- Roads are displayed based on their road class (for example, line symbols representing divided highways, main streets, residential streets, unpaved roads, and trails).
- Streams and water bodies are drawn in blue to indicate water.
- City streets are labeled with their name and often some address range information.
- Special point and line symbols denote specific features such as rail lines, airports, schools, hospitals, and special facilities.

In a GIS, descriptive attributes are managed in tables, which are based on a series of simple, essential relational database concepts. A relational database provides a simple, universal data model for storing and working with attribute information. DBMSs are inherently open because their simplicity and flexibility enables support for a broad range of applications. Key relational concepts include:

- Descriptive data is organized into tables.
- Tables contain rows.
- All rows in a table have the same columns.
- Each column has a type, such as integer, decimal number, character, date, and so on.
- A series of relational functions and operators (SQL) is available to operate on the tables and their data elements.

The illustration below shows two tables and how their records can be related to one another using a common field. In the example, the parcels feature class table is linked to the owners table through the common Property ID field.



Attributes of Parcels

| OBJECTID | SHAPE | Property ID | Parcel ID | Zoning |
|---|---|---|---|---|
| 4 | Polygon | 1004 | 2361 | Residential |
| 5 | Polygon | 1005 | 2362 | Residential |
| 8 | Polygon | 1008 | 2365 | Residential |
| 9 | Polygon | 1009 | 2366 | Residential |
| 10 | Polygon | 1010 | 2367 | Residential |
| 11 | Polygon | 1011 | | |
| 12 | Polygon | 1012 | | |
| 13 | Polygon | 1013 | | |
| 14 | Polygon | 1014 | | |

Record 0

Attributes of owners

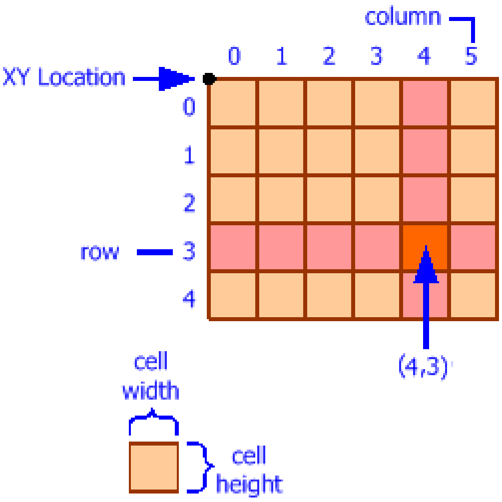
| Property ID | Owner | Deed date |
|---|---|---|
| 1004 | THOMMAON DAN | 1912-06-26 00:00:00 |
| 1005 | CRIDER ANJA | 1917-05-09 00:00:00 |
| 1008 | CHINNAMY ELIZABETH | 1918-10-30 00:00:00 |
| 1009 | LIEBENTHAL MATTHEW | 1921-06-14 00:00:00 |
| 1010 | EBERT DANIELA | 1921-07-02 00:00:00 |
| 1011 | VAN LIU | 1921-07-09 00:00:00 |
| 1012 | AFRONI DAN | 1923-05-02 00:00:00 |
| 1013 | WINCHELL JEFFREY | 1924-04-12 00:00:00 |
| 1014 | MCCARTHY BIJU | 1925-04-10 00:00:00 |

Record 0 Show: All Selected Re

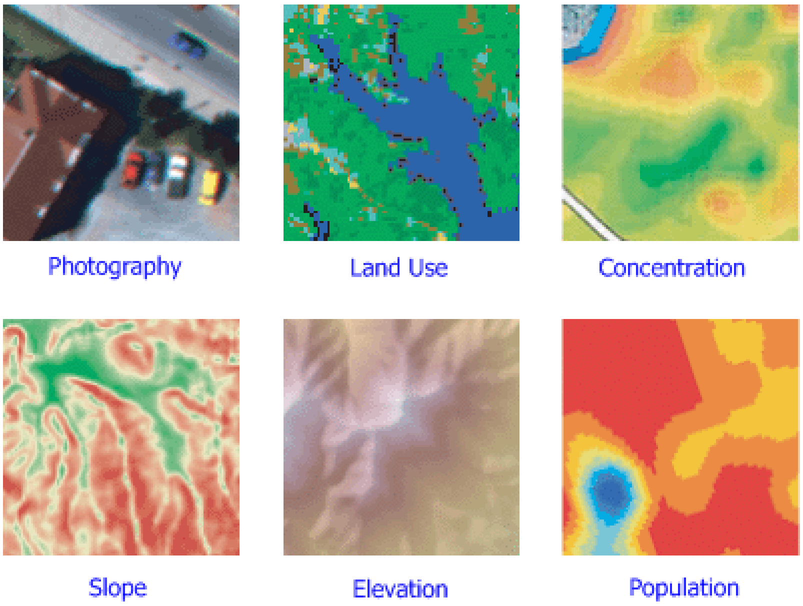
## Imagery

Aerial imagery is a raster data structure obtained from various sensors carried in satellites and aircraft. Imagery is managed as a raster data type composed of cells organized in a grid of rows and columns. In addition to the map projection, the coordinate system for a raster dataset includes its cell size and a reference coordinate (usually the upper left or lower left corner of the grid).

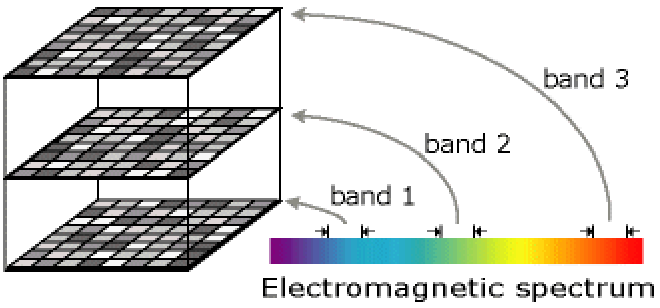
These properties enable a raster dataset to be described by a series of cell values starting in the upper left row. Each cell location can be automatically located using the reference coordinate, the cell size, and the number of rows and columns.




Typical image sources include cameras capable of capturing aerial photographs that can be georeferenced and corrected to ground locations (such as digital ortho photography).




Imagery is also used to collect data in both the visible and non-visible portions of the electromagnetic spectrum. One system is the multispectral scanner carried in LANDSAT satellites that records imagery in seven *bands* (or ranges) along the electromagnetic spectrum. The measures for each band are recorded in a separate grid. The stack of seven grids makes up a multiband image.

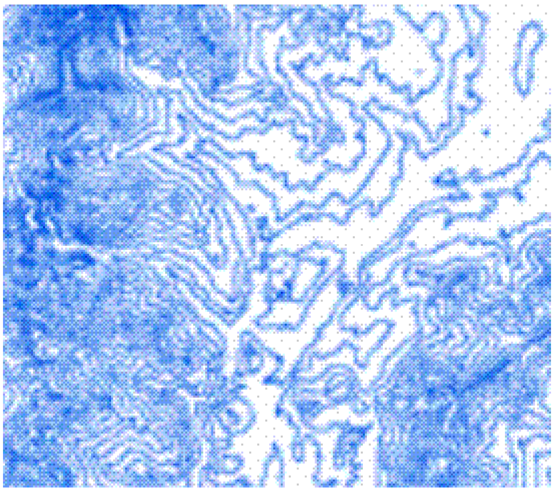



### Surfaces

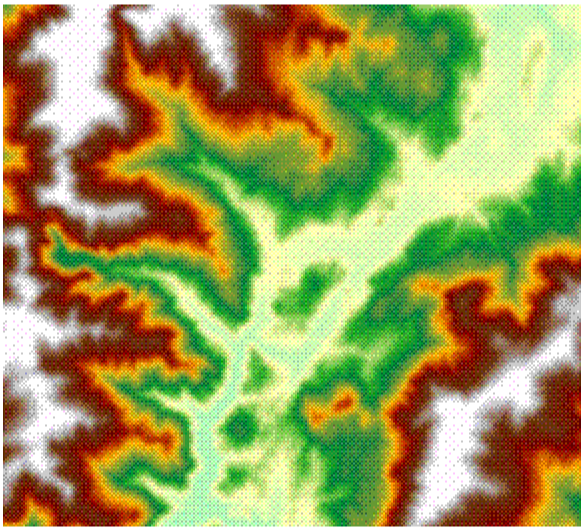
A surface describes an occurrence that has a value for every point on the earth. For example, surface elevation is a continuous layer of values for ground elevation above mean sea level for the entire extent of the dataset. Other surface type examples include rainfall, pollution concentration, and sub-surface representations of geological formations.

Surface representation is somewhat challenging. With continuous datasets, it is impossible to represent all values for all locations. Various alternatives exist for representing surfaces using either features or rasters. Here are some example alternatives for surface representation:

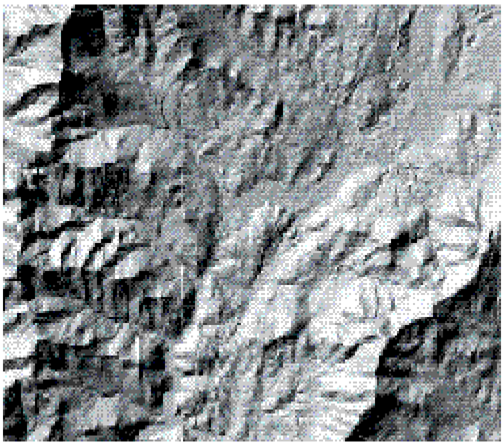
Contour lines—Isolines represent locations having an equal value, such as elevation contours.



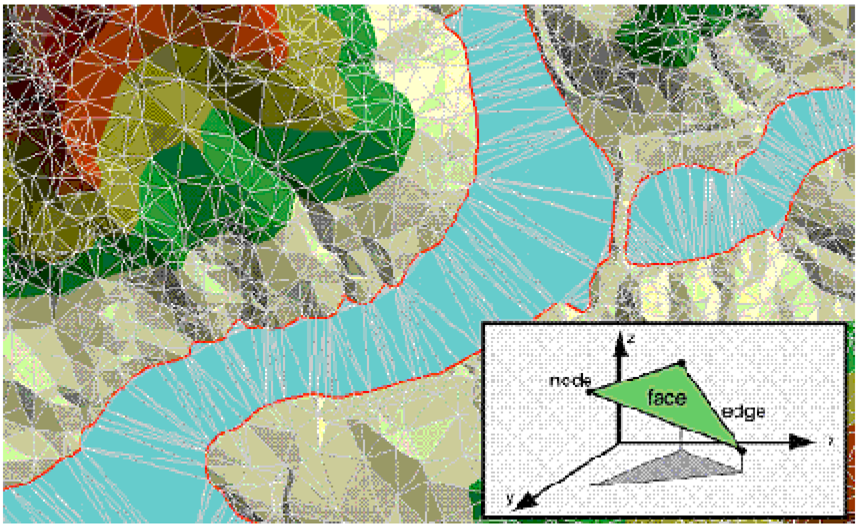
Contour bands—The areas where the surface value is within a specified range, such as bands of average annual rainfall between 25 CM and 50 CM per year.



Raster datasets—A matrix of cells where each cell value represents a measure of the continuous variable. For example, Digital Elevation Models (DEMs) are frequently used to represent surface elevation.



TIN layers—A Triangulated Irregular Network (TIN) is a data structure for representing surfaces as a connected network of triangles. Each triangle node has an XY coordinate and a Z or surface value.




The raster and TIN representations can be used to estimate the surface value for any location using interpolation.

### *How GIS represents and organizes geographic information*

Managing features, rasters, attributes, and surfaces in GIS

These four types of geographic information (features, rasters, attributes, and surfaces) are actually managed using three primary GIS data structures:

- Feature classes
- Attribute tables
- Raster datasets

Each of these primary datasets can be extended with additional capabilities to maintain data integrity (for example, using topology), to model geographic relationships (such as network connectivity and flow), and to add advanced behaviors (for example, using TINs).

**How the four types of map layers are represented in GIS**

| Map Layer Types | GIS Datasets |
|---|---|
| Features—points, lines, and polygons | Feature classes |
| Attributes | Tables |
| Imagery | Raster datasets |
| Surfaces | Both features and rasters can be used to provide a number of alternative surface representations:<br>• Feature classes (such as contours)<br>• Raster-based elevation datasets |

- TINs built from XYZ points and 3D line feature classes

A GIS has a collection of datasets

Typically, a GIS is used for handling several different datasets where each holds data about a particular feature collection (for example, roads) that is geographically referenced to the earth's surface.

A GIS database design is based upon a series of data themes, each having a specified geographic representation. For example, individual geographic entities can be represented as features (such as points, lines, and polygons); as imagery using rasters; as surfaces using features, rasters, or TINs; and as descriptive attributes.

In a GIS, homogeneous collections of geographic objects are organized into data themes such as parcels, wells, buildings, ortho imagery, and raster-based digital elevation models (DEMs). Precisely and simply defined geographic datasets are critical for useful geographic information systems, and the layer-based concept of data themes is a critical GIS concept.

**GIS datasets are collections of geographic representations**

A dataset is a collection of homogeneous features. Geographic representations are organized in a series of datasets or layers. Most datasets are collections of simple geographic elements such as a road network, a collection of parcel boundaries, soil types, an elevation surface, satellite imagery for a certain date, well locations, and so on.

In a GIS, spatial data collections are typically organized as feature class datasets or raster-based datasets.
Many data themes are best represented by a single dataset such as for soil types or well locations. Other themes, such as a transportation framework, are represented by multiple datasets (such as a separate feature class each for streets, intersections, bridges, highway ramps, railroads, and so on).

Raster datasets are used to represent georeferenced imagery as well as continuous surfaces such as elevation, slope, and aspect.

**Common GIS representations**

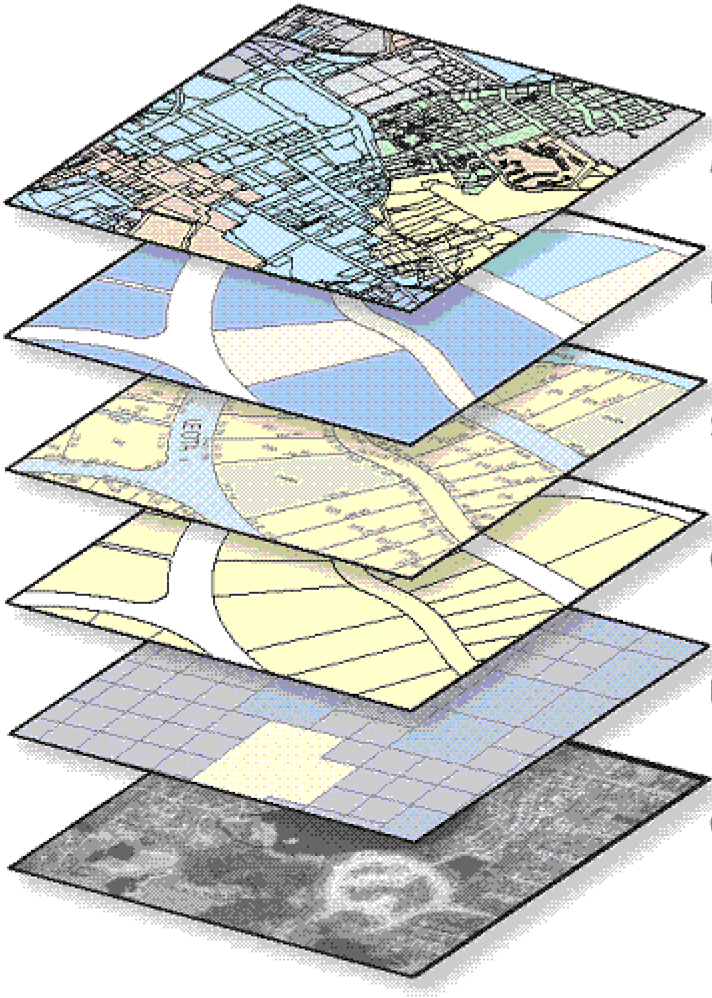
| Theme | Geographic representation |
|---|---|
| Hydrography | Lines |
| Road centerlines | Lines |
| Vegetation | Polygons |
| Urban areas | Polygons |
| Administrative boundaries | Polygons |
| Elevation contours | Lines |

| | |
|---|---|
| Well locations | Points |
| Orthophotography | Raster |
| Satellite imagery | Raster |
| Land parcels | Polygons |
| Parcel tax records | Tables |

**Datasets are the organizing principle in a GIS database**

The concept of a data theme was one of the early notions in GIS. Historically, GIS practitioners thought about how the geographic information in maps could be partitioned into a series of logical information layers—as more than a random collection of objects. They envisioned homogeneous collections of representations that could be managed as layers and that these data layers could be combined through georeferencing. These early GIS users organized information in various data themes that described the distribution of a phenomenon and how each should be portrayed across a geographic extent.

These layers also provided a protocol for collecting the representations. For example, a data theme could be defined that delineated various areas representing the dominant soil type (that is, a layer collection of soil type polygons). Each and every area (the polygons) in a specified extent could be assigned an explicit soil type, and the soil types could be described using properties or attributes of each polygon.




Each GIS will contain multiple themes for a common geographic area. The collection of themes acts as a stack of layers. Each theme can be managed as an information set independent of other themes. Each has its own representation (as a collection of points, lines, polygons, surfaces, rasters, and so on). Because layers are spatially referenced, they overlay one another and can be combined in a common map display. GIS analysis operations, such as polygon overlay, can fuse information between data layers to discover and work with the derived spatial relationships.

### Implications of layer-based data themes

The GIS design concept of layer-based data themes has some key implications:

- All layers must be georeferenced to a place on the earth. This is accomplished by defining the geographic coordinate system for each dataset.
- Layers can be combined in many ways—some that are visual (such as the ordered layers in a map)—and some that employ tool-based operators or "commands". Geographic operators can be used to identify and work with the relationships both within as well as between layers. One

key aspect of GIS is that it provides a sophisticated series of operators that work against the georeferenced collection of data layers. GIS users often refer to this as geoprocessing.

- GIS layers can be combined from many sources and many users. In fact, most users are dependent on one another for portions of the data they want to use. Interoperability is fundamental in GIS.

**GIS uses many datasets and data types**

A GIS will use numerous datasets, each containing its specific representation, often from many organizations. A number of alternative file formats and schemas will be used across a range of systems, but users still have the need to share and re-use each other's data.

Therefore, it is important for GIS datasets to be:

- Simple to use and easy to understand
- Combined easily with other geographic datasets (for example, each has a well-defined spatial reference)
- Effectively compiled and validated
- Clearly documented for content, intended uses, and purposes

Any effective GIS database or file base will adhere to these common principles and concepts regardless of its format. Each GIS requires a mechanism for describing geographic data in these terms along with a comprehensive set of tools to use and manage this information.

### *How ArcGIS users work with geographic information*

- Users work with geographic data in two fundamental ways:
- As datasets, which are homogeneous collections of features, rasters, or attributes—such as parcels, wells, buildings, orthophoto imagery, and raster-based digital elevation models (DEMs)
- As individual elements—such as individual features, rasters, and attribute values—contained within each dataset

In a GIS, homogeneous collections of geographic objects are organized into datasets about common subjects, such as parcels, wells, roads, buildings, orthophoto imagery, and raster-based digital elevation models (DEMs).

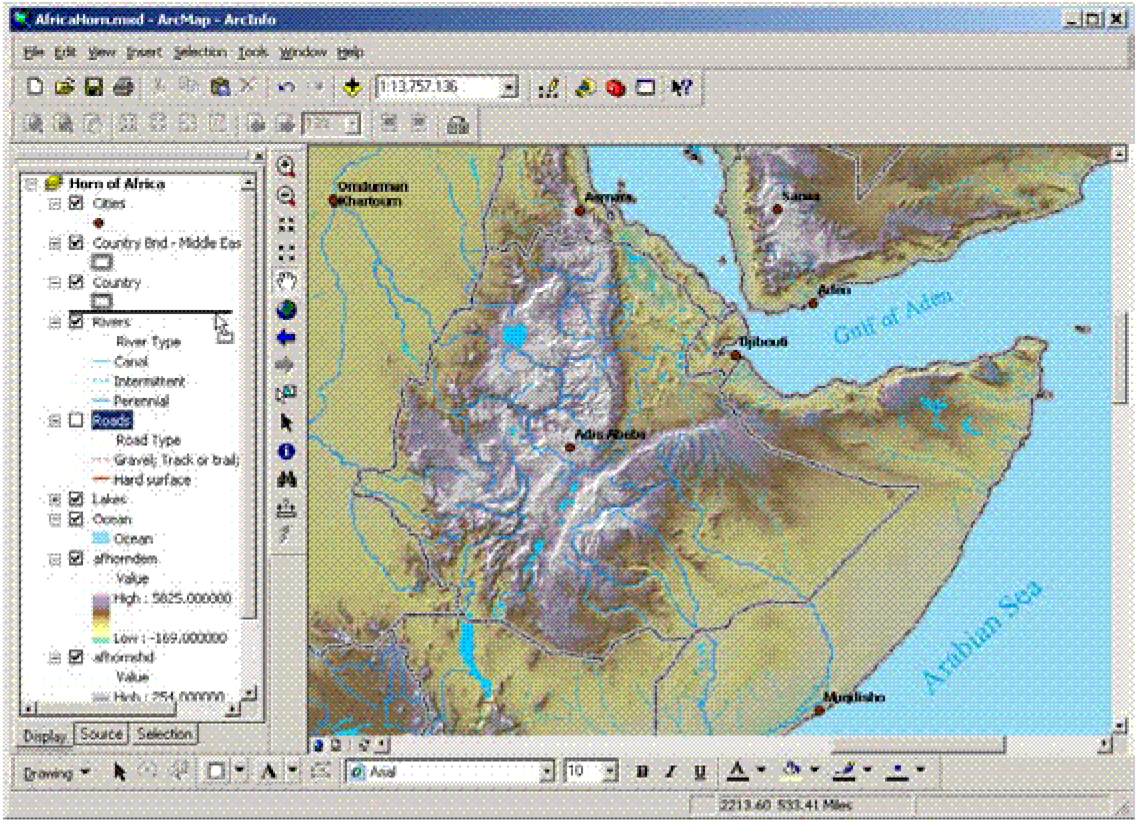
## Users work with GIS datasets

Geographic datasets are the primary object collections that users work with in a GIS. They also represent the most common method for data sharing among GIS users.

Datasets are used as the basis for most GIS operations, and provide the primary data sources for:

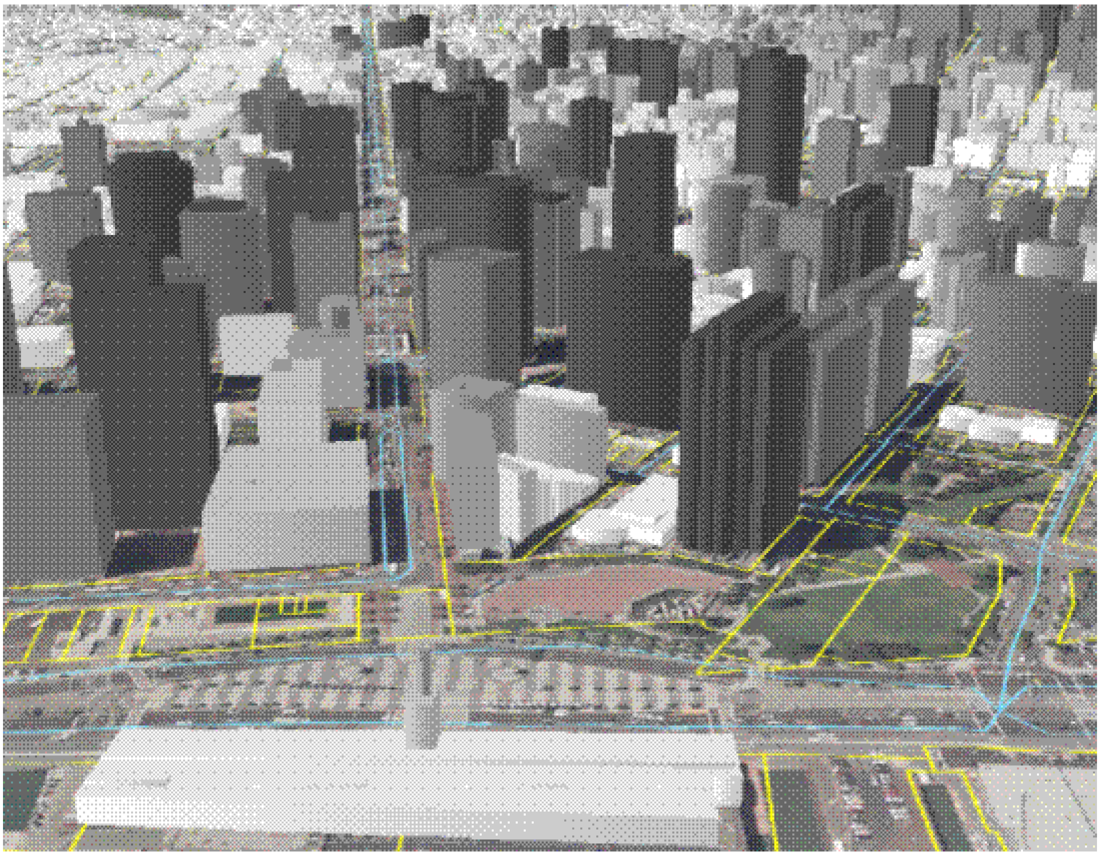
- Maps
- Globes and 3D scenes
- Geoprocessing inputs and derived datasets
- Data sharing with other users

<u>Datasets provide the key inputs for mapping and visualization</u>. Each layer in a map references a dataset and specifies how it will be drawn and labeled.

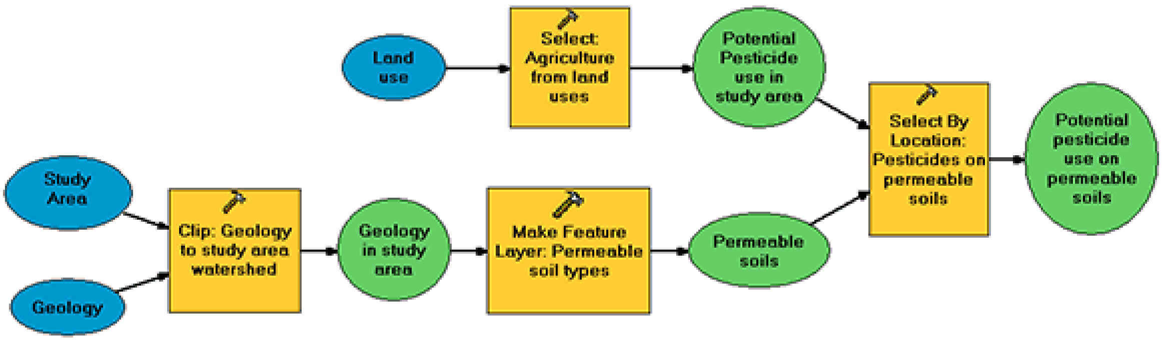



The map display above was created by drawing numerous datasets—feature classes of cities, country boundaries, rivers, and water bodies—on top of a raster dataset of shaded relief.

Datasets are also used as sources for layers in ArcGlobe and ArcScene views.



Datasets are the primary inputs and outputs for geoprocessing. ArcGIS includes a rich set of geoprocessing tools. Each tool takes datasets as inputs, performs a transformation on these datasets, and creates results—called derived datasets. A sequence of operations can be assembled into a *process* to automate workflows, do analysis, and to automate many critical GIS tasks—hence the term geoprocessing.




Datasets are the primary means for data sharing. GIS users primarily share their information as individual datasets. Datasets typically come in any number of formats: CAD files, image files, tables, shapefiles, GML files, and so on. A key goal of ArcGIS is for users to work with all the commonly external file formats as well as ESRI supported formats, such as the geodatabase.

Datasets can be listed in ArcCatalog and can be copied and distributed to other GIS users:

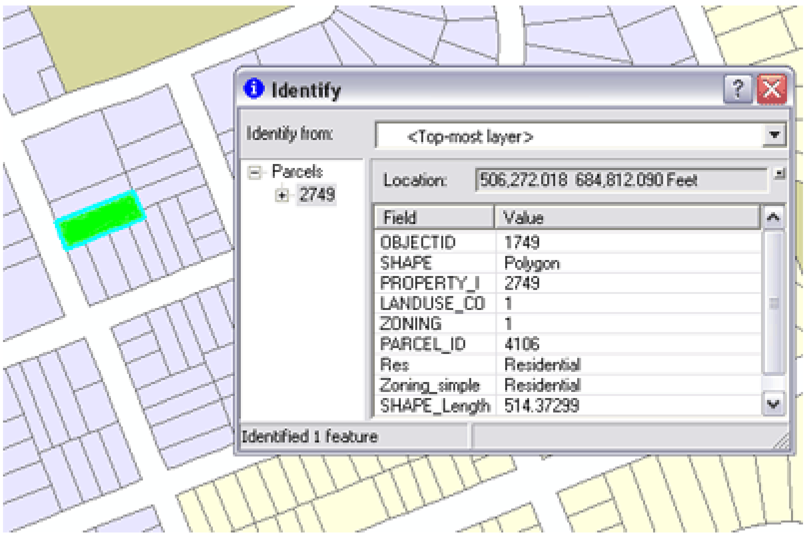
| | | |
|---|---|---|
| Cities | Feature Class | Thu 9/23/2004 3:54 PM |
| CitiesAnno | Feature Class | Thu 9/23/2004 11:06 AM |
| Roads | Feature Class | Mon 10/4/2004 10:55 AM |
| RoadsAnno | Feature Class | Thu 9/30/2004 8:58 AM |
| ParkBoundaries | Feature Class | Tue 9/28/2004 8:56 AM |
| States | Feature Class | Thu 9/23/2004 3:54 PM |
| Streams | Feature Class | Thu 9/23/2004 11:06 AM |
| UtahRelief | Raster Dataset | Mon 10/4/2004 10:55 AM |
| 150mNaturalColor | Raster Dataset | Thu 9/30/2004 8:58 AM |

ArcGIS supports datasets in its native geodatabase as well as multiple GIS file formats. ArcGIS works with geographic datasets that are managed in geodatabases as well as in numerous GIS file formats. Geodatabase datasets represent the native data structure for ArcGIS and are the primary data format used for editing and data management.

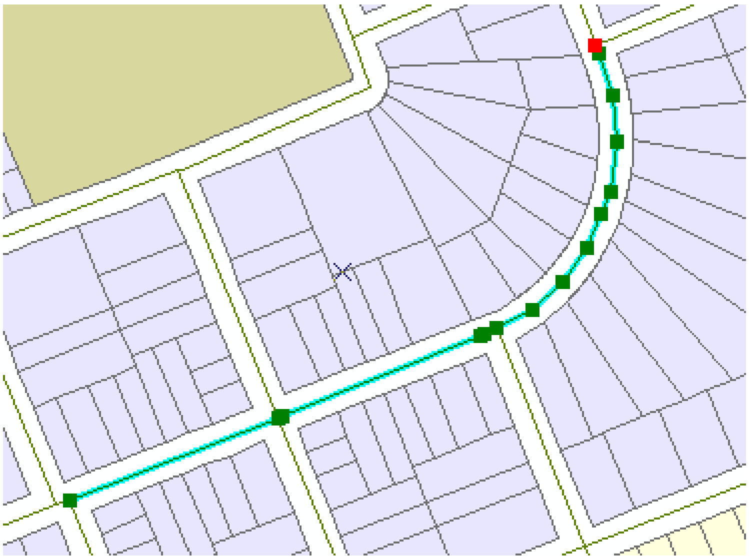
Users work with individual data elements held in each dataset

- In addition to working with datasets, users also work with the individual elements contained in datasets. These elements include individual features, rows and columns in attribute tables, and individual cells in raster datasets. For example:

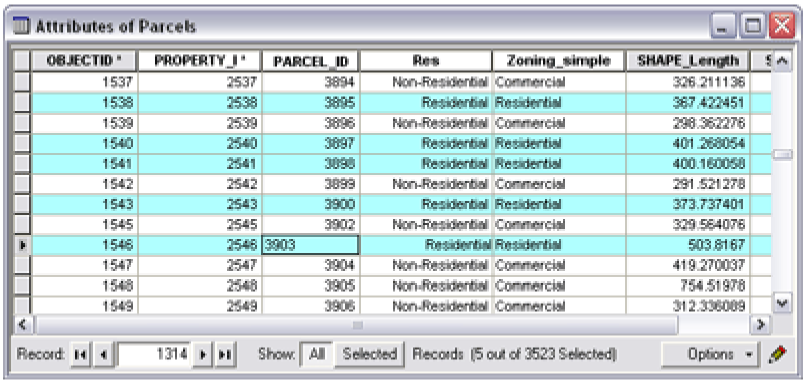
When you identify a parcel by pointing at it, you're working with the individual data elements in a dataset:




You work with individual data elements when you edit features, a road centerline in this case:



In tables, users work with descriptive information contained in rows and columns.



Attributes of Parcels

| OBJECTID * | PROPERTY_I * | PARCEL_ID | Res | Zoning_simple | SHAPE_Length |
|---|---|---|---|---|---|
| 1537 | 2537 | 3894 | Non-Residential | Commercial | 326.211136 |
| 1538 | 2538 | 3895 | Residential | Residential | 367.422451 |
| 1539 | 2539 | 3896 | Non-Residential | Commercial | 298.362276 |
| 1540 | 2540 | 3897 | Residential | Residential | 401.268054 |
| 1541 | 2541 | 3898 | Residential | Residential | 400.160058 |
| 1542 | 2542 | 3899 | Non-Residential | Commercial | 291.521278 |
| 1543 | 2543 | 3900 | Residential | Residential | 373.737401 |
| 1545 | 2545 | 3902 | Non-Residential | Commercial | 329.564076 |
| 1546 | 2546 | 3903 | Residential | Residential | 503.8167 |
| 1547 | 2547 | 3904 | Non-Residential | Commercial | 419.270037 |
| 1548 | 2548 | 3905 | Non-Residential | Commercial | 754.51978 |
| 1549 | 2549 | 3906 | Non-Residential | Commercial | 312.336089 |

Record: 1314 Show: All Selected Records (5 out of 3523 Selected) Options

***Users work with many data types and data formats in ArcGIS***

Users work with many data types and formats in GIS

ArcGIS supports geographic datasets that are managed in geodatabases as well as in numerous GIS file formats. Geodatabase datasets represent the native data structure for ArcGIS and are the primary data format used for editing and data management. Yet, many additional datasets can be used.

A number of additional file formats are supported. These can be used in ArcGIS much like geodatabase datasets—to create layers in ArcMap and ArcGlobe; as inputs for Geoprocessing operations; to be viewed and queried in charts, maps, globes, and tables; and converted to and from many other GIS formats.

The following table lists some of the dataset file types commonly used in ArcGIS.

**Some commonly used external data files in ArcGIS**

| | | |
|---|---|---|
| ESRI | Coverage | ArcInfo Workstation coverages |
| | Grid | ArcInfo GRID raster format |
| | Tin | ArcInfo triangulated irregular network (TIN) format |
| | Shapefile (SHP) | ESRI shapefile format |
| Vector | TIGER/Line | U.S. Census Bureau's TIGER/Line Files |
| | MIF/MID | MapInfo Vector Interchange File<br>MapInfo Table Interchange for MIF |
| | TAB | MapInfo Native Dataset |
| | VPF | National Geospatial Intelligence Agency's Vector Product File format |
| | GML | Open Geospatial Consortium's GML Interchange Specification |
| Raster | IMG | Leica ERDAS Imagine image files |
| | BMP | Bitmap raster format |
| | TIF | TIFF raster format |
| | JPG | JPEG raster compression format |
| | JP2 | JPEG 2000 raster format |
| | SID | MrSID raster format |
| CAD | DXF | CAD transfer file. Uses ASCII or binary drawing file interchange. |
| | DGN | MicroStation design file format |
| | DWG | AutoCAD drawing file format |
| Tables | XLS | Excel spreadsheets |
| | DBF | dBase data file format |
| | Info | Arc/Info Workstation INFO tables |
| | MDB | File format for Microsoft's Access database |
| | TXT | Text file often used to hold attribute columns delimited by commas |

| | | |
|---|---|---|
| | | or tabs |

In addition to these file and RDBMS data sources, ArcGIS can work with numerous additional formats through data conversion. GIS data can also be accessed through networks using Web services and various XML schemas. XML support includes, among others, ArcXML, SOAP, the Open Geospatial Consortium's WMS and WFS protocols, and Geodatabase XML.

### The ArcGIS Data Interoperability extension

The ArcGIS Data Interoperability extension provides direct read access to dozens of additional spatial data formats not already supported in ArcGIS. For example, you can use the Data Interoperability extension to add support for various GML profiles as well as advanced data formats in DWG/DXF, MicroStation Design, MapInfo MID/MIF, and TAB file types.

You can convert to and from these data types and geodatabases using this extension. More importantly, you can use Data Interoperability to directly use these formats in ArcGIS. Users can drag and drop these and many other external data sources into ArcGIS for general use in mapping, geoprocessing, metadata management, and 3D globe use. For example, you can make use of all the mapping functions available to native ESRI formats inside ArcMap for these data sources—such as viewing features and attributes, identifying features, and making selections.

The ArcGIS Data Interoperability extension is developed and maintained collaboratively by ESRI and Safe Software Inc., the leading GIS interoperability vendor, and is based on Safe Software’s popular Feature Manipulation Engine (FME) product.

The ArcGIS Data Interoperability extension also includes FME Workbench, which contains a series of data transformation tools to build converters for many complex vector data formats.

## 3.2. SPATIAL RELATIONSHIP AND TOPOLOGY

GIs is unique in its ability to allow users to create, maintain, and analyze geographic or spatial data. The term spatial data implies that data exist - which not only describe landscape features (e.g., condition, composition, structure of forests) but also reference the location where the features can be found. A GIs allows one to manipulate spatial data and to analyze quickly a large volume of spatial data. A GIs stores spatial data in a digital database files, often referred to as themes, maps, covers, tables, layers, or GIs databases.

Other software programs perform GIs-like tasks (e.g., database management, graphics, or computer-assisted drafting [CAD] software) but does not describe individual data with reference to earth location.

### About topology

What is topology? Topology, or topological coding, provides the intelligence in the data structure relative to the spatial relationships among landscape features (Lillesand and Kiefer 2000). For example, in a vector GIs database containing polygons, topological coding keeps track of each line that forms each polygon, as well as the nodes each line shares with each other line. In addition, the polygons that are formed on either side of each line (since polygons may share a boundary defined by a line) are known. Thus, for example, with topology one can understand which timber stands are next to which other timber stands.

Topology has historically been viewed as a spatial data structure used primarily to ensure that the associated data forms a consistent and clean topological fabric. With advances in object-oriented GIS development, an alternative view of topology has evolved. The geodatabase supports an approach to modeling geography that integrates the behavior of different feature types and supports different types of key relationships. In this context, topology is a collection of rules and relationships that, coupled with a set of editing tools and techniques, enables the geodatabase to more accurately model geometric relationships found in the world.

Topology, implemented as feature behavior and rules, allows a more flexible set of geometric relationships to be modeled than topology implemented as a data structure. It also allows topological relationships to exist between more discrete types of features within a feature dataset. In this alternative view, topology may still be employed to ensure that the data forms a clean and consistent topological fabric, but also more broadly, it is used to ensure that the features obey the key geometric rules defined for their role in the database.

### Why use topology?

Topology is used most fundamentally to ensure data quality and allow your geodatabase to more realistically represent geographic features. A geodatabase provides a framework within which features can have behavior such as subtypes, default values, attribute domains, validation rules, and structured relationships to tables or other features. This behavior enables you to more accurately model the world and maintain referential integrity between objects in the geodatabase. Topology may be considered an extension of this framework for behavior that allows you to control the geometric relationships between features and maintain their geometric integrity. Unlike other feature behavior, topology rules are managed at the level of the topology and dataset, not for individual feature classes.

### How do I work with topology?

Different people work with topology in different ways, depending upon their role in an organization and its GIS design and management work flow.

Initially, creating a topology requires a geodatabase designer. A topology organizes the spatial relationships between features in a set of feature classes. The designer analyzes an organization's data modeling needs, identifies the key topological relationships required in the geodatabase, and defines the rules that will constrain different features' topological relationships.

Once the participating feature classes have been added to the topology and the rules defined, the topology is validated. Data quality managers use the topology tools to analyze; visualize; report; and, where necessary, repair the spatial integrity of the database after it is initially created, as well as after editing. Topology provides these users with a set of validation rules for the topologically related features. It also provides a set of editing tools that let users find and fix integrity violations.

As the geodatabase is used and maintained, new features are added, and existing features are modified. Data editors update features in the geodatabase and use the topology tools to construct and maintain relationships between features within the constraints imposed by the database designer. Depending on the work flow of the organization, the topology may be validated after each edit session or on a schedule.

You can also impose a map topology on your data. A map topology allows you to simultaneously edit simple features that overlap or touch each other. You can create a map topology with an ArcView license (creating a geodatabase topology requires an ArcInfo or ArcEditor license) and apply a map topology to simple features in a shapefile or to simple feature classes in a geodatabase.

### *An overview of topology in the geodatabase*

If you have features that are coincident (i.e., share the same location of coordinates, boundaries, or nodes), chances are that using a geodatabase topology can help you better manage your geographic data.

Most users care deeply about the spatial integrity of their feature data, and geodatabase topologies help you to better manage your data integrity. Using a topology provides a strong mechanism to perform integrity checks on your data and will help you to validate and maintain better feature representations in your geodatabase.

Also, many users like to use topologies for various analytical operations (e.g., to find adjacent features, to work with coincident boundaries between features, and to navigate along connected features). Topologies enable richer analytical functions in your GIS.

Topology is the arrangement for how point, line, and polygon features share geometry. Topology is employed in order to:

- Constrain how features share geometry. For example, adjacent polygons such as parcels have shared edges; street centerlines and census blocks share geometry; adjacent soil polygons share edges; etc.
- Define and enforce data integrity rules (e.g., no gaps should exist between polygons; there should be no overlapping features; and so on).
- Support topological relationship queries and navigation (e.g., to navigate feature adjacency and connectivity).
- Support sophisticated editing tools (tools that enforce the topological constraints of the data model).
- Construct features from unstructured geometry (e.g., to construct polygons from lines).

### *Geometric elements of a topology*

When you create a topology, you specify the feature classes that participate in the topology. These feature classes may have point, line, or polygon features in them. In the topology, the geometric relationships are between the parts of the features rather than the features themselves. Polygons in a topology have:

- Edges that define the boundary of the polygons
- Nodes where edges intersect

- Vertices that define the shape of the edges

Similarly, line features are made up of an edge, at least two nodes that define the endpoints of the edge, and vertices that define the shape of the edge. Point features behave as nodes when they are coincident with other features in a topology.

### *An overview of topology in ArcGIS*

If you have features that are coincident (i.e., share the same location of coordinates, boundaries, or nodes), chances are that using a geodatabase topology can help you better manage your geographic data.

Most users care deeply about the spatial integrity of their feature data, and geodatabase topologies help you to better manage your data integrity. Using a topology provides a strong mechanism to perform integrity checks on your data and will help you to validate and maintain better feature representations in your geodatabase.

Also, many users like to use topologies for various analytical operations (e.g., to find adjacent features, to work with coincident boundaries between features, and to navigate along connected features). Topologies enable richer analytical functions in your GIS.

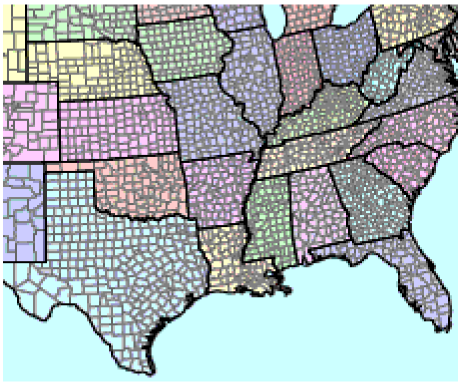
Topology is the arrangement for how point, line, and polygon features share geometry. Topology is employed in order to:

- Constrain how features share geometry. For example, adjacent polygons such as parcels have shared edges; street centerlines and census blocks share geometry; adjacent soil polygons share edges; etc.
- Define and enforce data integrity rules (e.g., no gaps should exist between polygons; there should be no overlapping features; and so on).
- Support topological relationship queries and navigation (e.g., to navigate feature adjacency and connectivity).
- Support sophisticated editing tools (tools that enforce the topological constraints of the data model).
- Construct features from unstructured geometry (e.g., to construct polygons from lines).

The following topics cover key concepts and ArcGIS support for topology. They also cover a number of common workflows for building and using topologies in ArcGIS. See Topology basics and Topology in ArcGIS for conceptual information.

### *Topology basics*

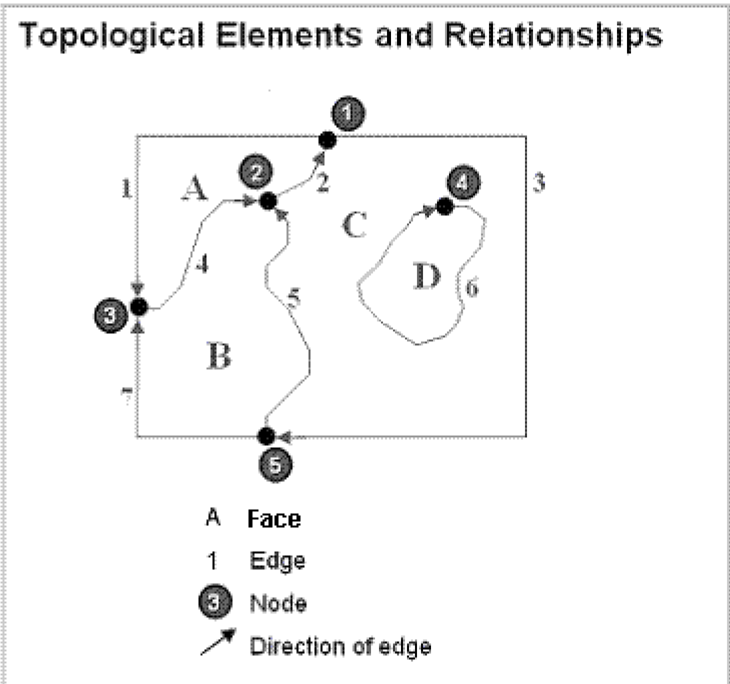
- A GIS topology is a set of rules and behaviors that model how points, lines, and polygons share coincident geometry. For example:
- Adjacent features, such as two counties, will have a common boundary between them. They "share" this edge.
- The set of county polygons within each state must completely cover the state polygon and share edges with the state boundary.



These are examples of topological rules and behaviors that are commonly used to manage coincident geometry in a geodatabase.

### *Why topology?*

Topology has long been a key GIS requirement for data management and integrity. In general, a topological data model represents spatial objects (point, line, and area features) as an underlying graph of topological primitives—nodes, faces, and edges. These primitives, together with their relationships to one another and to the features whose boundaries they represent, are defined by representing the feature geometries in a planar graph of topological elements.




Topology is fundamentally used to ensure data quality and to aid in data compilation. Topology is also used for analyzing spatial relationships in many situations -- such as dissolving the boundaries between adjacent polygons with the same attribute values or traversing along a network of the elements in a topology graph.

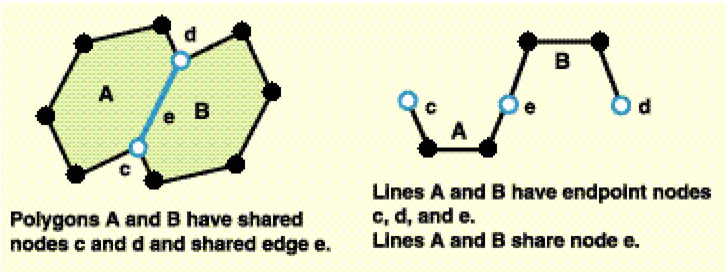
Topology can also be used to model how the geometry from a number of feature classes can be integrated. Some refer to this as *vertical integration* of feature classes.

Generally, topology is employed to do the following:

- Manage coincident geometry (constrain how features share geometry). For example, adjacent polygons, such as parcels have shared edges; street centerlines and the boundaries of census blocks have coincident geometry; adjacent soil polygons share edges, etc.
- Define and enforce data integrity rules (such as no gaps should exist between parcel features, parcels should not overlap, road centerlines should connect at their endpoints).
- Support topological relationship queries and navigation (for example, to provide the ability to identify adjacent and connected features, find the shared edges, and navigate along a series of connected edges).
- Support sophisticated editing tools that enforce the topological constraints of the data model (such as the ability to edit a shared edge and update all the features that share the common edge).
- Construct features from unstructured geometry (e.g., the ability to construct polygons from lines sometimes referred to as "spaghetti").

### *Ways that features share geometry in a topology*

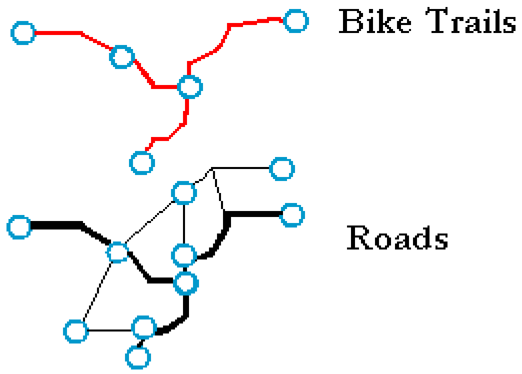
- Features can share geometry within a topology. Here are some examples among adjacent features:




- Area features can share boundaries (polygon topology).
- Line features can share endpoints (edge–node topology).

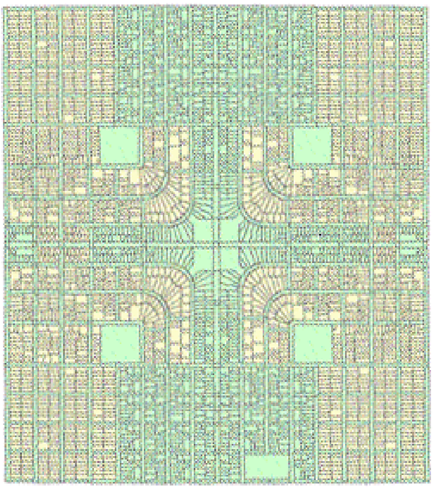
In addition, shared geometry can be managed *between* feature classes using a geodatabase topology. For example:

- Line features can share segments with other line features. For example, parcels can nest within blocks:

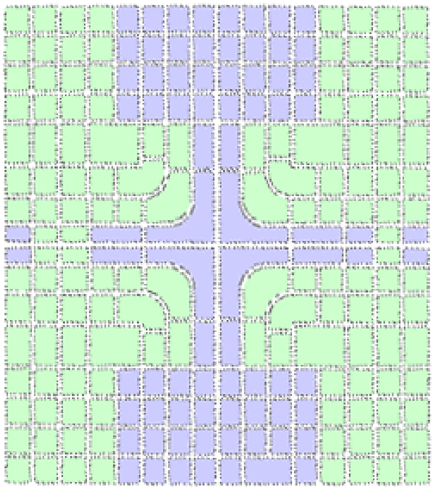



The shared geometry between Roads and Bike Trails are shown as thick lines.

- Area features can be coincident with other area features.



Parcels



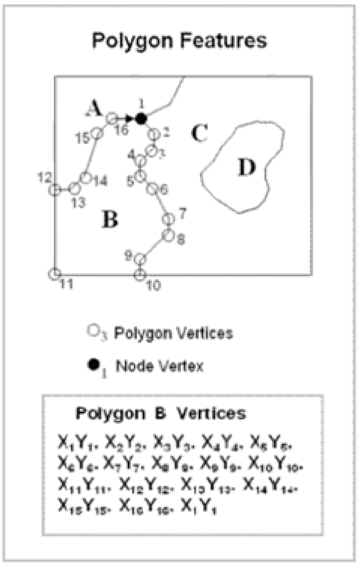
Blocks

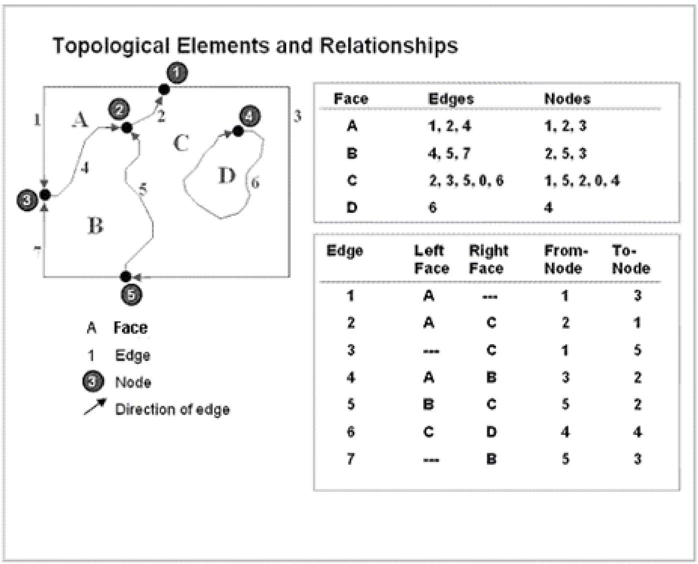
- Line features can share endpoint vertices with other point features (node topology).

- Point features can be coincident with line features (point events).

### *Two views: Features and topological elements*

- The following illustration shows how a layer of polygons can be described and used:
- As collections of geographic features (points, lines, and polygons); or
- As a graph of topological elements (nodes, edges, faces, and their relationships).




| Face | Edges | Nodes |
|---|---|---|
| A | 1, 2, 4 | 1, 2, 3 |
| B | 4, 5, 7 | 2, 5, 3 |
| C | 2, 3, 5, 0, 6 | 1, 5, 2, 0, 4 |
| D | 6 | 4 |

| Edge | Left Face | Right Face | From-Node | To-Node |
|---|---|---|---|---|
| 1 | A | --- | 1 | 3 |
| 2 | A | C | 2 | 1 |
| 3 | --- | C | 1 | 5 |
| 4 | A | B | 3 | 2 |
| 5 | B | C | 5 | 2 |
| 6 | C | D | 4 | 4 |
| 7 | --- | B | 5 | 3 |

This means that there are two alternatives for working with features—one in which features are defined by their coordinates and another in which features are represented as an ordered graph of their topological elements.

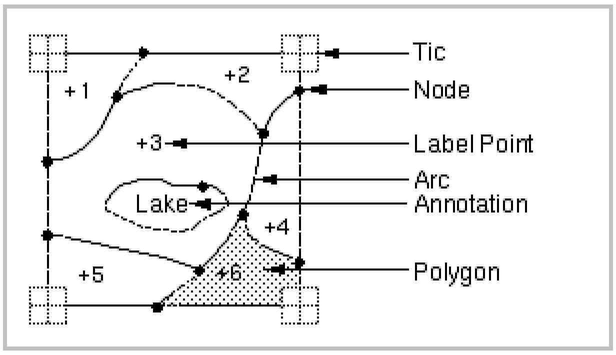
The evolution of geodatabase topology from ArcInfo coverages

NOTE: Reading this large topic is not necessary to implement geodatabase topologies. However, you may want to spend some time reading this if you are interested in the historical evolution and motivations for how topology is managed in the geodatabase.

### *The genesis of "Arc-node" and "Geo-relational"*

ArcInfo coverage users have a long history and appreciation for the role that topology plays in maintaining the spatial integrity of their data.

Here are the elements of the ArcInfo coverage data model.




In a coverage, the feature boundaries and points were stored in a few main files that were managed and owned by ArcInfo Workstation. The "ARC" file held the linear or polygon boundary geometry as topological edges, which were referred to as "arcs". The "LAB" file held point locations, which were used as label points for polygons or as individual point features such as for a wells feature layer. Other files were used to define and persist the topological relationships between each of the edges and the polygons.

For example, one file called the "PAL" file ("Polygon-arc list") listed the order and direction of the arcs in each polygon. In ArcInfo, software logic was used to assemble the coordinates for each polygon for display, analysis, and query operations. The ordered list of edges in the PAL file was used to look up and assemble the edge coordinates held in the ARC file. The polygons were assembled during run-time when needed.

The coverage model had several advantages:

- It used a simple structure to maintain topology.
- It enabled edges to be digitized and stored only once and shared by many features.
- It could represent polygons of enormous size (with thousands of coordinates) because polygons were really defined as an ordered set of edges (called "arcs")
- The Topology storage structure of the coverage was intuitive. Its physical topological files were readily understood by ArcInfo users.

NOTE: An interesting historical fact: "Arc" when coupled with the table manager named "Info" was the genesis of the product name "ArcInfo" and hence all subsequent "Arc" products in the ESRI product family-- ArcView, ArcIMS, ArcGIS, etc.

### *Coverages also had some disadvantages:*

- Some operations were slow because many features had to be assembled on the fly when they needed to be used. This included all polygons and multi-part features such as regions (the coverage term for multi-part polygons) and routes (the term for multi-part line features).
- Topological features (such as polygons, regions, and routes) were not ready-to-use until the coverage topology was built. If edges were edited, the topology had to be rebuilt. (Note: "Partial processing" was eventually used, which required rebuilding only the changed portions of the coverage topology.) In general, when edits are made to features in a topological dataset, a geometric analysis algorithm must be executed to rebuild the topological relationships regardless of the storage model.
- Coverages were limited to single-user editing. Because of the need to ensure that the topological graph was in synchronization with the feature geometries, only a single user could update a topology at a time. Users would tile their coverages and maintain a tiled database for editing. This enabled individual users to "lock down" and edit one tile at a time. For general data use and deployment, users would append copies of their tiles into a mosaicked data layer. In other words, the tiled datasets they edited were not directly used across the organization. They had to be converted, which meant extra work and extra time.

### *Shapefiles and simple geometry storage*

In the early 1980s, coverages were seen as a major improvement over the older polygon and line-based systems in which polygons were held as complete loops. In these older systems, all of the coordinates for a feature were stored in each feature's geometry. Before the coverage and ArcInfo came along, these simple polygon and line structures were used. These data structures were simple, but had the disadvantage of "double digitized boundaries". That is, two copies of the coordinates of the adjacent portions of polygons with shared edges would be contained in each polygon's geometry. The main disadvantage was that GIS software at the time could not maintain shared edge integrity. Plus, storage costs were enormous and each byte of storage came at a premium. During the early 1980s, a 300 MB disk drive was the size of a washing machine and cost $30,000! Holding two or more representations of coordinates was expensive and the computations took too much compute time. Thus, the use of a coverage topology had real advantages.

During the mid 1990s, interest in simple geometric structures grew because disk storage and hardware costs in general were coming down while computational speed was growing. At the same time, existing GIS datasets were more readily available, and the work of GIS users was evolving from primarily data compilation activities to include data use, analysis, and sharing.

Users wanted faster performance for data use (for example, don't spend computer time to derive polygon geometries when we need them. Just deliver the feature coordinates of these 1,200 polygons as fast as possible). Having the full feature geometry readily available was more efficient. Thousands of GISs were in use and numerous datasets were readily available.

Around this time, ESRI had developed and published its ESRI Shapefile format. Shapefiles used a very simple storage model for feature coordinates. Each shapefile represented a single feature class (of points, lines, or polygons) and used a simple storage model for the feature's coordinates. Shapefiles could be easily created from ArcInfo coverages as well as many other GIS systems. They were widely adopted as a de facto standard and are still massively used and deployed to this day.

A few years later, ArcSDE pioneered a similar simple storage model in relational database tables. A feature table could hold one feature per row with the geometry in one of its columns along with other feature attribute columns.

A sample feature table of state polygons is shown below. Each row represents a state. The shape column holds the polygon geometry of each state.

Shape Column

| OBJECTID * | SHAPE * | AREA | STATE_NAME | STATE_ABBR |
|---|---|---|---|---|
| 45 | Polygon | 30867.398 | South Carolina | SC |
| 48 | Polygon | 55814.731 | Florida | FL |
| 26 | Polygon | 56299.387 | Illinois | IL |
| 50 | Polygon | 6381.227 | Hawaii | HI |
| 18 | Polygon | 4976.566 | Connecticut | CT |
| 44 | Polygon | 58629.222 | Georgia | GA |
| 24 | Polygon | 157776.31 | California | CA |
| 40 | Polygon | 264435.873 | Texas | TX |
| 34 | Polygon | 39819.882 | Virginia | VA |
| 10 | Polygon | 84520.49 | Minnesota | MN |
| 27 | Polygon | 66.063 | District of Columbia | DC |
| 17 | Polygon | 45360.118 | Pennsylvania | PA |
| 19 | Polygon | 1044.881 | Rhode Island | RI |
| 37 | Polygon | 70003.325 | Oklahoma | OK |
| 6 | Polygon | 97803.199 | Wyoming | WY |
| 38 | Polygon | 49048.024 | North Carolina | NC |
| 15 | Polygon | 77330.258 | Nebraska | NE |
| 3 | Polygon | 32161.925 | Maine | ME |
| 11 | Polygon | 97073.594 | Oregon | OR |
| 22 | Polygon | 110669.975 | Nevada | NV |
| 4 | Polygon | 70812.056 | North Dakota | ND |
| 2 | Polygon | 147244.653 | Montana | MT |
| 1 | Polygon | 67290.061 | Washington | WA |
| 12 | Polygon | 9259.527 | New Hampshire | NH |
| 42 | Polygon | 51715.786 | Alabama | AL |
| 23 | Polygon | 84871.909 | Utah | UT |

Feature

This simple features model fits the SQL processing engine very well. Through the use of relational databases, we began to see GIS data scale to unprecedented sizes and numbers of users without degrading performance. We were beginning to leverage RDBMS for GIS data management.

Shapefiles became ubiquitous and, using ArcSDE, this simple features mechanism became the fundamental feature storage model in RDBMSs. (To support interoperability, ESRI was the lead author of the OGC and ISO simple features specification).

Simple feature storage had clear advantages:

- The complete geometry for each feature is held in one record. No assembly is required.
- The data structure (physical schema) is very simple, fast, and scalable.
- It is easy for programmers to write interfaces.
- It is interoperable. Many wrote simple converters to move data in and out of these simple geometries from numerous other formats. Shapefiles were widely applied as a data use and interchange format.

Its disadvantages were that maintaining the data integrity that was readily provided by topology was not as easy to implement for simple features. As a consequence, users applied one data model for editing and maintenance (such as coverages) and used another for deployment (such as shapefiles or ArcSDE layers).

Users began to use this hybrid approach for editing and data deployment. For example, users would edit their data in coverages, CAD files, or other formats. Then, they would convert their data into shapefiles for deployment and use. Thus, even though the simple features structure was an excellent direct use format, it did not support the topological editing and data management of shared geometry. Direct use databases would use the simple structures, but another topological form was used for editing. This had advantages for deployment. But the disadvantage was that data would become out of date and have to be refreshed. It worked, but there was a lag time for information update. Bottom line—topology was missing.

What GIS required and what the geodatabase topology model implements now is a mechanism that stores features using the simple feature geometry, but enables topologies to be used on this simple, open data structure. This means that users can have the best of both worlds—a transactional data model that enables topological query, shared geometry editing, rich data modeling, and data integrity, but also a simple, highly scalable data storage mechanism that is based upon open, simple feature geometry.

This direct use data model is fast, simple, and efficient. It can also be directly edited and maintained by any number of simultaneous users.

### *The topology framework in ArcGIS*

In effect, topology has been considered as more than a data storage problem. The complete solution includes:

- A complete data model (objects, integrity rules, editing and validation tools, a topology and geometry engine that can process datasets of any size and complexity, and a rich set of topological operators, map display, and query tools).
- An open storage format using a set of record types for simple features and a topological interface to query simple features, retrieve topological elements, and navigate their spatial relationships (e.g., find adjacent areas and their shared edge, route along connected lines).
- The ability to provide the features (points, lines, and polygons) as well as the topological elements (nodes, edges, and faces) and their relationships to one another.
- A mechanism that can support:
    - Massively large datasets with millions of features.
    - Ability to perform editing and maintenance by many simultaneous editors.
    - Ready-to-use, always available feature geometry.
    - Support for topological integrity and behavior.
    - A system that goes fast and scales for many users and many editors.
    - A system that is flexible and simple.
    - A system that leverages the RDBMS SQL engine and transaction framework.
    - A system that can support multiple editors, long transactions, historical archiving, and replication.

In a geodatabase topology, the validation process identifies shared coordinates between features (both in the same feature class and across feature classes). A clustering algorithm is used to ensure that the shared coordinates have the same location. These shared coordinates are stored as part of each feature's simple geometry.

This enables very fast and scalable lookup of topological elements (nodes, edges, and faces). This has the added advantage of working quite well and scaling with the RDBMS's SQL engine and transaction management framework.

During editing and update, as features are added, they are directly usable. The updated areas on the map, called "dirty areas", are flagged and tracked as updates are made to each feature class. At any time, users can choose to topologically analyze and validate the dirty areas to generate clean topology. Only the topology for the dirty areas needs rebuilding, saving processing time.

The results are that topological primitives (nodes, edges, faces) and their relationships to one another and their features can be efficiently discovered and assembled. This has several advantages:

- Simple feature geometry storage is used for features. This storage model is open, efficient, and scales to large sizes and numbers of users.
- This simple features data model is transactional and is multi-user. By contrast the older topological storage models will not scale and have difficulties supporting multiple editor transactions and numerous other GIS data management workflows.
- Geodatabase topologies fully support all of the long transaction and versioning capabilities of the geodatabase. Geodatabase topologies need not be tiled, and many users can simultaneously edit the topological database—even their individual versions of the same features if necessary.
- Feature classes can grow to any size (hundreds of millions of features) with very strong performance.
- This topology implementation is additive. You can typically add this to an existing schema of spatially related feature classes. The alternative is that you must redefine and convert all of your existing feature classes to new data schemas holding topological primitives.
- There need only be one data model for geometry editing and data use, not two or more.
- It is interoperable because all feature geometry storage adheres to simple features specifications from the OpenGIS Consortium and ISO.
- Data modeling is more natural because it is based on user features (such as parcels, streets, soil types, and watersheds) instead of topological primitives (such as nodes, edges, and faces). Users will begin to think about the integrity rules and behavior of their actual features instead of the integrity rules of the topological primitives. For example, how do parcels behave? This will enable stronger modeling for all kinds of geographic features. It will improve our thinking about streets, soils types, census units, watersheds, rail systems, geology, forest stands, landforms, physical features, and on and on.
- Geodatabase topologies provide the same information content as persisted topological implementations—either you store a topological line graph and discover the feature geometry (like ArcInfo coverages) or you store the feature geometry and discover the topological elements and relationships (like geodatabases).

In cases where users want to store the topological primitives, it is easy to create and post topologies and their relationships to tables for various analytical and interoperability purposes (such as users who want to post their features into an Oracle Spatial warehouse which stores tables of topological primitives).

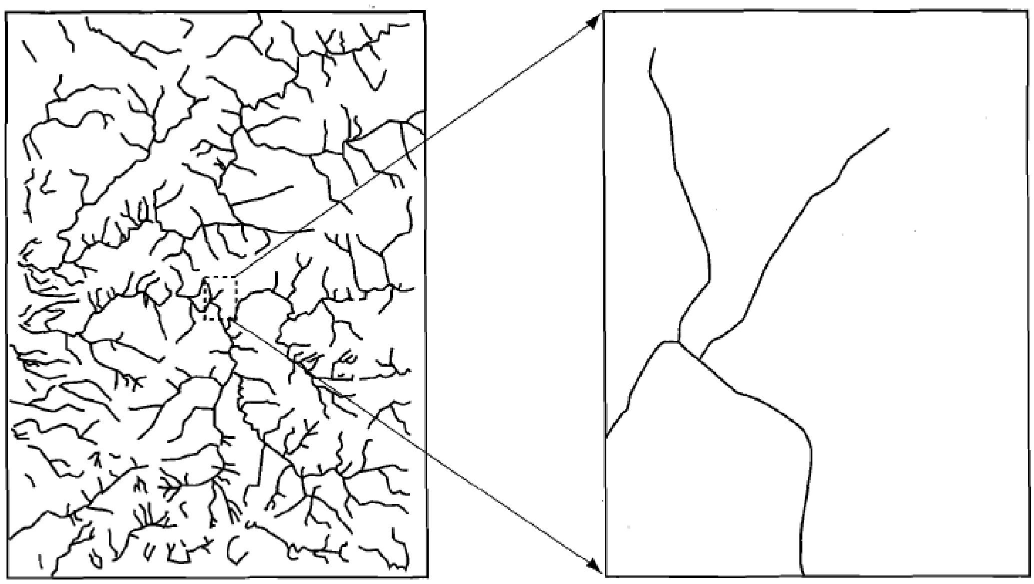
At a pragmatic level, the ArcGIS topology implementation works. It scales to extremely large geodatabases and multi-user systems without loss of performance. It includes rich validation and editing tools for building and maintaining topologies in geodatabases. It includes rich and flexible data modeling tools that enable users to assemble practical, working systems on file systems, in any relational database, and on any number of schemas.

## 3.3. SCALE AND RESOLUTION

GIS databases are often characterized in terms of their **scale**, or **resolution**. Issues of scale are usually associated with vector GIS databases, whereas issues of resolution are associated with raster GIS databases. Scale and resolution both refer to characteristics of the landscape features represented in GIS databases. Typically, this relates to the source material from which the GIS databases were created. Source material includes aerial photographs, existing maps, satellite data, and information gathered from survey instruments, such as total stations or GPS receivers. Many sources of vector data are derived from remote sensing techniques, particularly from aerial photographs. The scale that is associated with the vector GIS databases typically relates to photographic scale, a function of camera height, lens length, and photo size. Scale is often expressed as a ratio, or representative fraction, such as 1:24,000 or 1:100,000 (Muehrcke and Muehrcke 1998). The ratio expression is unitless and implies that 1 unit of measurement on a map or photo represents 24,000 or 100,000 units on the ground. Sometimes, confusion exists as to the correct use of the terms large scale and small scale. The ratio 1:24,000 is a larger ratio than 1: 100,000 (1 is a larger portion of 24,000 than of 100,000); thus, 1:24,000 is a larger scale than 1:100,000. If one examined both 1:24,000 and 1:100,000 scale maps printed on the same size paper, the 1:24,000 map would show less area but greater detail than the 1: 100,000 map. Scale can also be referred to in terms of relative units, such as 1 cm = 1 km, or through the use of a scale that graphically illustrates approximate ground distances.

With imagery derived from satellite and aerial platforms, the ability of the electromagnetic sensor on the platform to delineate landscape features on the ground determines the resolution. A 1m resolution image implies that the sensors used to collect the imagery captured a value for each square meter of the landscape. For raster CIS databases that were developed by scanning from maps or photographs, such as a DRG or DOQ, the size of the raster grid cell in representing landscape features determines the resolution. If each raster grid cell spans a 30 m ground distance, the raster GIS database is said to have a "30 m resolution." This means that each raster grid cell represents 900 m2 (30 m X 30 m) of ground area.

Although scales, or resolutions', are associated with spatial databases, some users mistakenly believe that they can improve the detail of GIS databases by focusing on small land areas. Users need to be aware of the fact that the scale, or resolution, of a GIS database remains static, regardless of how closely one views an area of the landscape.



1:100,000 map    1:24,000 map

## 3.4. VECTOR, RASTER AND DIGITAL TERRAIN MODEL

There are three basic spatial data types used with GIS (points, lines, and areas):

*Points represent anything that can be described as a discrete x, y location

*Lines represent anything having a length

*Areas, or polygons, describe anything having boundaries

These data types comprise the vector model, which is the model you will deal with most often in GIS.

Vector data model:

Discrete features, such as customer locations, are usually represented using the vector model. Features can be discrete locations or events, lines, or areas. Lines, such as streams or roads, are represented as a series of coordinate pairs. Areas are defined by borders, and are represented by closed polygons. When you analyze vector data, much of your analysis involves working with (summarizing) the attributes in the layer's data table.

Raster data model:

Continuous numeric values, such as elevation, and continuous categories, such as vegetation types, are represented using the raster model. The raster data model represents features as a matrix/lattice of cells in continuous space. A point is one cell, a line is a continuous row of cells, and an area is represented as continuous touching cells.

Tabular data:

Contain information describing a map feature in the form of a table or spreadsheet. For example, a GIS database of customer locations may be linked to address and personnel information. GIS links this tabular data to associated spatial data.

Brief geodatabase types:

A geodatabase is a database that is in some way referenced to locations on the earth. Coupled with this data is usually data known as attribute data. Attribute data generally defined as additional information, which can then be tied to spatial data.

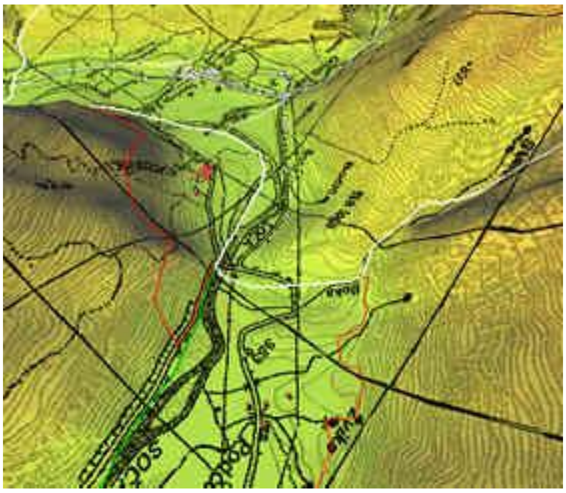
GIS data can be separated into two categories: spatially referenced data which is represented by vector and raster forms (including imagery) and attribute tables which is represented in tabular format. Within the spatial referenced data group, the GIS data can be further classified into two different types: vector and raster. Most GIS software applications mainly focus on the usage and manipulation of vector geodatabases with added components to work with raster-based geodatabases.

### *GIS file formats*

A GIS file format is a standard of encoding geographical information into a file. They are created mainly by government mapping agencies (such as the USGS or National Geospatial-Intelligence Agency) or by GIS software developers.

# Raster

A raster data type is, in essence, any type of digital image represented by reducible and enlargeable grids. Anyone who is familiar with digital photography will recognize the Raster graphics pixel as the smallest individual grid unit building block of an image, usually not readily identified as an artifact shape until an image is produced on a very large scale. A combination of the pixels making up an image color formation scheme will compose details of an image, as is distinct from the commonly used points, lines, and polygon area location symbols of scalable vector graphics as the basis of the vector model of area attribute rendering. While a digital image is concerned with its output blending together its grid based details as an identifiable representation of reality, in a photograph or art image transferred into a computer, the raster data type will reflect a digitized abstraction of reality dealt with by grid populating tones or objects, quantities, cojoined or open boundaries, and map relief schemas. Aerial photos are one commonly used form of raster data, with one primary purpose in mind: to display a detailed image on a map area, or for the purposes of rendering its identifiable objects by digitization. Additional raster data sets used by a GIS will contain information regarding elevation, a digital elevation model, or reflectance of a particular wavelength of light, Landsat, or other electromagnetic spectrum indicators.



Digital elevation model, map (image), and vector data

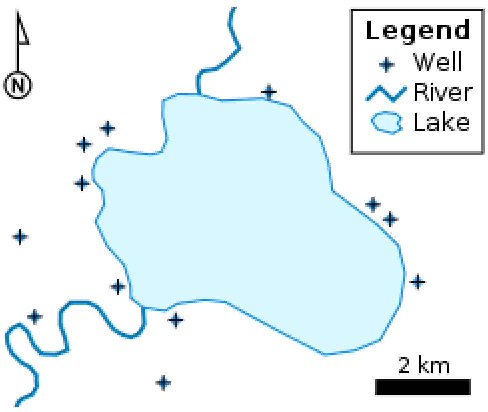
Raster data type consists of rows and columns of cells, with each cell storing a single value. Raster data can be images (raster images) with each pixel (or cell) containing a color value. Additional values recorded for each cell may be a discrete value, such as land use, a continuous value, such as temperature, or a null value if no data is available. While a raster cell stores a single value, it can be extended by using raster bands to represent RGB (red, green, blue) colors, colormaps (a mapping between a thematic code and RGB value), or an extended attribute table with one row for each unique cell value. The resolution of the raster data set is its cell width in ground units.

Raster data is stored in various formats; from a standard file-based structure of TIFF, JPEG, etc. to binary large object (BLOB) data stored directly in a relational database management system (RDBMS) similar to other vector-based feature classes. Database storage, when properly indexed, typically allows for quicker retrieval of the raster data but can require storage of millions of significantly sized records.

# Vector

In a GIS, geographical features are often expressed as vectors, by considering those features as geometrical shapes. Different geographical features are expressed by different types of geometry:

- Points




A simple vector map, using each of the vector elements: points for wells, lines for rivers, and a polygon for the lake

Zero-dimensional points are used for geographical features that can best be expressed by a single point reference—in other words, by simple location. Examples include wells, peaks, features of interest, and trailheads. Points convey the least amount of information of these file types. Points can also be used to represent areas when displayed at a small scale. For example, cities on a map of the world might be represented by points rather than polygons. No measurements are possible with point features.

- Lines or polylines

One-dimensional lines or polylines are used for linear features such as rivers, roads, railroads, trails, and topographic lines. Again, as with point features, linear features displayed at a small scale will be represented as linear features rather than as a polygon. Line features can measure distance.

- Polygons

Two-dimensional polygons are used for geographical features that cover a particular area of the earth's surface. Such features may include lakes, park boundaries, buildings, city boundaries, or land uses. Polygons convey the most amount of information of the file types. Polygon features can measure perimeter and area.

Each of these geometries are linked to a row in a database that describes their attributes. For example, a database that describes lakes may contain a lake's depth, water quality, pollution level. This information can be used to make a map to describe a particular attribute of the dataset. For example, lakes could be coloured depending on level of pollution. Different geometries can also be compared. For example, the GIS could be used to identify all wells (point geometry) that are within one kilometre of a lake (polygon geometry) that has a high level of pollution.

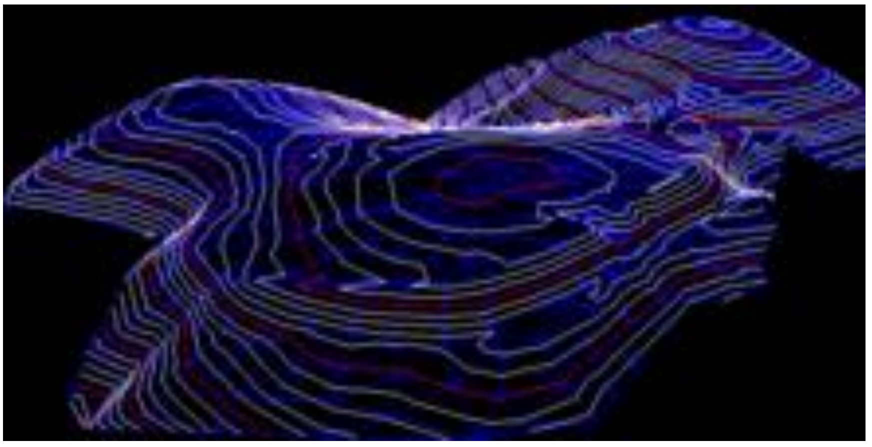
Vector features can be made to respect spatial integrity through the application of topology rules such as 'polygons must not overlap'. Vector data can also be used to represent continuously varying phenomena. Contour lines and triangulated irregular networks (TIN) are used to represent elevation or other continuously changing values. TINs record values at point locations, which are connected by lines to form an irregular mesh of triangles. The face of the triangles represent the terrain surface.

## Non-spatial data

Additional non-spatial data can also be stored along with the spatial data represented by the coordinates of a vector geometry or the position of a raster cell. In vector data, the additional data contains attributes of the feature. For example, a forest inventory polygon may also have an identifier value and information about tree species. In raster data the cell value can store attribute information, but it can also be used as an identifier that can relate to records in another table.

# Triangulated irregular network

A **triangulated irregular network** (**TIN**) is a digital data structure used in a geographic information system (GIS) for the representation of a surface. A TIN is a vector-based representation of the physical land surface or sea bottom, made up of irregularly distributed nodes and lines with three-dimensional coordinates (*x*, *y*, and *z*) that are arranged in a network of nonoverlapping triangles. TINs are often derived from the elevation data of a rasterized digital elevation model (DEM). An advantage of using a TIN over a raster DEM in mapping and analysis is that the points of a TIN are distributed variably based on an algorithm that determines which points are most necessary to an accurate representation of the terrain. Data input is therefore flexible and fewer points need to be stored than in a raster DEM, with regularly distributed points. A TIN may be less suited than a raster DEM for certain kinds of GIS applications, such as analysis of a surface's slope and aspect. TINs were first invented by Phil Mellor whilst studying Sociology at University of Edinburgh in 1947.



TIN overlaid with contour lines

A TIN comprises a triangular network of vertices, known as mass points, with associated coordinates in three dimensions connected by edges to form a triangular tessellation. Three-dimensional visualizations are readily created by rendering of the triangular facets. In regions where there is little variation in surface height, the points may be widely spaced whereas in areas of more intense variation in height the point density is increased.

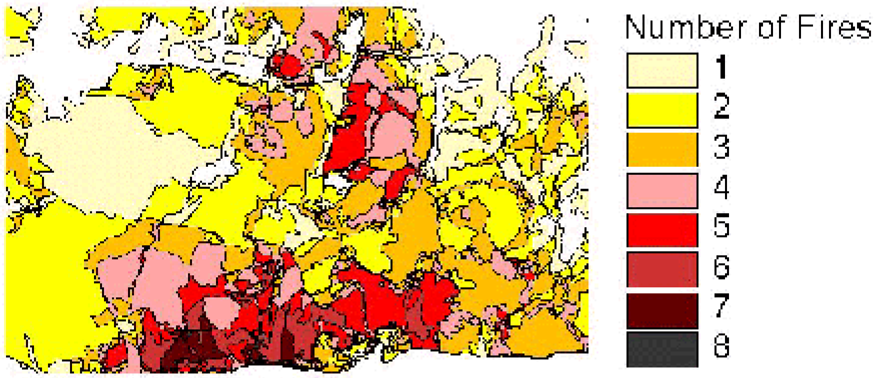
A TIN is typically based on a Delaunay triangulation, but its utility will be limited by the selection of input data points: well-chosen points will be located so as to capture significant changes in surface form, such as topographical summits, breaks of slope, ridges, valley floors, pits, and cols.

Although usually associated with three-dimensional data (*x*, *y*, and *z*) and topography, TINs are also useful for the description and analysis of general horizontal (*x* and *y*) distributions and relationships.

The first triangulated irregular network program for GIS was written by W. Randolph Franklin, under the direction of David Douglas and Thomas Peucker (Poiker), at Simon Fraser University in 1973

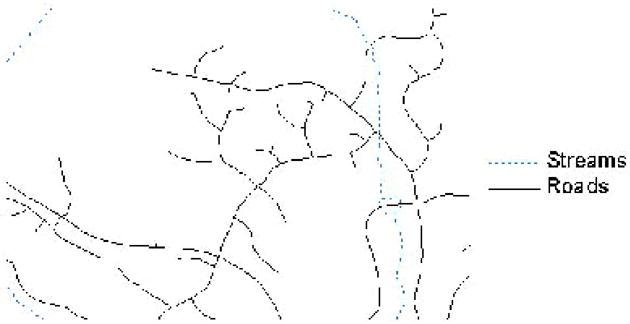
### *Vector data*

Vector data is split into three types: polygon, line (or arc) and point data. Polygons are used to represent areas such as the boundary of a city (on a large scale map), lake, or forest. Polygon features are two dimensional and therefore can be used to measure the area and perimeter of a geographic feature. Polygon features are most commonly distinguished using either a thematic mapping symbology (color schemes), patterns, or in the case of numeric gradation, a color gradation scheme could be used.



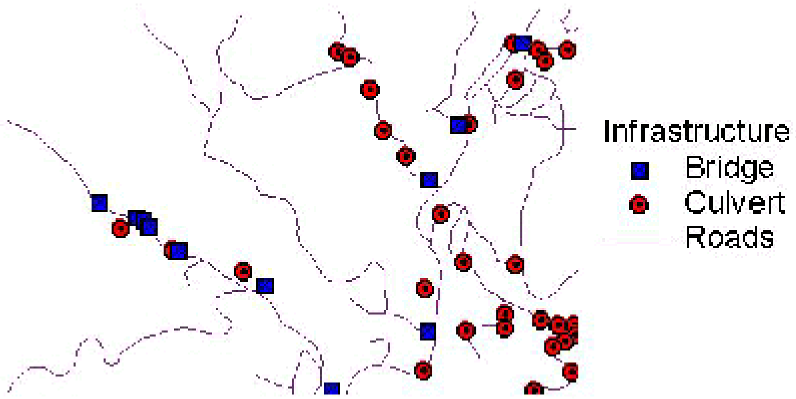

In this view of a polygon based dataset, frequency of fire in an area is depicted showing a graduate color symbology.

Line (or arc) data is used to represent linear features. Common examples would be rivers, trails, and streets. Line features only have one dimension and therefore can only be used to measure length. Line features have a starting and ending point. Common examples would be road centerlines and hydrology. Symbology most commonly used to distinguish arc features from one another are line types (solid lines versus dashed lines) and combinations using colors and line thicknesses. In the example below roads are distinguished from the stream network by designating the roads as a solid black line and the hydrology a dashed blue line.




Streams are shown as dashed blue lines and roads as solid black lines in this example.

Point data is most commonly used to represent nonadjacent features and to represent discrete data points. Points have zero dimensions, therefore you can measure neither length or area with this dataset. Examples would be schools, points of interest, and in the example below, bridge and culvert locations. Point features are also used to represent abstract points. For instance, point locations could represent city locations or place names.

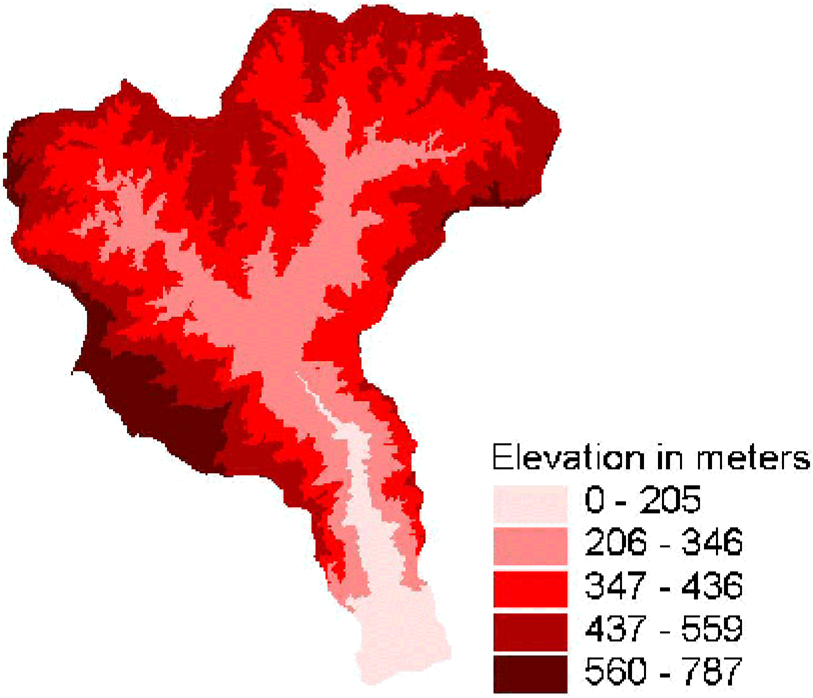



GIS point data showing the location of bridges and culverts.

Both line and point feature data represent polygon data at a much smaller scale. They help reduce clutter by simplifying data locations. As the features are zoomed in, the point location of a school is more realistically represented by a series of building footprints showing the physical location of the campus. Line features of a street centerline file only represent the physical location of the street. If a higher degree of spatial resolution is needed, a street curbwidth file would be used to show the width of the road as well as any features such as medians and right-of-ways (or sidewalks).

### *Raster Data*

Raster data (also known as grid data) represents the fourth type of feature: surfaces. Raster data is cell-based and this data category also includes aerial and satellite imagery. There are two types of raster data: continuous and discrete. An example of discrete raster data is population density. Continuous data examples are temperature and elevation measurements. There are also three types of raster datasets: thematic data, spectral data, and pictures (imagery).

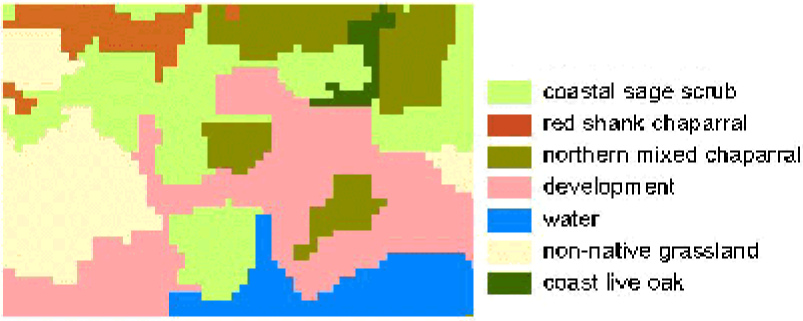



Digital Elevation Model (DEM) showing elevation.

This example of a thematic raster dataset is called a Digital Elevation Model (DEM). Each cell presents a 30m pixel size with an elevation value assigned to that cell. The area shown is the Topanga Watershed in California and gives the viewer and understand of the topography of the region.


This image shows a portion of Topanga, California taken from a USGS DOQ.

Each cell contains one value representing the dominate value of that cell. Raster datasets are intrinsic to most spatial analysis. Data analysis such as extracting slope and aspect from Digital Elevation Models occurs with raster datasets. Spatial hydrology modeling such as extracting watersheds and flow lines also uses a raster-based system. Spectral data presents aerial or satellite imagery which is then often used to derive vegetation geologic information by classifying the spectral signatures of each type of feature.



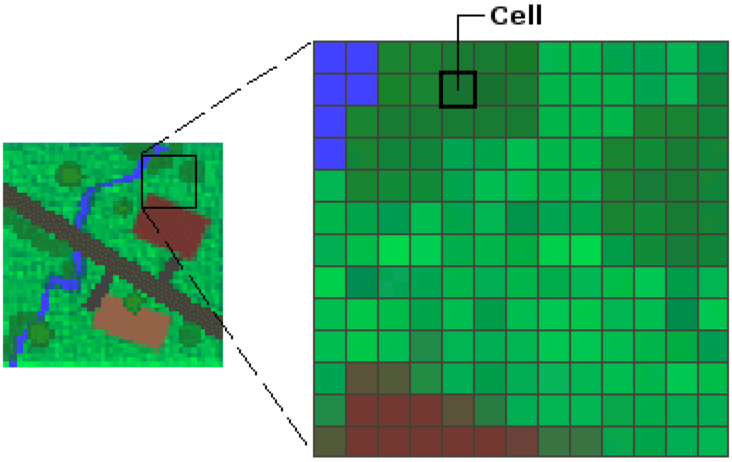

Raster data showing vegetation classification. The vegetation data was derived from NDVI classification of a satellite image.

What results from the effect of converting spatial data location information into a cell based raster format is called stairstepping. The name derives from the image of exactly that, the square cells along the borders of different value types look like a staircase viewed from the side.

Unlike vector data, raster data is formed by each cell receiving the value of the feature that dominates the cell. The stairstepping look comes from the transition of the cells from one value to another. In the image above the dark green cell represents chamise vegetation. This means that the dominate feature in that cell area was chamise vegetation. Other features such as developed land, water or other vegetation types may be present on the ground in that area. As the feature in the cell becomes more dominantly urban, the cell is attributed the value for developed land, hence the pink shading.

### *What is raster data?*

- In its simplest form, a raster consists of a matrix of cells (or pixels) organized into rows and columns (or a grid) where each cell contains a value representing information, such as temperature. Rasters are digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps.




Data stored in a raster format represents real-world phenomena, such as

- Thematic data (also known as discrete), representing features such as land-use or soils data
- Continuous data, representing phenomena such as temperature, elevation or spectral data such as satellite images and aerial photographs
- Pictures, such as scanned maps or drawings and building photographs

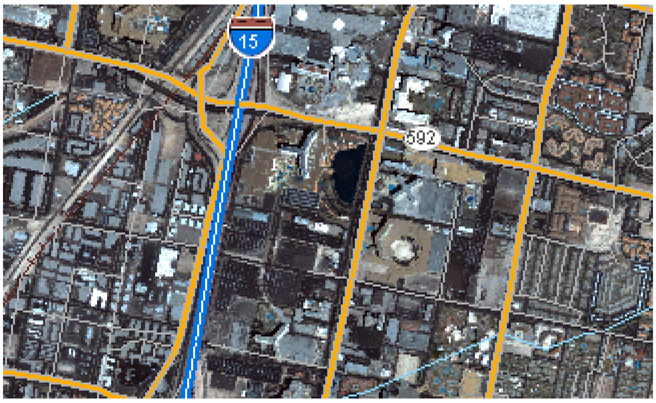
Thematic and continuous rasters may be displayed as data layers along with other geographic data on your map but are often used as the source data for spatial analysis with the ArcGIS Spatial Analyst extension. Picture rasters are often used as attributes in tables—they can be displayed with your geographic data and are used to convey additional information about map features.

While the structure of raster data is simple, it is exceptionally useful for a wide range of applications. Within a GIS, the uses of raster data fall under four main categories:

### *Rasters as basemaps*

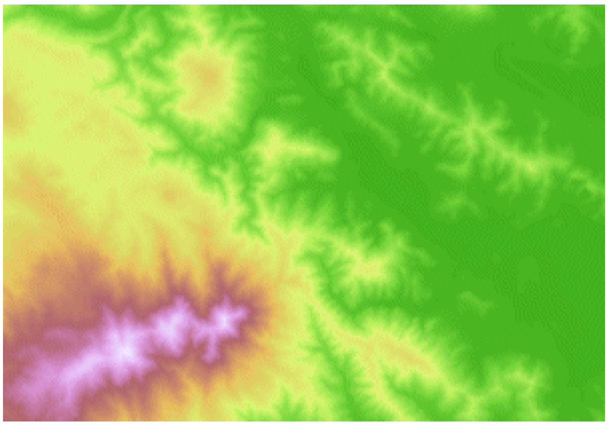
A common use of raster data in a GIS is as a background display for other feature layers. For example, orthophotographs displayed underneath other layers provide the map user with confidence that map layers are spatially aligned and represent real objects, as well as representing additional information. Three main sources of raster basemaps are orthophotos from aerial photography, satellite imagery, and scanned maps.

Below is a raster used as a basemap for road data.




### *Rasters as surface maps*

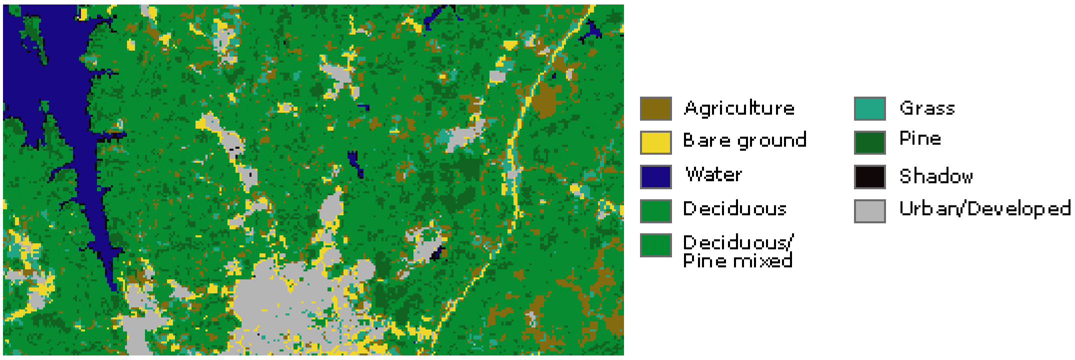
Rasters are well suited for representing data that changes continuously across a landscape (surface). They provide an effective method of storing the continuity as a surface. They also provide a regularly spaced representation of surfaces. Elevation values measured from the earth's surface are the most common application of surface maps, but other values, such as rainfall, temperature, concentration, and population density, can also define surfaces that can be spatially analyzed. The raster below displays elevation—using green to show lower elevation and red, pink, and white cells to show higher elevation.



### *Rasters as thematic maps*

Rasters representing thematic data can be derived from analyzing other data. A common analysis application is classifying a satellite image by land-cover categories. Basically, this activity groups the values of multispectral data into classes (such as vegetation type) and assigns a categorical value. Thematic maps can also result from geoprocessing operations that combine data from various sources, such as vector, raster, and terrain data. For example, you can process data through a geoprocessing model to create a raster dataset that maps suitability for a specific activity.

Below is an example of a classified raster dataset showing land use.




### *Rasters as attributes of a feature*

Rasters used as attributes of a feature may be digital photographs, scanned documents, or scanned drawings related to a geographic object or location. A parcel layer may have scanned legal documents identifying the latest transaction for that parcel, or a layer representing cave openings may have pictures of the actual cave openings associated with the point features. Below is a digital picture of a very large, old tree that could be used as an attribute to a landscape layer that a city may maintain.

### *Why store data as a raster?*

- Sometimes you don't have the choice of storing your data as a raster; for example, imagery is only available as a raster. However, there are many other features (such as points) and measurements (such as rainfall) that could be stored as a feature (vector) data type (or both).

The advantages of storing your data as a raster are

- A simple data structure—A matrix of cells with values representing a coordinate and sometimes linked to an attribute table
- A powerful format for advanced spatial and statistical analysis
- The ability to represent continuous surfaces and to perform surface analysis
- The ability to uniformly store points, lines, polygons, and surfaces
- The ability to perform fast overlays with complex datasets

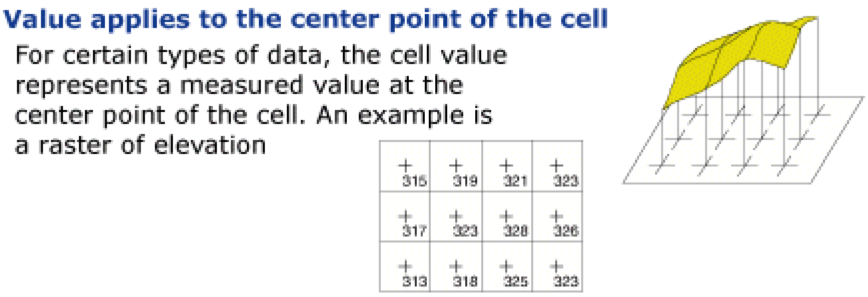
There are other considerations for storing your data as a raster that may convince you to use a vector-based storage option. For example

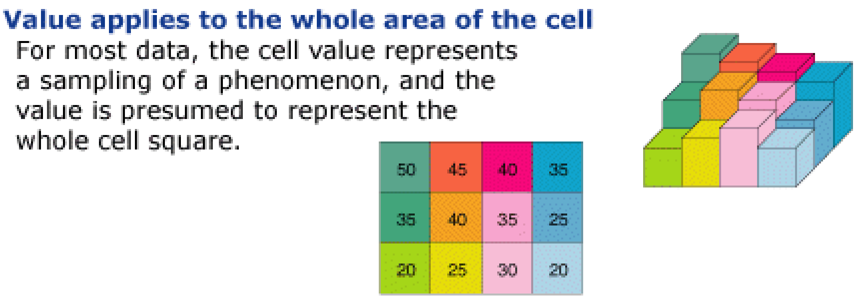
- There can be spatial inaccuracies due to the limits imposed by the raster dataset cell dimensions.
- Raster datasets are potentially very large datasets. Resolution increases as the size of the cell decreases; however, normally cost also increases in both disk space and processing speeds. For a given area, changing cells to one-half the current size requires as much as four times the storage space, depending on the type of data and storage techniques used.
- There is also a loss of precision that accompanies restructuring data to a regularly spaced raster-cell boundary.

### *General characteristics of raster data*

In raster datasets, each cell (which is also known as a pixel) has a value. The cell values represent the phenomenon portrayed by the raster dataset such as a category, magnitude, height, or spectral value. The category could be a land-use class such as grassland, forest, or road. A magnitude might represent gravity, noise pollution, or percent rainfall. Height (distance) could represent surface elevation above mean sea level, which can be used to derive slope, aspect, and watershed properties. Spectral values are used in satellite imagery and aerial photography to represent light reflectance and color.

Cell values can be either positive or negative, integer, or floating point. Integer values are best used to represent categorical (discrete) data, and floating-point values to represent continuous surfaces. For additional information on discrete and continuous data, see Discrete and continuous data. Cells can also have a NoData value to represent the absence of data.

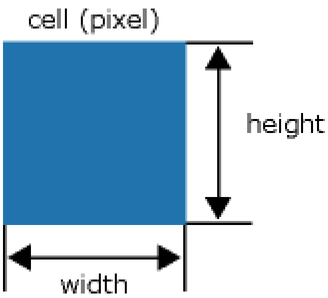


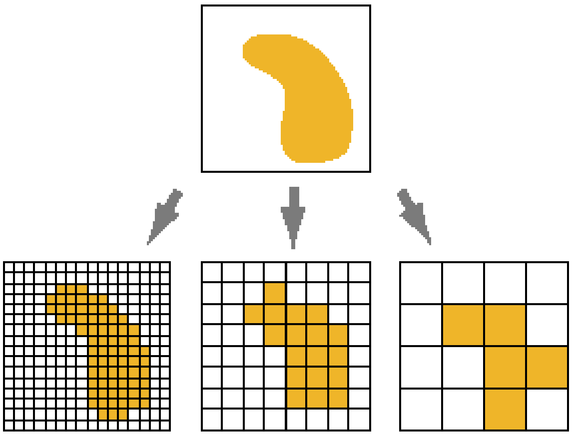

Rasters are stored as an ordered list of cell values. For example, 80, 74, 62, 45, 45, 34, and so on.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 80 | 74 | 62 | 45 | 45 | 34 | 39 | 56 |
| 80 | 74 | 74 | 62 | 45 | 34 | 39 | 56 |
| 74 | 74 | 62 | 62 | 45 | 34 | 39 | 39 |
| 62 | 62 | 45 | 45 | 34 | 34 | 34 | 39 |
| 45 | 45 | 45 | 34 | 34 | 30 | 34 | 39 |

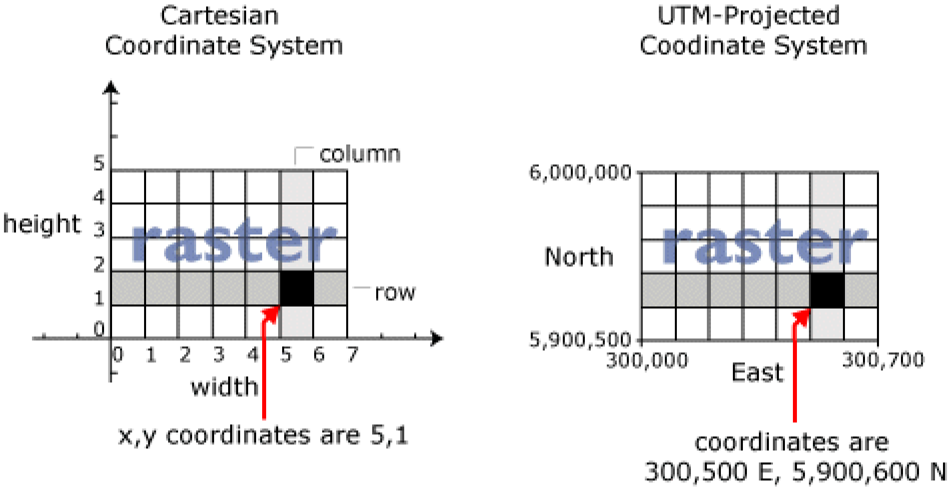
The area (or surface) represented by each cell consists of the same width and height and is an equal portion of the entire surface represented by the raster. For example, a raster representing elevation (that is, digital elevation model) may cover an area of 100 square kilometers. If there were 100 cells in this raster, each cell would represent one square kilometer of equal width and height (that is, 1 km x 1 km).



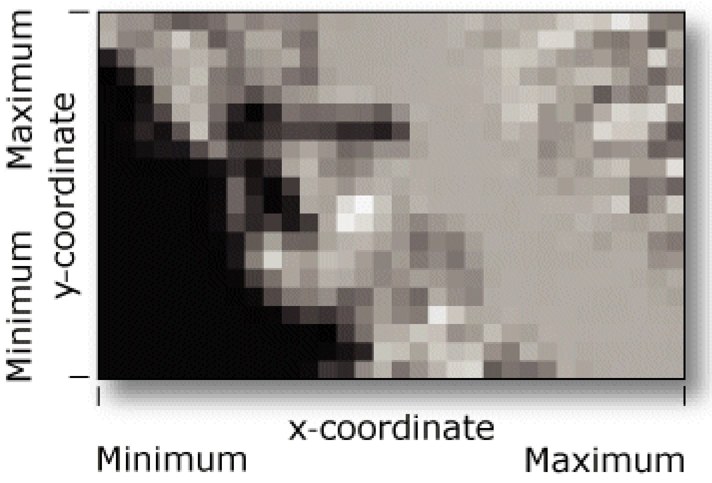

The dimension of the cells can be as large or as small as needed to represent the surface conveyed by the raster dataset and the features within the surface, such as a square kilometer, square foot, or even a square centimeter. The cell size determines how coarse or fine the patterns or features in the raster will appear. The smaller the cell size, the smoother or more detailed the raster will be. However, the greater the number of cells, the longer it will take to process and it will increase the demand for storage space. If a cell size is too large, information may be lost or subtle patterns may be obscured. For example, if the cell size is larger than the width of a road, the road may not exist within the raster dataset. In the diagram below, you can see how this simple polygon feature will be represented by a raster dataset at various cell sizes.



The location of each cell is defined by the row or column where it is located within the raster matrix. Essentially, the matrix is represented by a [Cartesian coordinate system](), in which the rows of the matrix are parallel to the x-axis and the columns to the y-axis of the Cartesian plane. Row and column values begin with 0. In the example below, if the raster is in a [Universal Transverse Mercator (UTM)]()-[projected coordinate system]() and has a cell size of 100, the cell location at 5,1 would be 300,500 East, 5,900,600 North.




Often you need to specify the extent of a raster. The extent is defined by the top, bottom, left, and right coordinates of the rectangular area covered by a raster, as shown below.




# Vector and raster Advantages and disadvantage

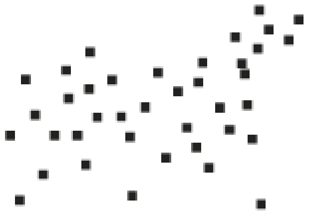
- There are some important advantages and disadvantages to using a raster or vector data model to represent reality:
- Raster datasets record a value for all points in the area covered which may require more storage space than representing data in a vector format that can store data only where needed.
- Raster data is computationally less expensive to render than vector graphics
- There are transparency and aliasing problems when overlaying multiple stacked pieces of raster images
- Vector data allows for visually smooth and easy implementation of overlay operations, especially in terms of graphics and shape-driven information like maps, routes and custom fonts, which are more difficult with raster data.
- Vector data can be displayed as vector graphics used on traditional maps, whereas raster data will appear as an image that may have a blocky appearance for object boundaries. (depending on the resolution of the raster file)
- Vector data can be easier to register, scale, and re-project, which can simplify combining vector layers from different sources.
- Vector data is more compatible with relational database environments, where they can be part of a relational table as a normal column and processed using a multitude of operators.
- Vector file sizes are usually smaller than raster data, which can be tens, hundreds or more times larger than vector data (depending on resolution).
- Vector data is simpler to update and maintain, whereas a raster image will have to be completely reproduced. (Example: a new road is added).
- Vector data allows much more analysis capability, especially for "networks" such as roads, power, rail, telecommunications, etc. (Examples: Best route, largest port, airfields connected to two-lane highways). Raster data will not have all the characteristics of the features it displays.

### *TIN concepts*

- A TIN data model is composed of nodes, edges, triangles, hull polygons, and topology.

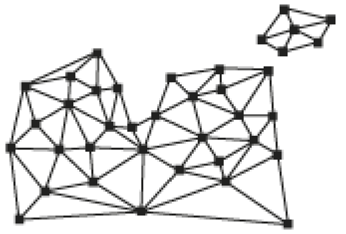
### Nodes

- Nodes are the fundamental building blocks of a TIN. The nodes originate from the points and line vertices contained in the input data sources. Every node is incorporated in the TIN triangulation. Every node in the TIN surface model must have a z-value.



### Edges

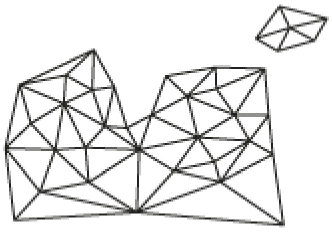
- Every node is joined with its nearest neighbors by edges to form triangles, which satisfy the Delaunay criterion. Each edge has two nodes, but a node may have two or more edges. Because edges have a node with a z-value at each end, it is possible to calculate a slope along the edge from one node to the other.



Each feature in the input data sources used to build the TIN is processed in accordance with its surface feature type. Breakline features are always maintained as edges in the TIN triangulation. These breakline TIN edges are flagged internally as either hard or soft edges.

### Triangles

- Each triangular facet describes the behavior of a portion of the TIN's surface. The x, y, and z coordinate values of a triangle's three nodes can be used to derive information about the facet, such as slope, aspect, surface area, and surface length. Considering the entire set of triangles as a whole, it is possible to derive additional information about the surface, including volume, surface profiles, and visibility analysis.

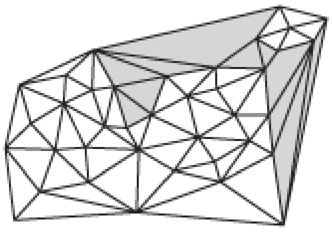


Because each facet summarizes a certain surface behavior, it is important to ensure that the sample points are selected adaptively to give the best possible surface fit. A TIN surface model can yield poor results if important regions of the surface are sampled inadequately.

## Hull

The hull of a TIN is formed by one or more polygons containing the entire set of data points used to construct the TIN. The hull polygons define the zone of interpolation of the TIN. Inside or on the edge of the hull polygons, it is possible to interpolate surface z-values, perform analysis, and generate surface displays. Outside the hull polygons, it is not possible to derive information about the surface. The hull of a TIN can be formed by one or more polygons, which can be nonconvex.

A nonconvex hull must be user-defined by including Clip and Erase exclusion features during the construction of the TIN. These features explicitly define the edge of the surface. When no exclusion features are used to define the hull, the TIN generator creates a convex hull to define the bounding edges of the TIN. A convex hull is a polygon with the property that any line connecting any two points of the TIN must itself lie inside or define the edge of the convex hull. The definition of a nonconvex hull is essential to prevent the generation of erroneous information in regions of the TIN outside the actual dataset but inside the convex hull. Consider the diagram below.



Without the use of clip features, incorrect values may be interpolated in the shaded regions.

## Topology

- The topological structure of a TIN is defined by maintaining information defining each triangle's nodes, edge numbers, type, and adjacency to other triangles. For each triangle, a TIN records:
- The triangle number
- The numbers of each adjacent triangle
- The three nodes defining the triangle
- The x,y coordinates of each node
- The surface z-value of each node
- The edge type of each triangle edge (hard or soft)

In addition, the TIN maintains a list of all the edges that form the TIN's hull and information defining the TIN's projection and units of measure.

### *How a TIN is stored*

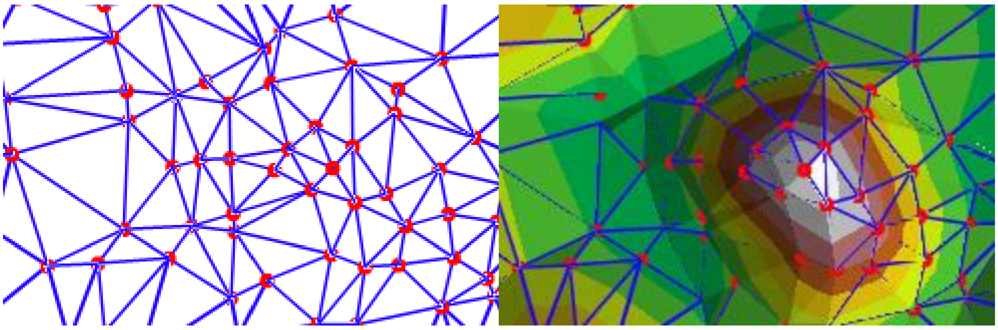
Like a coverage, a TIN is stored as a directory of files. Note, however, that a TIN is not a coverage and has no associated INFO files. A TIN directory contains seven files containing information about the TIN surface. These files are encoded in binary format and are not readable by standard text display or editing programs.

### *About TIN surfaces*

Triangular Irregular Networks (TIN) have been used by the GIS community for many years and are a digital means to represent surface morphology. TINs are a form of vector based digital geographic data and are constructed by triangulating a set of vertices (points). The vertices are connected with a series of edges to form a network of triangles. There are different methods of interpolation to form these triangles, such as Delaunay triangulation or distance ordering. ArcGIS supports the Delaunay triangulation method.

The resulting triangulation satisfies the Delaunay triangle criterion, which ensures that no vertex lies within the interior of any of the circumcircles of the triangles in the network. If the Delaunay criterion is satisfied everywhere on the TIN, the minimum interior angle of all triangles is maximized. The result is that long, thin triangles are avoided as much as possible.

The edges of TINs form contiguous, nonoverlapping triangular facets and can be used to capture the position of linear features that play an important role in a surface, such as ridgelines or stream courses. The graphics below show the nodes and edges of a TIN (left) and the nodes, edges, and faces of a TIN (right).



Because nodes can be placed irregularly over a surface, TINs can have a higher resolution in areas where a surface is highly variable or where more detail is desired and a lower resolution in areas that are less variable.

The input features used to create a TIN remain in the same position as the nodes or edges in the TIN. This allows a TIN to preserve all the precision of the input data while simultaneously modeling the values between known points. You can include precisely located features on a surface—such as mountain peaks, roads, and streams—by using them as input features to the TIN nodes.

TIN models are less widely available than raster surface models and tend to be more expensive to build and process. The cost of obtaining good source data can be high, and processing TINs tends to be less efficient than processing raster data because of the complex data structure.

TINs are typically used for high-precision modeling of smaller areas, such as in engineering applications, where they are useful because they allow calculations of planimetric area, surface area, and volume.

## 3.5. SPATIAL DATABASE DESIGN WITH THE CONCEPT OF GEO-DATABASE

# SPATIAL DATABASES

## Concept, Design and Management

A spatial database system may be defined as a database system that offers spatial data types in its data model and query language, and supports spatial data types in its implementation, providing at least spatial indexing and spatial join methods.

Spatial database systems offer the underlying database technology for geographic information systems and other applications. We survey data modeling, querying, data structures and algorithms, and system architecture for such systems. The emphasis is on describing known technology in a coherent manner, rather than listing open problems.

## Spatial Database Concept

In various fields there is a need to manage *geometric*, *geographic*, or *spatial* data, which means data related to *space*. The space of interest can be, for example, the two-dimensional abstraction of (parts of) the surface of the earth or a 3d-space representing a digital terrain model. At least since the advent of relational database systems there have been attempts to manage such data in database systems.

Characteristic for the technology emerging to address these needs is the capability to deal with *large collections of relatively simple geometric objects*, for example, a set of 100 000 polygons. Several terms have been used for database systems offering such support like *pictorial*, *image, geometric, geographic*, or spatial database *system*. The terms “pictorial” and “image” database system arise from the fact that the data to be managed are often initially captured in the form of digital raster images (e.g. remote sensing by satellites, or computer tomography in medical applications).

The term “spatial database system” has become popular during the last few years, and is associated with a view of a database as containing sets of objects in space rather than images or pictures of a space. Indeed, the requirements and techniques for dealing with objects in space that have identity and well-defined extents, locations, and relationships are rather different from those for dealing with raster images.

A spatial database therefore has the following characteristics:

a. A spatial database system is a database system.
b. It offers spatial data types (SDTs) in its data model and query language.
c. It supports spatial data types in its implementation, providing at least spatial indexing and efficient algorithms for spatial join.

Nobody cares about a special purpose system that is not able to handle all the standard data modeling and querying tasks. Hence a spatial database system is a full-fledged database system with *additional* capabilities for handling spatial data. Therefore spatial indexing is mandatory. It should also support connecting objects from different classes through some spatial relationship.

## Spatial Database Design

- A spatial database includes collections of information about the spatial location, relationship and shape of topological geographic features and the data in the form of attributes. The design of the spatial database is the formal process of analyzing facts about the real world into a structured model. Database design is characterized by the following phases: requirement analysis, logical design and physical design. In other words, you basically need a plan, a design layout and then the data to complete the process.
- Having a solid well designed spatial database is the key to performing good Spatial Analysis. The database can be complex and designed with expensive sophisticated software or can be merely a simple well organized collection of data that can be utilized in a geographic form.
- Three main categories of spatial modeling functions that can be applied to geographic features within a GIS are: (1) geometric models, such as calculating the Euclidean distance between features, generating buffers, calculating areas and perimeters, and so on; (2) coincidence models, such as topological overlay; and (3) adjacency models (path finding, redistricting, and allocation). All three model categories support operations on spatial data such as points, lines, polygons, tins, and grids. Functions are organized in a sequence of steps to derive the desired information for analysis.
- Almost all entities of geographic reality have at least a 3-dimensional spatial character, but not all dimensions may be needed. E.g. a highway pavement actually has a depth which might be important, but is not as important as the width, which is not as important as the length. Representation should be based on the types of manipulations that might be undertaken. Map-scale of the source document is important in constraining the level of detail represented in a database. E.g. on a 1:100,000 map individual houses or fields are not visible

### Steps in database design

1. Conceptual
    a. software and hardware independent
    b. describes and defines included entities
    c. identifies how entities will be represented in the database
    d. i.e. selection of spatial objects - points, lines, areas, raster cells
    e. requires decisions about how real-world dimensionality and relationships will be represented

f. these can be based on the processing that will be done on these objects
g. e.g. should a building be represented as an area or a point?
h. e.g. should highway segments be explicitly linked in the database?

2. Logical
    a. software specific but hardware independent
    b. sets out the logical structure of the database elements, determined by the data base management system used by the software
3. Physical
    a. both hardware and software specific
    b. requires consideration of how files will be structured for access from the disk

### Characteristics of a Good Database Design

In order that the GIS database provides the best service it should be:

- Contemporaneous – the data should be updated regularly so as to yield information that pertains to the same time-frame for all its measured variables
- Flexible and extensible so that additional datasets may be added as necessary for the intended applications
    - the categories of information and subcategories within them should contain all of the data needed to analyze or model the behavior of the resource using conventional methods and models
- Positionally accurate – if for example the boundary between the residential and agricultural land has changed, this may be incorporated with ease.
- Exactly compatible with other information that may be overlain with it
- Internally accurate, portraying the nature of phenomena without error - requires clear definitions of phenomena that are included
- Readily updated on a regular schedule
- Accessible to whoever needs it

## Spatial Database Management

Many factors influence a successful Geographic Information System (GIS) implementation. None however are more fundamental than having the right management strategies and software to implement these. The spatial database is the foundation by which all data is uniformly created and converted. But maintaining the integrity and currency of the data is of fundamental importance. A classic mistake made by many organizations is thinking that a generic spatial database design will be sufficient for their needs. That is simply not the case. The spatial database is the end result of a series of processes that determine the specific functional requirements for the user and the key applications. Interoperability of data is also a critical area of concern in the development of spatial data information systems. As we move from newly created data to assimilation of all existing data, a properly designed spatial database is insurance for end user success.

A good spatial database management software package should be able to:

1. Scale and rotate coordinate values for "best fit" projection overlays and changes.
2. Convert (interchange) between polygon and grid formats.
3. Permit rapid updating, allowing data changes with relative ease.
4. Allow for multiple users and multiple interactions between compatible data bases.
5. Retrieve, transform, and combine data elements efficiently.
6. Search, identify, and route a variety of different data items and score these values with assigned weighted values, to facilitate proximity and routing analysis.
7. Perform statistical analysis, such as multivariate regression, correlations, etc.
8. Overlay one file variable onto another, i.e., map super positioning.
9. Measure area, distance, and association between points and fields.
10. Model and simulate, and formulate predictive scenarios, in a fashion that allows for direct interactions between the user group and the computer program.

### *An overview of geodatabase design*

Geodatabase design is based on a common set of fundamental GIS design steps, so it's important to have a basic understanding of these GIS design goals and methods. This section provides an overview.

GIS design involves organizing geographic information into a series of data themes—layers that can be integrated using geographic location. So it makes sense that geodatabase design begins by identifying the data themes to be used, then specifying the contents and representations of each thematic layer.

This involves defining

- How the geographic features are to be represented for each theme (for example, as points, lines, polygons, or rasters) along with their tabular attributes
- How the data will be organized into datasets such as feature classes, attributes, raster datasets, and so forth
- What additional spatial and database elements will be needed for integrity rules, for implementing rich GIS behavior (such as topologies, networks, and raster catalogs), and for defining spatial and attribute relationships between datasets

### *Representation*

Each GIS database design is based on deciding what the geographic representations will be for each dataset. Individual geographic entities can be represented as

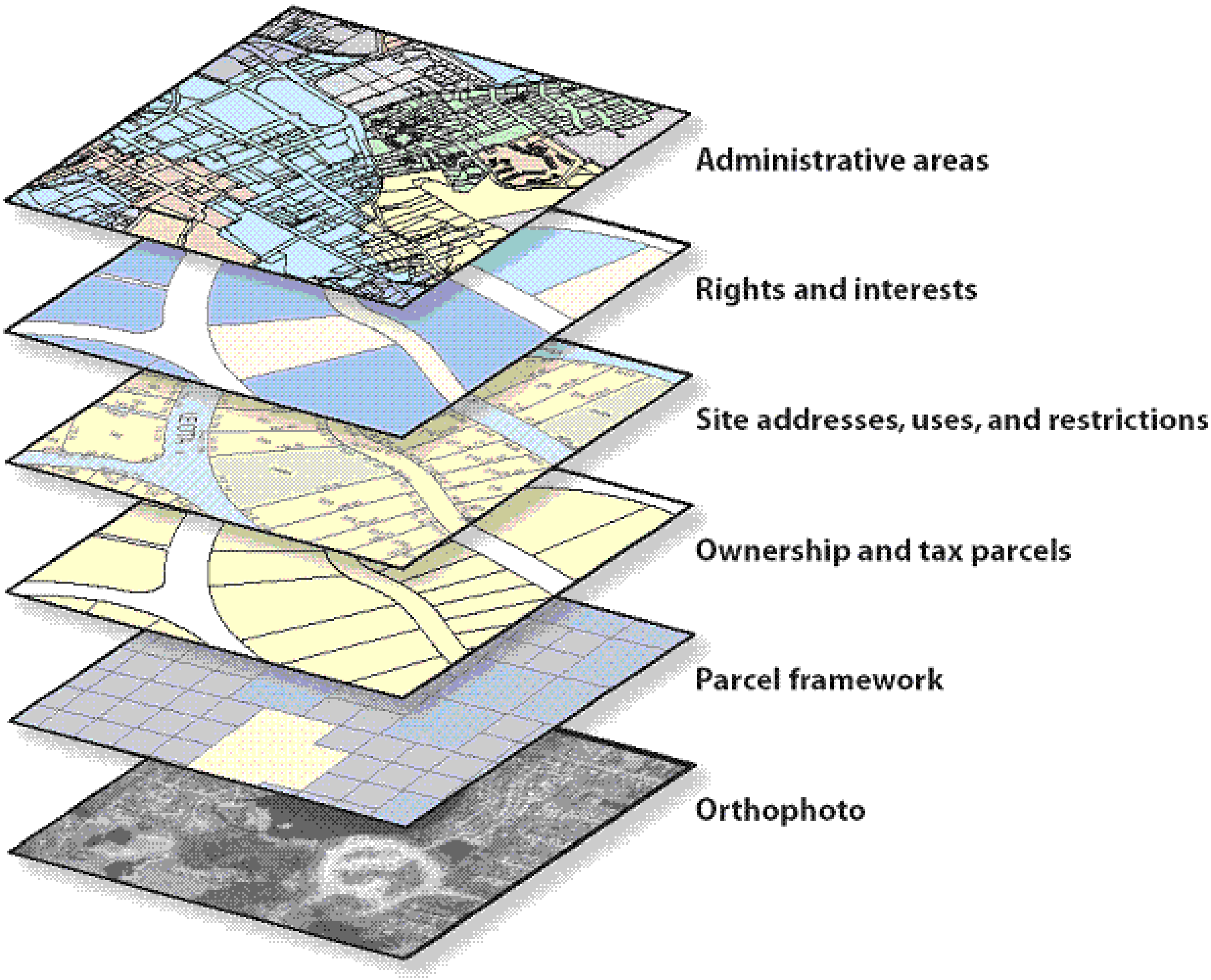
- Feature classes (sets of points, lines, and polygons)
- Imagery and rasters
- Continuous surfaces that can be represented using features (such as contours), rasters (Digital Elevation Models [DEM]), or triangulated irregular networks (TINs)
- Attribute tables for descriptive data

### *Data themes*

Geographic representations are organized in a series of data themes (sometimes referred to as *thematic layers*). A key concept in a GIS is one of data layers or themes. A data theme is a collection of common geographic elements such as a road network, a collection of parcel boundaries, soil types, an elevation surface, satellite imagery for a certain date, well locations, and so on.

The concept of a thematic layer was one of the early notions in GIS. Practitioners thought about how the geographic information in maps could be partitioned into logical information layers—as more than a random collection of individual objects (such as a road, a bridge, a hill, a house, a peninsula). These early GIS users organized information in thematic layers that described the distribution of a phenomenon and how it should be portrayed across a geographic extent. These layers also provided a protocol (capture rules) for collecting the representations (as feature sets, raster layers, attribute tables, and so on).

In GIS, thematic layers are one of the main organizing principles for GIS database design.



Each GIS will contain multiple themes for a common geographic area. The collection of themes acts as layers in a stack. Each theme can be managed as an information set independent of other themes. Each has its own representations (points, lines, polygons, surfaces, rasters, and so on). Because the various independent themes are spatially referenced, they overlay one another and can be combined in a common map display. Plus, GIS analysis operations such as overlay can fuse information between themes.

### *GIS datasets are collections of representations for a data theme*

Geographic data collections can be represented as feature classes and raster-based datasets in a GIS database.

Many themes are represented by a single collection of homogeneous features such as a feature class of soil type polygons and a point feature class of well locations. Other themes, such as a transportation framework, are represented by multiple datasets (such as a set of spatially related feature classes for streets, intersections, bridges and highway ramps and so on).

Raster datasets are used to represent continuous surfaces, such as elevation, slope, and aspect, as well as to hold satellite imagery, aerial photography, and other gridded datasets (such as land cover and vegetation types).

Both the intended use and existing data sources influence spatial representations in a GIS. When designing a GIS database, users have a set of applications in mind. They understand what questions will be asked of the GIS. Defining these uses helps to determine the content specification for each theme and how each is to be represented geographically. For example, there are numerous alternatives for representing surface elevation: as contour lines and spot height locations (such as hilltops, peaks), as a continuous terrain surface (a TIN), or as shaded relief. Any or all of these may be relevant for each particular GIS database design. The intended uses of the data will help to determine which of these representations will be required.

Frequently, the geographic representations will be predetermined to some degree by the available data sources for the theme. If a pre-existing data source was collected at a particular scale and representation, it will often be necessary to adapt your design to use it.

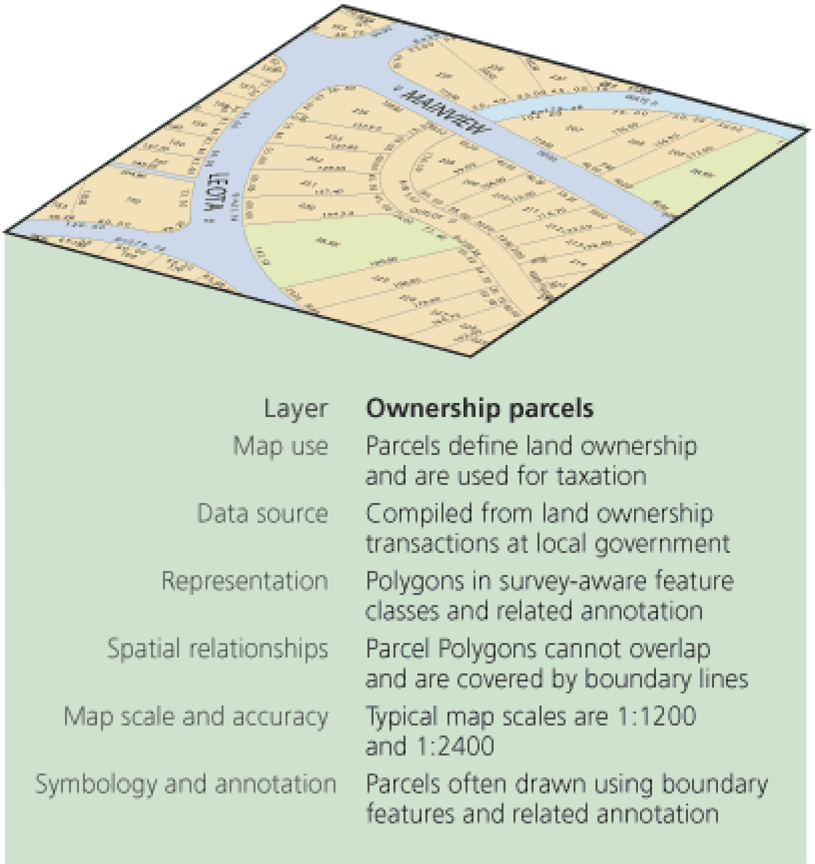
### *Geodatabase design steps*

Design starts with thematic layers.

First, you identify the thematic layers you'll need for your particular applications and information requirements. Then, you define each thematic layer in more detail. The characterization of each thematic layer will result in a specification of standard geodatabase data elements such as feature classes, tables, relationship classes, raster datasets, subtypes, topologies, domains, and so on.

When identifying thematic layers in your design, try to characterize each theme in terms of its visual representations, its expected uses in the GIS, its likely data sources, and its levels of resolution. For example, at what scales and extents will you need to use this information, and how will its elements be represented at each scale? These characteristics help describe the high-level contents expected from each theme.

Here is an example description of a data theme for ownership parcels in a cadastral application.




Once you have identified the key thematic layers in your design, the next step is to develop specifications for representing the contents of each thematic layer in the physical database.

- List the map scales and extents that you'll need to work with.
- For each, describe how the geographic features are to be represented (for example, as points, lines, polygons, rasters, or tabular attributes).
- How should the data be organized into feature classes, tables, and relationships?
- How will spatial and database integrity rules be used to implement GIS behavior?

The 11 steps presented below outline a general GIS database design process. The initial design steps 1 through 3 help you to identify and characterize each thematic layer. In steps 4 through 7, you begin to develop representation specifications, relationships, and ultimately, geodatabase elements and their properties. In steps 8 and 9, you will define the data capture procedures and assign data collection responsibilities. In the final stage (steps 10 and 11), you will test and refine your design through a series of initial implementations. You will also document your design.

**Eleven steps to geodatabase design**

1. **Identify the information products that you will create and manage with your GIS.**
Your GIS database design should reflect the work of your organization. Consider compiling and maintaining an inventory of map products, analytical models, Web mapping applications, data flows, database reports, key responsibilities, 3D views, and other mission-based requirements for

your organization. List the data sources you currently use in this work. Use these to drive your data design needs.

**2. Identify the key data themes based on your information requirements.**

Define more completely some of the key aspects of each data theme. Determine how each dataset will be used—for editing, for GIS modeling and analysis, representing your business workflows, and for mapping and 3D display. Specify the map use, the data sources, the spatial representations for each specified map scale; data accuracy and collection guidelines for each map view and 3D view; how the theme is displayed, its symbology, text labels, and annotation. Consider how each map layer will be displayed in an integrated fashion with other key layers. For modeling and analysis, consider how information will be used with other datasets (for example, how they are combined and integrated). This will help you to identify some key spatial relationships and data integrity rules. Ensure that these display and analysis properties are considered as part of your database design.

**3. Specify the scale ranges and spatial representations of each data theme at each scale.** Data is compiled for use at a specific range of map scales. Associate your geographic representation for each map scale. Geographic representation will often change between map scales (for example, from polygon to line or point). In many cases, you may need to generalize the feature representations for use at smaller scales. Rasters can be resampled using image pyramids.

**4. Decompose each representation into one or more geographic datasets.**

Discrete features are modeled as feature classes of points, lines, and polygons. You can consider advanced data types such as topologies, networks, and terrains to model the relationships between elements in a layer as well as across datasets.

For raster datasets, mosaics and catalog collections are options for managing very large collections.

Surfaces can be modeled using features, such as contours, as well as using rasters and terrains.

**5. Define the tabular database structure and behavior for descriptive attributes.**

Identify attribute fields and column types. Tables also might include attribute domains, relationships, and subtypes. Define any valid values, attribute ranges, and classifications (for use as domains). Use subtypes to control behaviors. Identify tabular relationships and associations for relationship classes.

**6. Define the spatial behavior and integrity rules for your datasets.**

For features, you can add behavior and capabilities for a number of purposes using topologies, address locators, networks, terrains, and so on. For example, use topologies to model the spatial relationships of shared geometry and to enforce integrity rules. Use address locators to support geocoding. For rasters, you can decide if you need a raster dataset or a raster catalog.

**7. Propose a geodatabase design.**

Define the set of geodatabase elements you want in your design for each data theme. Study existing designs for ideas and approaches that work.

**8 Design editing workflows and map display properties.**

Define the editing procedures and integrity rules (for example, all streets are split where they intersect other streets and street segments connect at endpoints).

Design editing workflows that help you to meet these integrity rules for your data. Define display properties for maps and 3D views. These will be used to define map layers.

**9. Assign responsibilities for building and maintaining each data layer.**

Determine who will be assigned the data maintenance work within your organization or assigned to other organizations. Understanding these roles is important. You will need to design how data conversion and transformation is used to import and export data from your partner organizations.

**10. Build a working prototype and review and refine your design**

Test your prototype design. Build a sample geodatabase copy of your proposed design using a file, personal, or ArcSDE Personal geodatabase. Build maps, run key applications, and perform editing operations to test the design's utility. Based on your prototype test results, revise and refine your design. Once you have a working schema, load a larger set of data (such as loading it into an ArcSDE geodatabase) to check out production, performance, scalability, and data management workflows. This is an important step. Settle on your design **before** you begin to populate your geodatabase.

**11. Document your geodatabase design.**

Documenting your geodatabase design is important. Various methods can be used to describe your database design and decisions. Use drawings, map layer examples, schema diagrams, simple reports, and metadata documents.

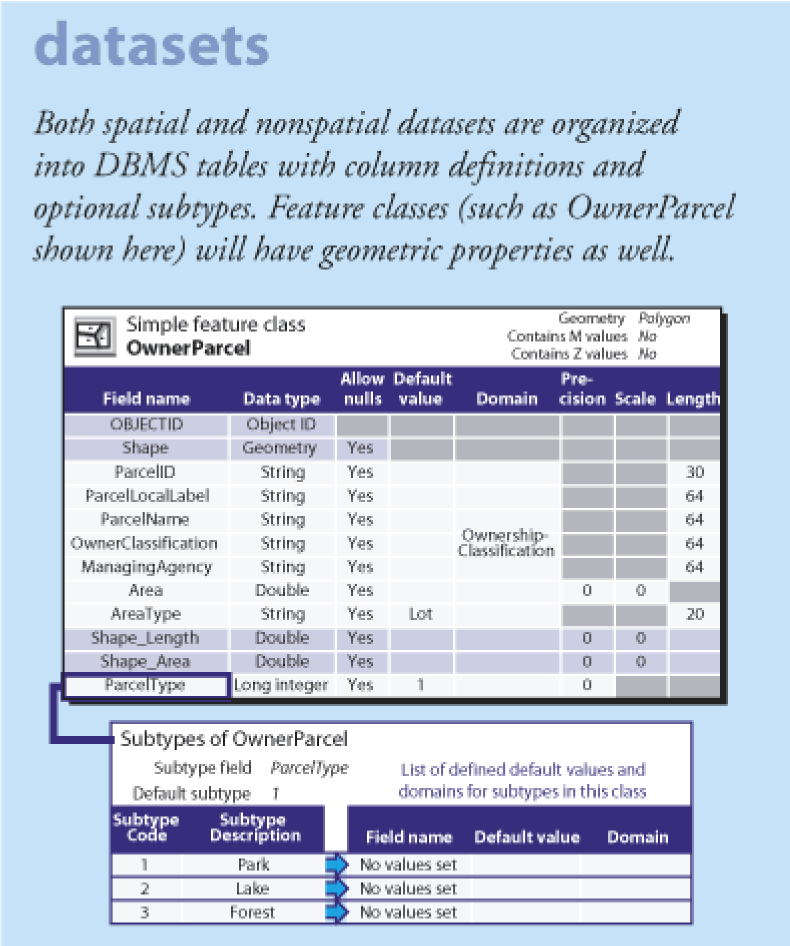
Some users like using UML. However, UML is not sufficient on its own. UML cannot represent all the geographic properties and decisions to be made. Also, UML does not convey the key GIS design concepts such as thematic organization, topology rules, and network connectivity. UML provides no spatial insight into your design.

### *Documenting your geodatabase design*

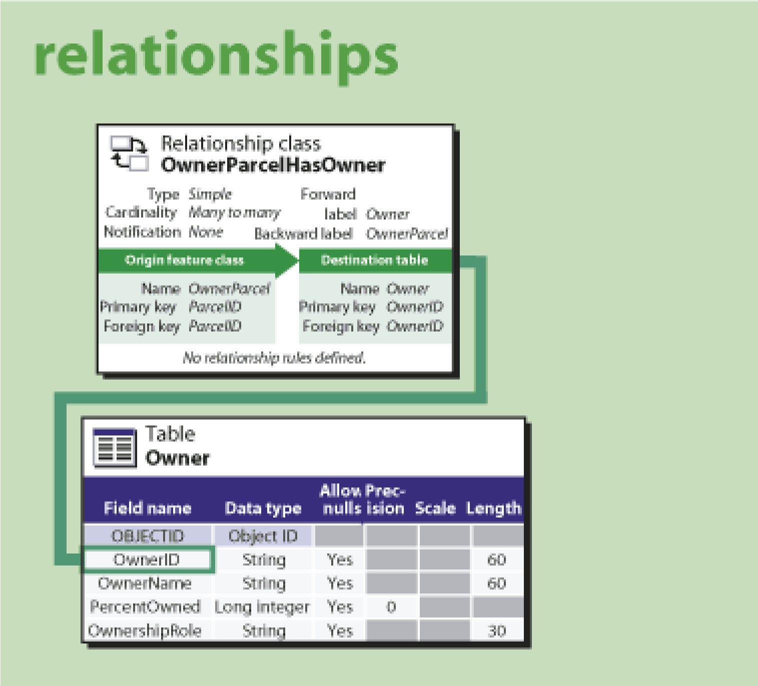
Documenting your geodatabase design is important. At the ArcGIS data models Web site (http://support.esri.com/datamodels), a series of diagrams is used to represent the key design concepts and to document the specifications of geodatabase elements, metadata, and map layers in each of the data model templates. This section provides a short overview for how various geodatabase elements are presented at the Web site and may be helpful as you document your own designs.

There are five key elements to represent the contents of your geodatabase design. These include:

1. **Datasets**—These are specifications for how to record the properties of feature classes, rasters, and attribute tables as well as the set of columns in each table. For spatial representations you'll see some geometric properties (such as point, line, and polygon and types of coordinates). Often, you'll see a specification for subtypes. These parts of the schema diagram are always shown in blue.

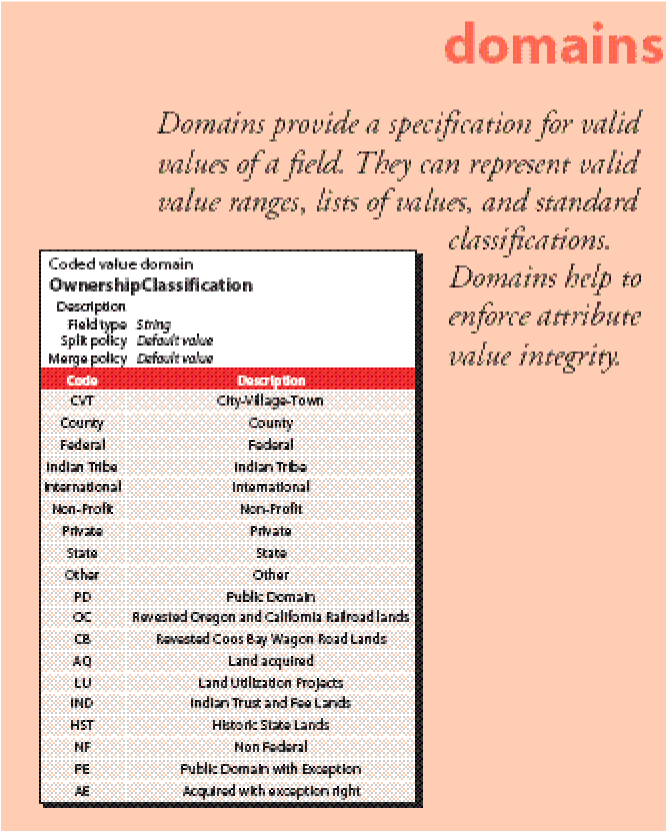



2. **Relationship classes**—Attribute relationships are widely used in GIS, just as they are in all DBMS applications. They define how rows in one table can be associated with rows in another table. Relationships have a direction of cardinality and other properties (for example, is this a one-to-one, one-to-many, or many-to-many relationship?). Relationships and their properties are shown in green.



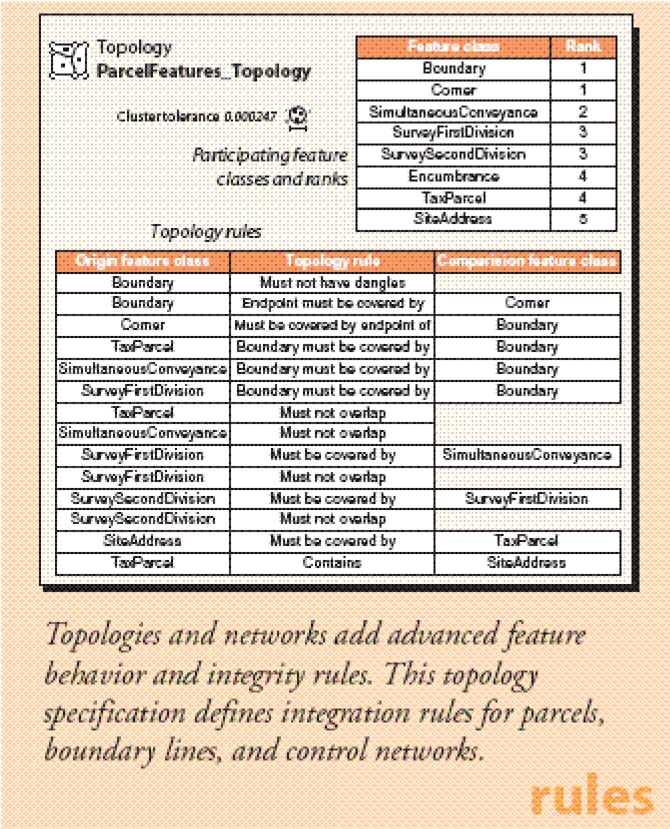

*Relationships associate rows in one table to rows in another table. This is a common relational database modeling technique.*

3. **Domains**—These represent the list or range of valid values for attribute columns. These rules control how the software maintains data integrity in certain attribute columns. Domains are shown in red.




4. **Spatial Rules**—A number of advanced data modeling capabilities are available for geodatabases. For example, data elements, such as topologies and their properties, are used to model how features share geometry with other features. Topologies, along with network datasets, address locators, terrains, cartographic representations, geometric networks, and many other advanced geodatabase types, provide a very critical and widely used GIS mechanism to enable spatial behaviors and to enforce integrity in GIS databases. These and other rules, such as networks, are shown in orange.

The best way to think about how to document and describe the set of extended data types in the geodatabase is to describe their rules and behaviors. The following is an example of how a topology can be documented:

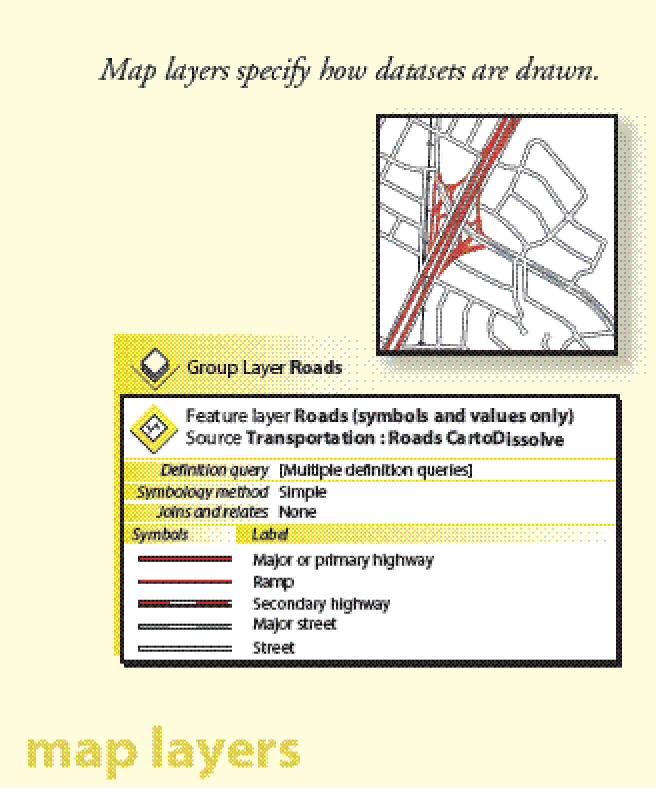



| Feature class | Rank |
|---|---|
| Boundary | 1 |
| Corner | 1 |
| SimultaneousConveyance | 2 |
| SurveyFirstDivision | 3 |
| SurveySecondDivision | 3 |
| Encumbrance | 4 |
| TaxParcel | 4 |
| SiteAddress | 5 |

| Origin feature class | Topology rule | Comparison feature class |
|---|---|---|
| Boundary | Must not have dangles | |
| Boundary | Endpoint must be covered by | Corner |
| Corner | Must be covered by endpoint of | Boundary |
| TaxParcel | Boundary must be covered by | Boundary |
| SimultaneousConveyance | Boundary must be covered by | Boundary |
| SurveyFirstDivision | Boundary must be covered by | Boundary |
| TaxParcel | Must not overlap | |
| SimultaneousConveyance | Must not overlap | |
| SurveyFirstDivision | Must be covered by | SimultaneousConveyance |
| SurveyFirstDivision | Must not overlap | |
| SurveySecondDivision | Must be covered by | SurveyFirstDivision |
| SurveySecondDivision | Must not overlap | |
| SiteAddress | Must be covered by | TaxParcel |
| TaxParcel | Contains | SiteAddress |

*Topologies and networks add advanced feature behavior and integrity rules. This topology specification defines integration rules for parcels, boundary lines, and control networks.*

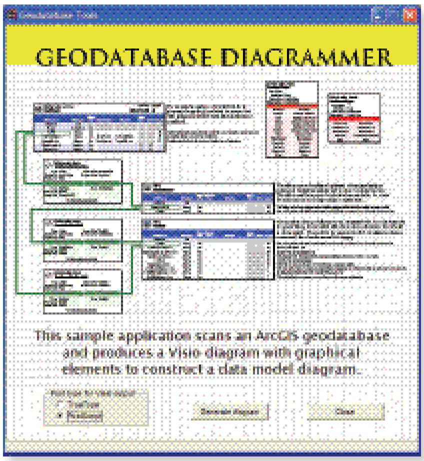
5. **Map Layers**—GIS includes interactive maps and other views. A critical part of each dataset is the specification for how it is symbolized and rendered in maps. These are typically defined as layer properties in ArcMap, which specify how features are assigned map symbology (colors, fill patterns, line and point symbols) and text labels. Layers are not managed in geodatabases but are an important aspect in helping to define some key dataset properties in a geodatabase schema. Layer specifications are shown in yellow. Layers can be stored as .lyr files or as elements in an ArcMap document (.mxd). See [Adding layers to a map]() for more information on map layers.



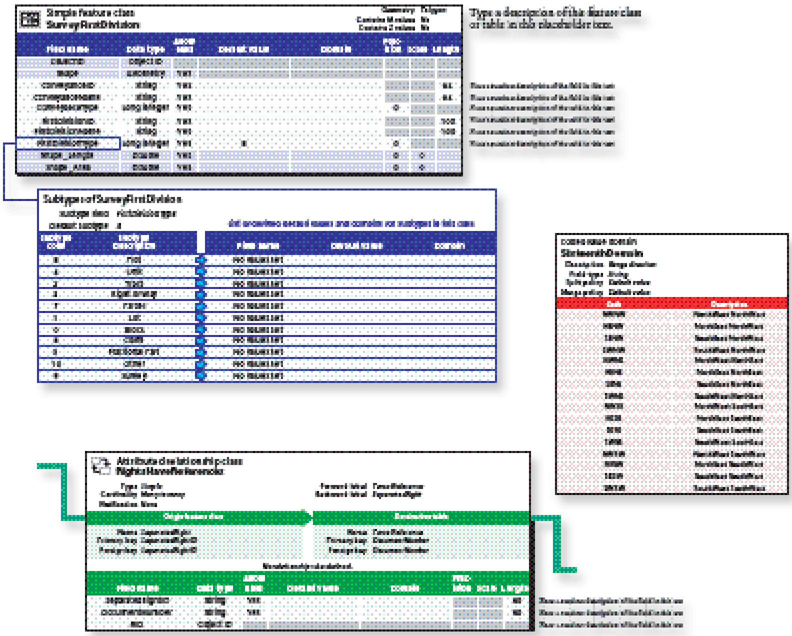

### *Using Microsoft Visio and the Geodatabase Diagrammer tool*

ESRI provides a diagramming utility as a download for users who want to generate graphics similar to these for their geodatabase designs. You can download a tool, Geodatabase Diagrammer, that will generate a series of Visio graphics of your datasets and elements in your geodatabase. Search for "Geodatabase Diagrammer" at http://arcscripts.esri.com.

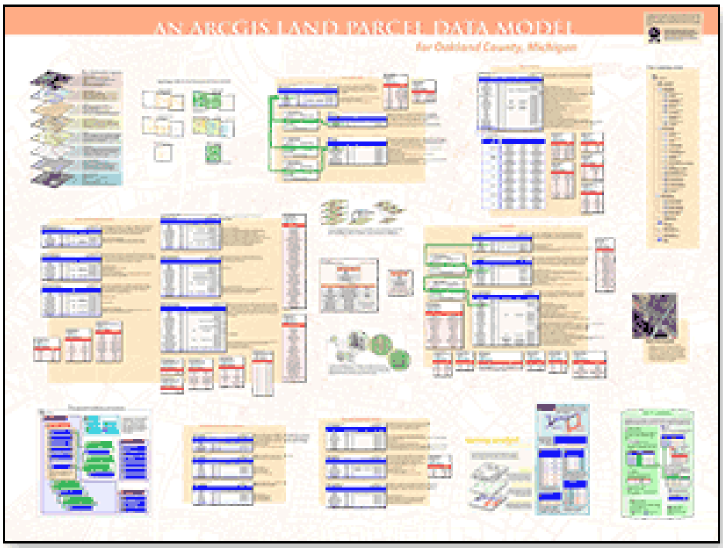
This tool is used to create graphical elements in Visio of your geodatabase contents. You can easily cut and paste graphics from Visio into Microsoft Word, PowerPoint, and any application that accepts .wmf files.




In ArcCatalog, select a geodatabase and start the Geodatabase Diagrammer command.



Geodatabase Diagrammer launches Microsoft Visio and interrorgates the structure of the geodatabase. Schema graphics are generated that you can insert in your own documentation.




You will derive great benefit from diagramming your data model—the active participation of stakeholders in examining your data model poster, the perspective you will gain on data model patterns, and the assurance that your schema is correct. The act of diagramming drives changes that will improve your data model.

### *Documenting additional properties of your geodatabase design*

Other key properties of your geodatabase design should be considered and documented including

- The definition of your coordinate system and spatial properties for each dataset

This includes such properties as the map projection, the coordinate system, spheroid, datum, x,y units, vertical coordinate system, and the use of z and m properties.

- Key tolerances and the coordinate resolution for each dataset
- The data sources and data compilation workflows

This includes translation scripts, geoprocessing and transformation models, and the workflows used to build and maintain the dataset.

- Metadata documentation for each dataset

### *Modeling feature classes*

- The following are some useful design tips for modeling geodatabase feature classes:

## Task 1. Design simple feature classes.

Almost without exception, every geodatabase will contain feature classes. You may want only a simple geodatabase design that contains just a collection of feature classes. However, most users will find the need to develop a more comprehensive data model that adds advanced geodatabase elements. You will make the decision to extend your simple feature class designs based on your system needs and goals; you'll extend your design to support essential GIS functionality and behavior. This section introduces many of these feature class capabilities and points you to help topics where you can get more information on each option.

Start by defining the common properties of simple feature classes. You can add to this later as needed, but focus on defining your basic design first.

A feature class is a collection of geographic features with the same geometry type (such as point, line, or polygon), a common set of attribute columns, and the same coordinate system.

**Example Feature Classes in ArcGIS**

| Feature Class | Representation | Notes |
|---|---|---|
| Street centerlines | Line | Street segments split at each intersection; usually contain address ranges and network properties |
| Wells | Point | |
| Soil types | Polygon | Usually have many descriptive attributes in related tables |
| Parcels | Polygon | Topologically integrated with parcel boundaries and corners |
| Parcel boundaries | Line | Has coordinate geometry and dimension attributes; participates in a topology with parcels and corners |
| Parcel corners | Point | Surveyed corners of parcels; participates in a topology with parcel polygons and boundaries |
| Parcel annotation | Annotation | Provides text labels for lot dimensions, taxation, and legal description information |
| Building footprints | Polygon | Contains outlines of buildings and structures |

Once you settle on a proposed list of feature classes, try to define the following for each:

- Pick a geometry type (also known as the feature class type) such as point, line, polygon, or annotation. You'll need to use one common geometry type for all the features in each feature class. See Feature class basics.
- Determine the attribute fields and column types. See Geodatabase field data types.
- Determine geometry properties. Will you have z-coordinates? M-coordinates? What kind of coordinate resolution? What kinds of line segments for lines and polygon feature classes? Most often, users only need the default, which uses simple, straight-line segments. However, sometimes you might need curved segments, for example, to represent cul-de-sacs and roads. See Feature class basics.
- Define the coordinate system for each feature class. See An overview of map projections.

- Do you need to use this dataset at multiple scales? How will the representations change at each map scale? You may find that you'll need alternative feature class representations for use at other scale ranges. In these cases, you can consider additional feature classes for representing the same data theme for each scale range.

Sometimes, you'll load feature data "as is" into your GIS. If this is the case, you may not need to do any of the following additional design tasks. However, it is important to evaluate the advantages of adding further GIS capabilities to the features in your geodatabase. These additional capabilities can potentially make data use and maintenance much easier in the long term. They will help you to maintain the integrity of your spatial information; will help in many ways for your data use; and, most important, will help you to understand how much confidence you can place in your data to meet your needs.

Some common reasons for extending your simple features data model:

- If you need to validate a dataset before you import and use it in your system (for example, to ensure the dataset adheres to a series of spatial integrity rules)
- If you will need to edit the data and maintain its spatial integrity
- If you want to use the feature class for advanced GIS work such as modeling and analysis

## Task 2. Organize related feature classes into feature datasets.

Use **feature datasets** to organize spatially related feature classes into a common feature dataset. Feature datasets are necessary if you want to

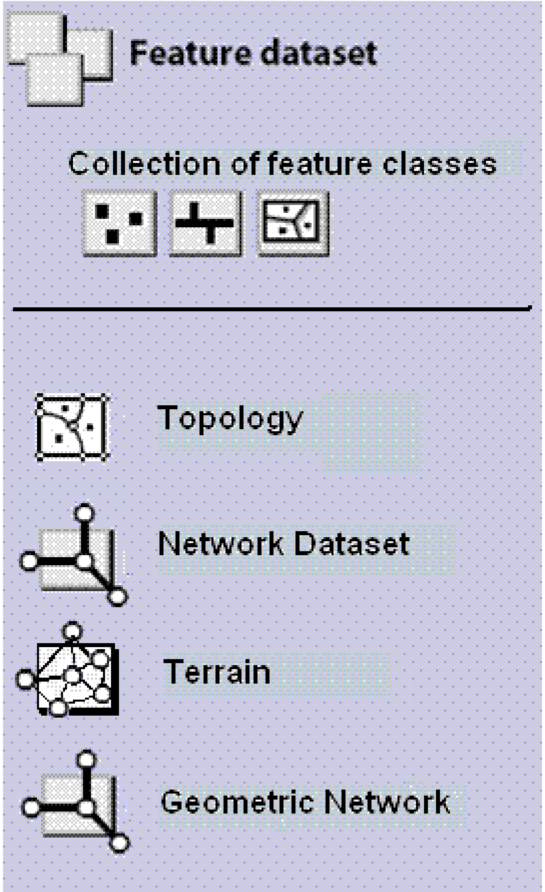
- Add a topology.
- Add a network dataset (must have Network Analyst to use)
- Add a geometric network
- Add a terrain dataset

A feature dataset is a collection of spatially or thematically related feature classes that share a common coordinate system. Feature datasets are used to hold feature classes that participate in a shared topology, a network dataset, a geometric network, or a terrain.

Sometimes users will organize a collection of feature classes for a common theme into a single feature dataset. For example, users might have a feature dataset for Water that contains Hydro Points (such as dams, bridges, and intakes), Hydro Lines (streams, canals, rivers), and Hydro Polygons (lakes, catchment areas, watersheds, etc.).

In some situations, people might use feature datasets as folders to hold a collection of simple feature classes. This technique is primarily used to organize how users share datasets. However, it is not a useful data structure for editing.

You will need to go through tasks 3 and 4 to decide on a final design for what feature classes should be organized within each feature dataset.




Feature datasets play a key role in establishing permissions for data editing. All the feature classes in a feature dataset will have the same permissions. This means that users can set permissions on feature datasets to identify which organization or group will maintain its contents. If different permissions need to be set on each feature class, then the feature classes should be organized in separate feature datasets (or feature classes), each with its own permission settings. In these cases, extract, transform, load (ETL) or Import/Export procedures can be used to move data updates between each dataset.

When to use feature datasets
Use feature datasets to spatially or thematically integrate related feature classes. Their primary purpose is for building a topology, a network dataset, a terrain dataset, or a geometric network.

You must use feature datasets to hold the set of feature classes that participate in any of the following geodatabase capabilities:

- Topology
- Network dataset
- Terrain
- Geometric network

## Task 3. Add geodatabase elements to facilitate data editing and to manage data integrity.

The geodatabase includes some optional data modeling capabilities that add integrity rules and editing behavior to your GIS. These capabilities help you automate much of your data management work and integrity checks.

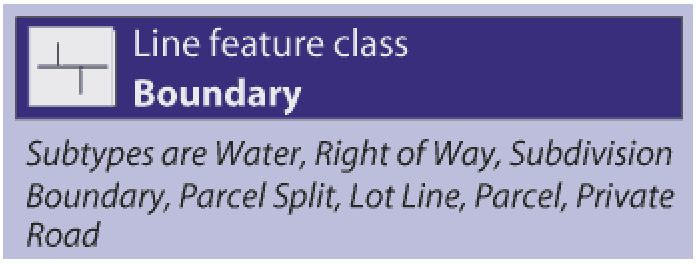
- Do you want to manage the integrity of attribute values? You can use domains, which are rules for assigning valid values in an attribute field.

Coded value domain
**TSDataType**

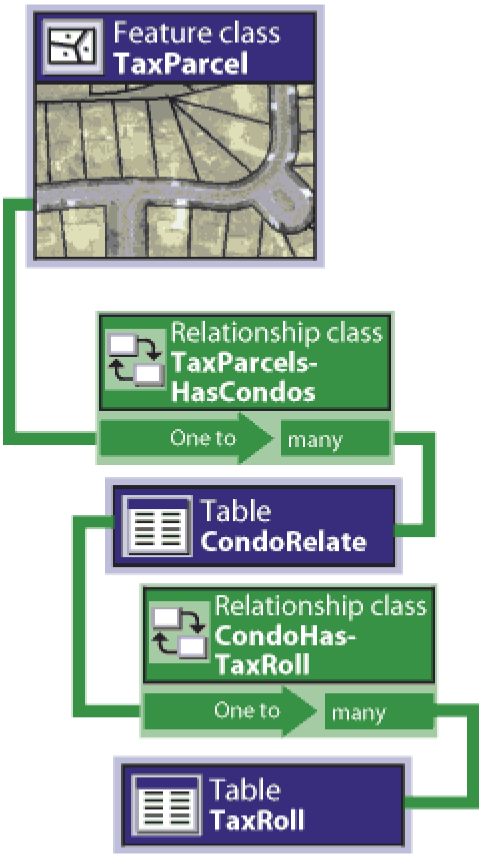
| | |
|---|---|
| Description | |
| Field type | *Long integer* |
| Split policy | *Default value* |
| Merge policy | *Default value* |

| Code | Description |
|---|---|
| 1 | Instantaneous |
| 2 | Cumulative |
| 3 | Incremental |
| 4 | Average |
| 5 | Maximum |
| 6 | Minimum |

- Do you want to use subtypes to help manage subsets of features in a feature class? Subtypes allow you to set up special behaviors for each subclass. They can be used to set default rules for managing feature subsets. For example, you can use subtypes to automatically assign default attribute values as new features are added during editing, to set spatial integrity rules for how new features connect to others, and to add other feature behaviors.

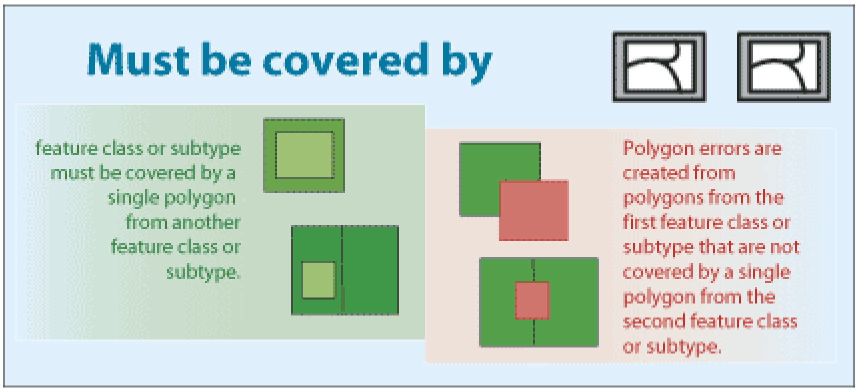



- Determine if there are related tables and if you need relationship classes. Relationship classes allow you to work with features in one table by selecting features in related tables, a very common relational database capability.




- Determine if there are spatial relationships between features in this feature class or with other feature classes that need to be modeled. For example, do you have parcels that share common boundaries? Do they share geometry with a feature class of parcel boundaries and another of parcel corners? Do you want to ensure that your road segments connect to one another or that electrical lines meet at junctions and switches? Do you have county boundaries that nest within states and do not overlap? Do you have vegetation classes that share boundaries with other environmental layers such as slope, aspect, and soil type polygons? In these kinds of cases, topology is very useful; in fact, it is essential.

The feature classes that participate in any topology must be organized into the same feature dataset. See topologies to read more about how you can use them within feature datasets to organize and manage the integrity of topological relationships during editing and update operations.



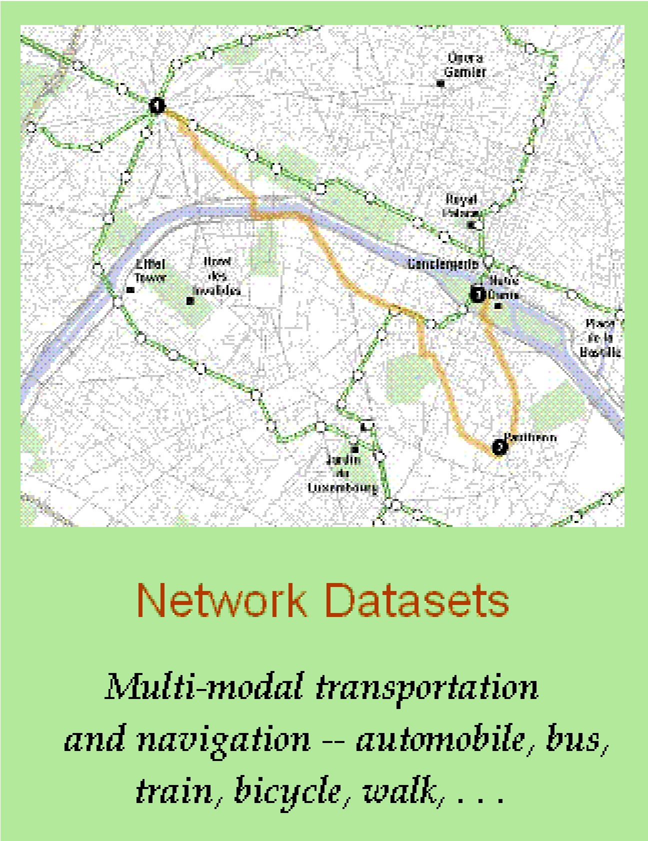

### Task 4. Add capabilities for advanced data uses, analytical models (such as network analysis and geocoding), and advanced cartography.

With each dataset, you may want to consider adding additional geodatabase capabilities that help you to further leverage each dataset. A number of alternative options are available, and you can apply any of these to add capabilities to your geodatabase.

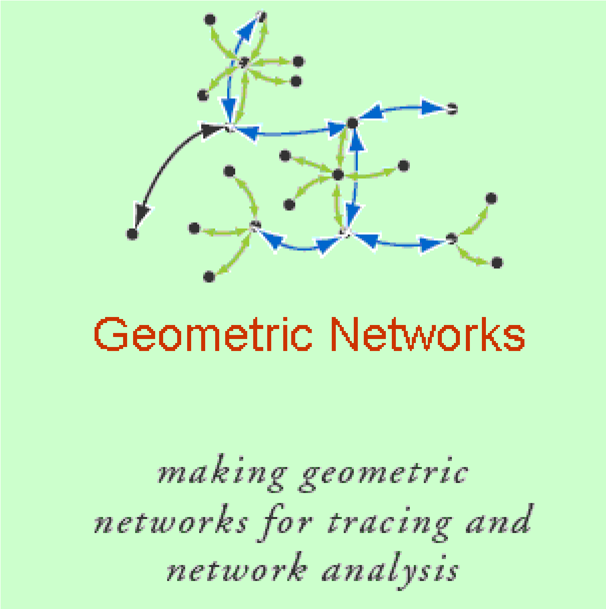
- Do you want to model and use topological relationships to navigate through the nodes, edges, and faces of a **topology**? Will shared feature geometry help you more realistically model your features? For example, the polygon and line boundaries of numerous terrain data layers, such as feature classes for vegetation, slope, aspect, soil types, geology, water bodies, watersheds, ecological zones, and other environmental layers, nest within one another. Integrating their common boundaries using topology enables you to build much more robust and consistent attribute combinations. These greatly impact suitability/capability models and the ability to gain real insight into a problem. Topologies can also help you integrate features such as parcel systems, census units, administrative boundaries, and many other information sets. GIS users sometimes refer to this as vertical integration of GIS data layers.

- Do you want to model a transportation network? The geodatabase uses a network dataset to model these situations. A network dataset is a collection of edges, turns, and junctions through which you can model navigation and the flow of goods and resources. Each network has a set of navigation properties. These include the "cost" to travel along each edge and transfer onto another edge as well as the ability to model one-way, left-turn, and other travel restrictions and multimodal networks (combining trips using an automobile, a bus, walking).

  A network dataset uses feature classes as data sources for edges, junctions, and turns. You specify the role each feature class will play in the network along with its navigation properties. The feature classes that participate in a network must be organized into the same feature dataset.




- Do you want to model utility networks? Electrical utilities and water, storm water, and sewer systems are modeled using a geometric network in the geodatabase. A geometric network is a set of connected edge and junction features used to model the flow of things such as electricity, water, gas, and storm water runoff. Each feature class is assigned a role in the geometric network as a collection of edges or junctions. The connectivity of the network is defined by feature properties and geometric coincidence. For example, valves (which are held as a point feature class) are connected to the endpoints of|segments (stored as line features). If the valve is open, water can flow through it in a specified direction.

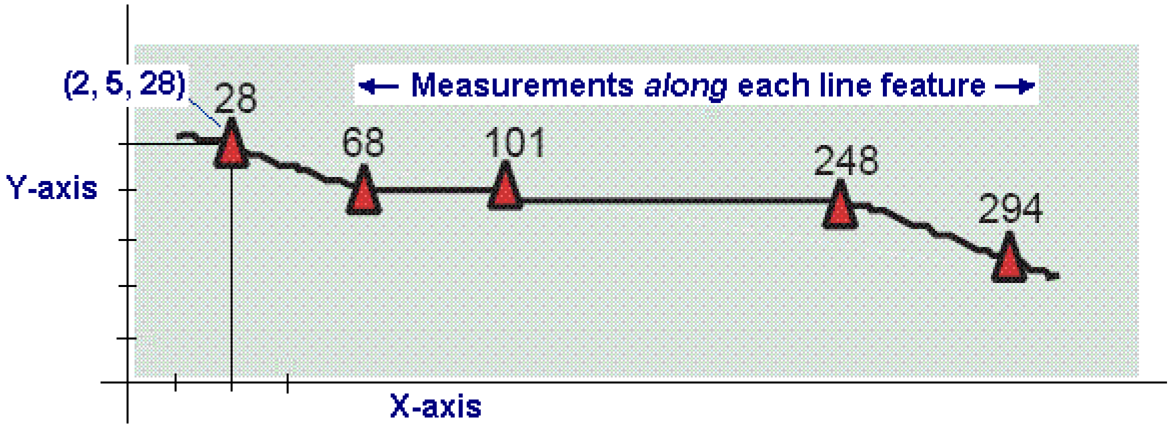



- Do you want to use geocoding? For address geocoding, you add an address locator to your geodatabase. A locator is a combination of one of more feature classes containing addressable features (such as address range information for street centerlines) and a set of address styles and matching rules. Each locator dataset is used as the source for matching a single address or a large file of addresses to find address locations.


You can create locators and save copies of them independent of the geodatabase.

This allows you to share your locators with many kinds of users for use in their own geocoding work.

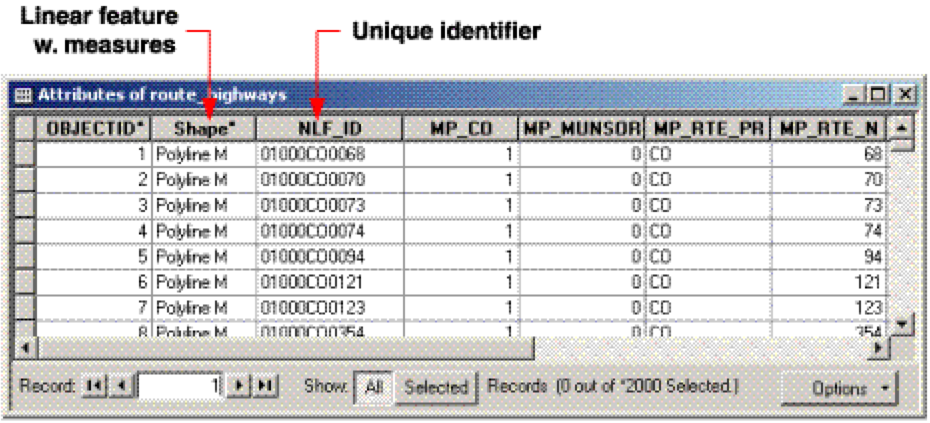
- Do you want to use linear referencing to locate events or facilities along transportation lines? Linear feature vertices can also include m-values. Some GIS applications employ a linear measurement system used to interpolate distances along linear features, such as along roads, stream lines, and pipelines. You can assign an m-value to each vertex in a feature. One very common example is a highway milepost measurement system used by departments of transportation for recording pavement conditions, speed limits, accident locations, and other incidents along highways. Two commonly used units of measure include milepost distance from a set location, such as a county line, and distance from a reference marker.



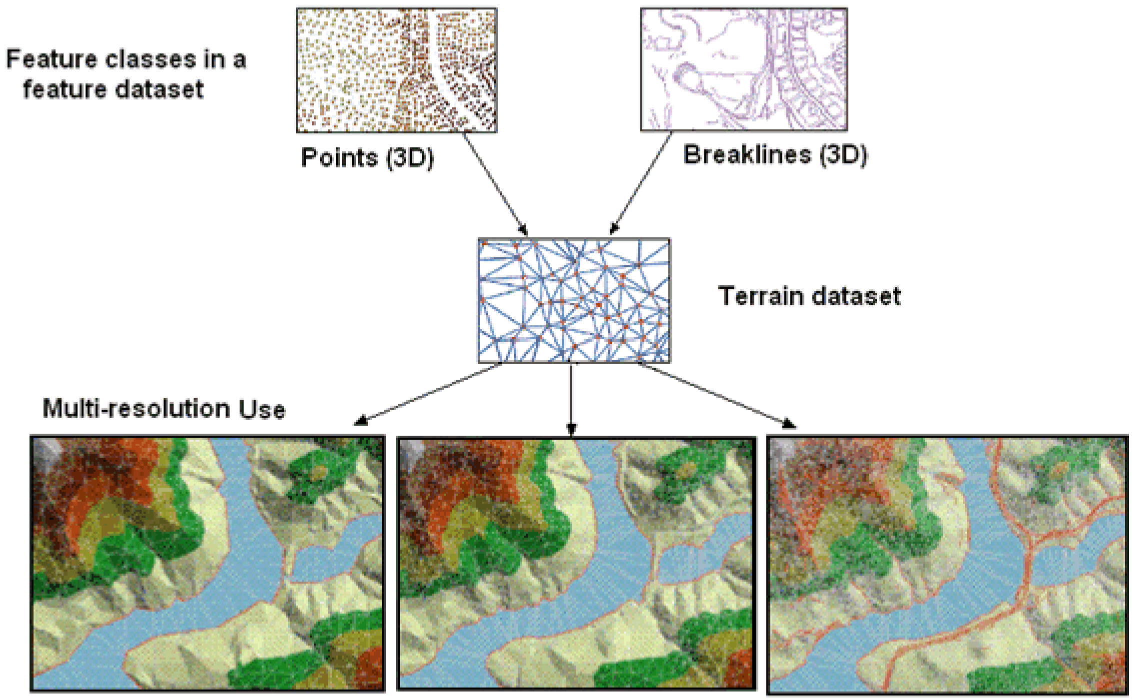

Vertices for measurements can be either (x,y,m) or (x,y,z,m).

Support for these data types is often referred to as linear referencing. The process of geolocating events that occur along these measurement systems is referred to as dynamic segmentation.

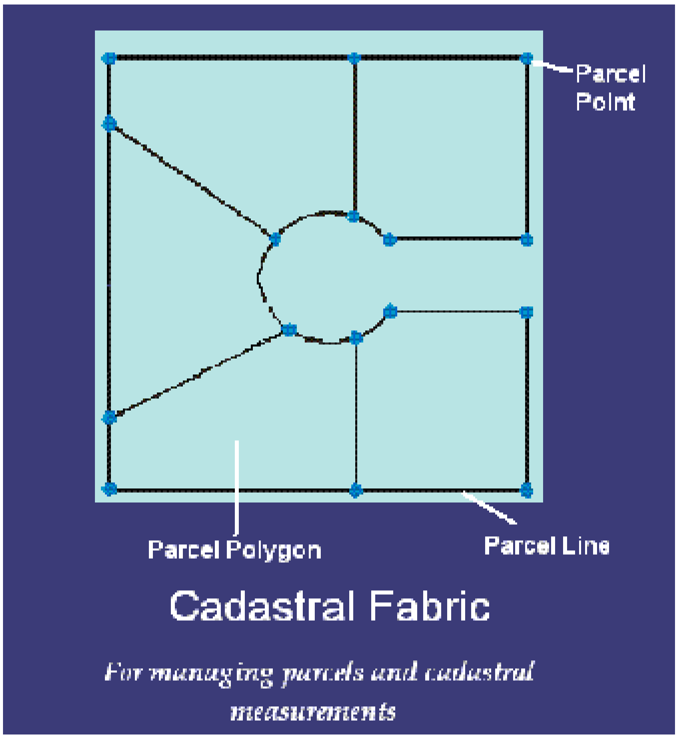
Measured coordinates form the building blocks for these systems. In the linear referencing implementation in ArcGIS, the term *route* refers to any linear feature, such as a city street, highway, river, or pipe, that has a unique identifier and a common measurement system along each linear feature. A collections of routes with a common measurement system can be built on a line feature class as follows:




- Do you want to model elevation using triangulated irregular networks? Or do you need to manage lidar or bathymetric point collections? The geodatabase has a terrain dataset to model surfaces using triangulated networks and to manage large multipoint collections such as lidar and bathymetry data. Terrains are used to manage massive 3D point collections (such as billion-point lidar collections) as well as other 3D features and to generate multiresolution TINs from these collections.

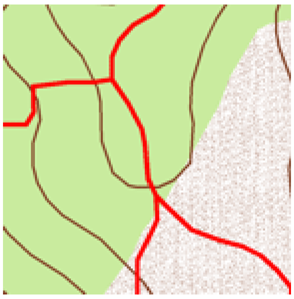


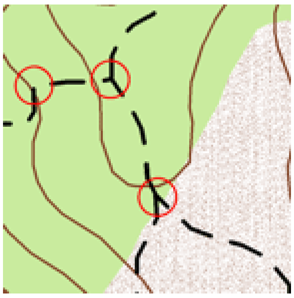

- Do you want to manage parcels or a cadastral database? A cadastral fabric is a dataset of connected parcels. In a cadastral fabric, parcels are represented by parcel line features, parcel point features, and parcel polygon features. Cadastral fabrics are created and managed using the ArcGIS Survey Analyst extension.

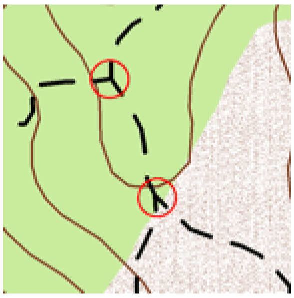


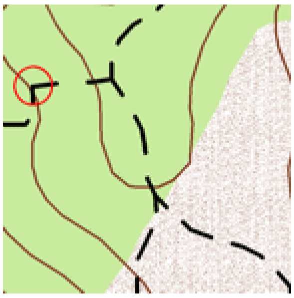

- Do you want to include cartographic representations and rules in your feature classes? A [cartographic representation]() can be added to a feature class to hold drawing rules or alternative graphic representation for the map display of features. In GIS, most users automate mapping by defining a set of map layers. A map layer is a set of rules for how to symbolize and label features on each map. Sometimes layers are not enough to convey the information properly. For example, you might have street centerlines that connect at intersections. But, if you want to show bridges, overpasses, tunnels, and so forth, you cannot easily show these on your map.

A cartographic representation enables users to apply special overrides, rules, and graphics to ensure that the map representation is clear. For example, in a map display, road symbols exaggerate the size of roads and can cause conflicts with other features such as streams and buildings. With cartographic representations, you can offset some feature symbols to remove the conflicts—without having to change the underlying geographic location of the features. You can move the road representations off of the rivers and offset buildings from road symbols.









### *A note about the use of UML for geodatabase design*

- ArcGIS supports the use of CASE tools to import Unified Markup Language (UML) models for geodatabase designs. However, support for all geographic data types, relationships, and behaviors is not complete in UML data modeling.

While UML is a useful tool for documenting the relational aspects of a geodatabase schema (such as table layouts and relationships), generally, it is not recommended to solely use UML for geodatabase design.

UML can be useful for relational database design (for example, for schemas that primarily contain feature classes, attribute tables, and a few other geodatabase properties). However, UML has generally not been useful for designing richer geographic behavior—topologies, networks, terrains, raster catalogs, map layers, map symbols, metadata, cartographic representations, semantic classifications, address locators, cadastral fabrics, linear referencing, and geoprocessing models. These data elements are used to define geographic behavior and associations.

Much of the richness of the geodatabase cannot be universally expressed in a UML design. More important, no special GIS insight is achieved through UML design. Graphing a hierarchy of object-oriented classes, subclasses, and inheritance in UML does not provide insight on how to model the spatial relationships in your geographic data; for example

- How the parcel boundary lines connect to form closed parcel polygons

- How parcel corners, boundaries, and areas share coincident geometry with one another
- What integrity rules you expect to be maintained as part of their geographic representation in the system.

Often, UML distracts designers from defining use cases that help you more clearly articulate critical geographic behaviors and spatial relationships.

Certainly, user communities can find some ways to express their geographic data elements as UML. In other words, you can *document* many (but not all) design aspects of your geodatabase using UML.

Additionally, many relational modelers depend heavily on UML and want their GIS designs to interoperate with their other DBMS designs. In these cases, you can share parts of your geodatabase schemas using UML.

In addition, many people primarily want to use UML as a means to share their schema and rules. ArcGIS has other mechanisms that can support schema documentation and sharing, such as via geodatabase XML.

Bottom line: UML is one of a number of methodologies (such as entity-relationship modeling) that can be used effectively for relational and tabular modeling. However, the use of UML alone is not sufficient. UML is not a replacement for the necessary work of geographic data modeling required in GIS—defining spatial behaviors and use cases of the spatial relationships you want your geodatabase to convey. The design steps described earlier in this Design section of the help (See Geodatabase design steps) will provide guidance on these other aspects of geodatabase design.

A useful tool for documenting your schema using graphic representations that ESRI uses is described in Using the geodatabase diagrammer tool.

### *Design Tips*

Geodatabase data models are designed to be used in practical application scenarios by a wide range of users. To ensure that each design is easy to understand and implement, each data model was built to support easy migration from existing data structures and has been designed to be flexible, extensible, and easily adapted by your organization. Here are a few final design tips to help you with your design implementations.

- **Build on your existing GIS designs.**

  Most existing database designs are suitable for moving forward. You can build on what has worked in the past and find new geodatabase capabilities that will improve on your past efforts.

- **Use generic geodatabase types whenever feasible.**

  Combining generic data structures with very rich GIS tools provides the best solutions that will scale and support multiple users and applications. Leverage the ArcGIS software logic as much as possible for your work. Only use customized GIS data structures as a last resort.

- **Integrate related feature classes using topology.**

  Legacy ArcInfo users with coverages will find many opportunities to integrate feature classes using topologies in the geodatabase. Learn how to use geodatabase topology and its rules. This will create real savings during editing, minimize the amount of customization work you'll need, and increase user productivity. Even small GIS organizations will see 40 percent increases in efficiency for data maintenance.

- **Combine GIS design concepts from this section with traditional relational database design methods.**

  Both database management system (DBMS) and GIS design methodologies are critical for good GIS design. One is not sufficient without the other. Learn to use and apply both sets of techniques.

- **Prototype and pilot your geodatabase design.**

  Prototyping a design using file, personal, or ArcSDE Personal geodatabases with ArcGIS Desktop is easy, fun, and effective. You’ll be surprised at how much insight you’ll gain through experimentation and how much more effective and efficient your design process will become.

During the final stage of design, you’ll want to test scalability and workflows that represent the work that your organization will perform with your geodatabase. Use this to make final adjustments to your design. Be practical in your final test phase and adjust your design as necessary.

# 4. CAPTURING THE REAL WORLD

## 4.1. DIFFERENT METHODS OF DATA CAPTURE

## 4.2. MAP PROJECTION AND SPATIAL REFERENCE

### *An overview of spatial references*

Geographic data for any particular area is stored in separate layers. For example, roads store in one layer, parcels in another, and buildings in a third. To enable the data in each layer to integrate when displayed and queried, each layer must reference locations on the earth's surface in a common way. Coordinate systems provide this framework. They also provide the framework needed for data in different regions to be referenced in different ways. Each layer in the geodatabase has a coordinate system that defines how its locations are georeferenced.

In the geodatabase, the coordinate system and other related spatial properties are defined as part of the spatial reference for each dataset. A spatial reference is the coordinate system used to store each feature class and raster dataset as well as other coordinate properties such as the coordinate resolution for x,y coordinates and optional z- and m- (Measure) coordinates. If required, you can define a vertical coordinate system for datasets with z-coordinates that represent surface elevation.

### *The properties of a spatial reference*

A spatial reference describes where features are located in the real world. You define a spatial reference when creating a geodatabase feature dataset or stand-alone feature class. The spatial reference includes a coordinate system for x-, y-, and z-values as well as tolerance and resolution values for x-, y-, z-, and m-values.

### *Coordinate System*

X,y coordinates are georeferenced with a geographic or projected coordinate system. A geographic coordinate system (GCS) is defined by a datum, an angular unit of measure (usually degrees), and a prime meridian. A projected coordinate system (PCS) consists of a linear unit of measure (usually meters or feet), a map projection, the specific parameters used by the map projection, and a geographic coordinate system.

A projected or geographic coordinate system can have a vertical coordinate system as an optional property. A vertical coordinate system (VCS) georeferences z-values, most commonly used to denote elevation. A vertical coordinate system includes a geodetic or vertical datum, a linear unit of measure, an axis direction, and a vertical shift.

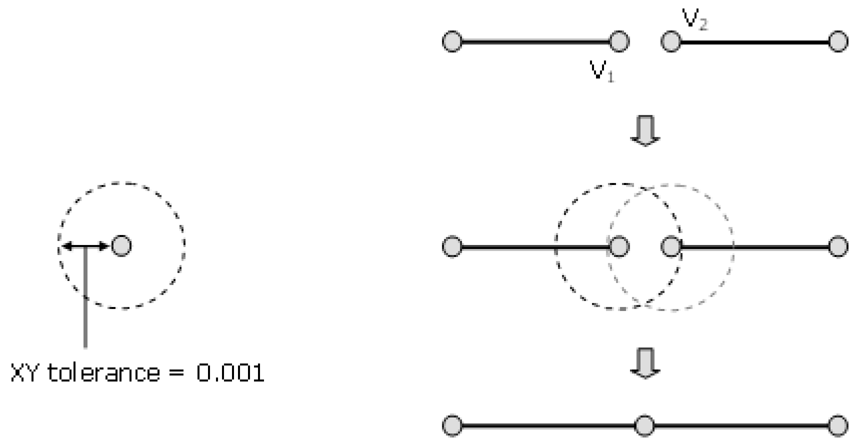
M, or measure, values do not have a coordinate system.

For a spatial reference that includes an unknown coordinate system (UCS) you specify a tolerance only. It is not possible to georeference a feature associated with a UCS. You cannot define a vertical coordinate system if the x,y coordinate system is unknown. If at all possible, you should not use an unknown coordinate system. Because the valid area of use and unit of measure are not known, the resolution and tolerance values may not be appropriate for the data.

#### *Tolerance*

A spatial reference also includes tolerance values. X-, y-, z-, and m-coordinates all have associated tolerance values that reflect the accuracy of the coordinate data. The tolerance value is the minimum distance between coordinates. If one coordinate is within the tolerance value of another, they are interpreted as being at the same location. This value is used in relational and topological operations when determining whether two points are close enough to be given the same coordinate value or if they are far enough apart to each have their own coordinate value.

For example, in the following graphic there are two line features of equal rank in the same feature class. During topology validation, if one vertex V2 is located within the XY tolerance of another vertex V1 (or vice-versa), then both are moved to a new location (e.g., the weighted average distance between the coordinates).

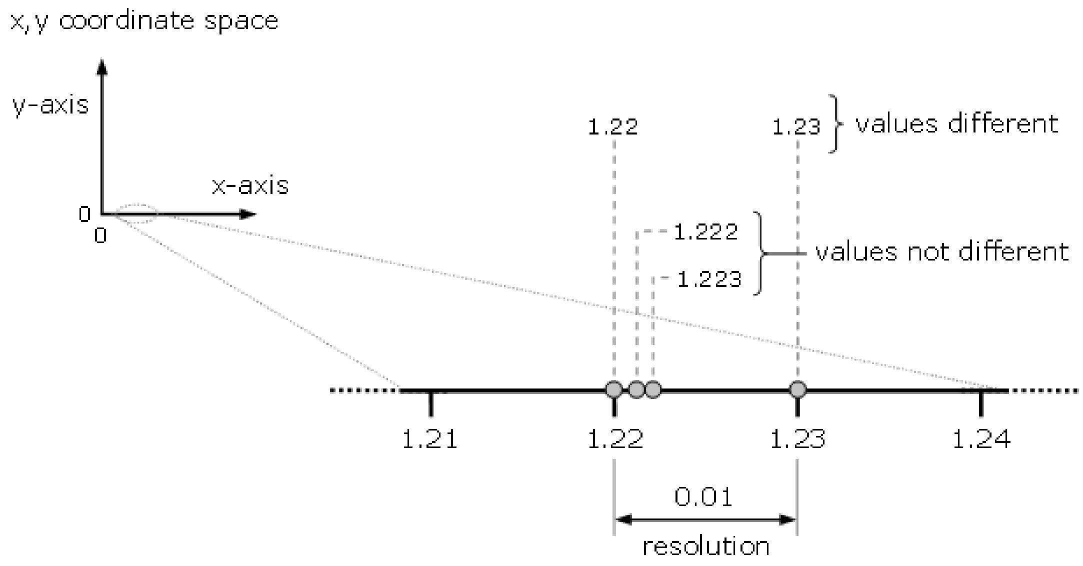



The default tolerance is set to 0.001 meters or its equivalent in map units. This is 10 times the default resolution value and is recommended in most cases. The minimum allowable tolerance value is twice the resolution value. Setting the tolerance value higher will result in lower accuracy in your coordinate data while setting it lower will result in higher accuracy.

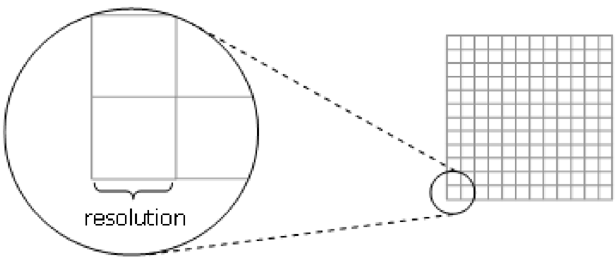
NOTE: Different tolerance values can produce different answers for relational and topological operations. For example, two geometries might be classified as disjoint (no points in common) with the minimum tolerance, but a larger tolerance might cause them to be classified as touching.

### Resolution

The resoultion represents the detail, or precision, in which a feature class depicts the location and shape of geographic features. It is the minimum distance, in map units, that separates unique x values or unique y values. For example, if a spatial reference has an XY resolution of 0.01, then x coordinates 1.22 and 1.23 are different, but x coordinates 1.222 and 1.223 are not, as seen in the graphic below. The latter pair of x coordinates are not considered different because the change in value is less than the XY resolution. The same would apply for y coordinates.

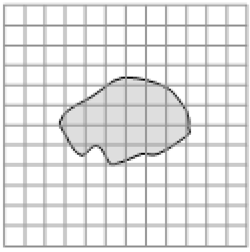



All coordinates of your feature class or feature dataset are georeferenced according to the chosen coordinate system and then snapped to a grid. This grid, known as the coordinate resolution grid, is defined by the resolution which determines the precision (the number of significant digits) of your coordinate values. The resolution establishes the fineness of the coordinate resolution grid that covers the extent of your feature class or feature dataset. All coordinates snap to this grid and the resolution defines how far apart the individual lines of the grid are.

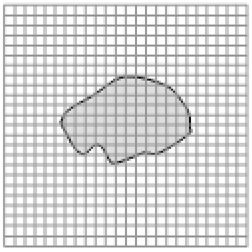



Feature coordinates are more accurately stored with smaller (finer) XY resolutions. However, XY resolutions that are extremely small may have a performance impact in terms of excessive disk use and worsened I/O performance. As XY resolution increases (e.g., becomes more coarse), the accuracy associated with feature coordinates diminishes; feature boundaries will be smoothed, simplified, or not shown at all.

In the following graphic, the grid with the large XY resolution will not be able to store the polygon feature with much precision. Conversely, the grid with the small XY resolution will store the polygon feature more precisely, better preserving its shape.



Large XY resolution
0.1



Small XY resolution
0.0001

Resolution values are in the same units as the associated coordinate system. For example, if a spatial reference is using a projected coordinate system with units of meters, the resolution value is defined in meters. You should use a resolution value that is at least 10 times smaller than the tolerance value. The default resolution value is 0.0001 meters (1/10 mm) or its equivalent in map units. For example, if a feature class is stored in state plane feet, the default precision will be 0.0003281 feet (0.003937 inches). If coordinates are in latitude-longitude, the default resolution is 0.000000001 degrees. For unknown coordinate systems, or for m-values, you will have to set resolution values appropriate to the type of data without explicitly setting the unit of measure.

### *What are Map Projections?*

The shape of the earth is roughly spherical where as maps are two dimensional. Map projection is a set of techniques designed to depict with reasonable accuracy the spherical earth in a two-dimensional (i.e. flat) media. Map projection types are created by an imaginary source of light projected inside the earth. Different map projection families have that source of "light" projected from a different angle within the earth.

A map projection is one of many methods used to represent the 3-dimensional surface of the earth or other round body on a 2-dimensional plane in cartography (mapmaking). This process is typically, but not necessarily, a mathematical procedure (some methods are graphically based).

### *Creation of a Map Projection*

The creation of a map projection involves three steps in which information is lost in each step:

1. selection of a model for the shape of the earth or round body (choosing between a sphere or ellipsoid)
2. transform geographic coordinates (longitude and latitude) to plane coordinates (eastings and northings).
3. reduce the scale (in manual cartography this step came second, in digital cartography it comes last)

### *Metric properties of maps*

Maps assume that the viewer has an orthogonal view of the map (they are looking straight down on every point). This is also called a perpendicular view or normal view. The metric properties or a map are

- area
- shape
- direction
- distance
- scale

### *Choosing a projection surface*

If a surface can be transformed onto another surface without stretching, tearing, or shrinking, then the surface is said to be an applicable surface. The sphere or ellipsoid are not applicable with a plane surface so any projection that attempts to project them on a flat sheet will have to distort the image (similar to the impossibility of making a flat sheet from an orange peel). A surface that can be unfolded or unrolled into a flat plane or sheet without stretching, tearing or shrinking is called a 'developable surface'. The cylinder, cone and of course the plane are all developable surfaces since they can be unfolded into a flat sheet without distorting the projected image (although the original projection of the earth's surface on the cylinder or cone would be distorted).

### *Orientation of the projection*

Once a choice is made between using a cylinder or cone is made, the orientation for that shape must be chosen (how the cylinder or cone is "placed" on the earth). The orientation of the projection surface can be normal (inline with the earth's axis), transverse (at right angles to the earth's axis) or oblique (any angle in between). These surfaces may also be either tangent or secant to the sphere or ellipsoid (if you see both a 1st and 2nd parallel on a projected map then the projection must be secant).

### *Using globes vs. projecting on a plane*

The globe is the only way to represent the earth without distorting one or more of the above-mentioned metric properties. Globes have the advantage of being true to metric properties and able to provide a true picture of spatial relationships on the earth's surface. The disadvantages of the globe are that it is impractical to make large-scale maps with it, it is difficult to measure on a globe, one can't see the whole world at once and it is difficult to handle and transport a globe around (unlike a folding map).

The flat map has the disadvantage of always distorting one or more of the metric properties and it is more difficult to get a true picture of the spatial relationships between objects. Flat maps have numerous advantages however; it is not practical to make large or even medium scale globes, it is easier to measure on a flat map, easy to carry around, and one can see the whole world at once.

Scale in particular is effected by the choice between using a globe vs. a plane. Only a globe can have a constant scale throughout the entire map surface and the scale for flat maps will vary from point to point and may also vary in different directions from a single point (as in Azimuthal maps). The scale for a flat map can only be true along one or two lines or points (tangent or secant points/lines). The 'scale factor is therefore used to measure the difference between the idealized scale and the actual scale at a particular point on the map.

### *Choosing a model for the shape of the Earth*

The projection is also affected by how the shape of the earth is approximated. In the following discussion on projection categories, a sphere is assumed, but the Earth is not exactly spherical but is closer in shape to an ellipsoid with a bulge around the equator. Selecting a model for a shape of the earth involves a choice between the advantages and disadvantages between using a sphere vs. an ellipsoid. Spherical models are useful for small-scale maps (features are small) such as world atlases and globes since the error at that scale is not usually noticeable or important enough to jusify using the more complicated ellipsoid. The ellipsoidal model is commonly used to construct topographic maps and for other large and medium scale maps that need to accurately depict the land surface.

A third model of the shape of the earth is called a geoid, which is a complex and more or less accurate representation of the global mean sea level surface that is obtained through a combination of terrestrial and satellite gravity measurements. This model is not used for mapping due to its complexity but is instead used for control purposes in the construction of geographic datums. A geoid is used to construct a datum by adding irregularities to the ellipsoid in order to better match the Earth's actual shape (it takes into account the large scale features in the Earth's gravity field associated with mantle convection patterns, as well as the gravity signatures of very large geomorphic features such as mountain ranges, plateaus and plains). Datums are always based on ellipsoids that best represent the geoid within the region the datum is going to be used for. Each ellipsoid has a distinct major and minor axis and different controls (modifications) are added to the ellipsoid in order to construct the datum, which is specialized and used for specific geographic regions (such as the North American Datum).

### Categories

Projection classification is based on type of projection surface that is used. The projections are described in terms of placing a gigantic planar surface in contact with the earth, followed by an implied scaling operation. These surfaces are classified as cylindrical (exm. Mercator projection), conic (exm. Albers projection), azimuthal or plane (polar region projections).

There are several different types of projections that aim to accomplish different goals while sacrificing data in other areas through distortion.

- Area preserving projection – equal area or equivalent projection
- Shape preserving – conformal, orthomorphic
- Direction preserving – conformal, orthomorphic, azimuthal (only from a the central point)
- Distance preserving – equidistant (shows the true distance between one or two points and every other point)

NOTE: It is impossible to construct a map projection that is both equal area and conformal.

The two major concerns that drive the choice for a projection are the compatibility of different data sets and the amount of tolerable metric distortions. On small areas (large scale) data compatibility issues are more important since metric distortions are minimal at this level. In very large areas (small scale), on the other hand, distortion is a more important factor to consider.

### Azimuthal projections

Azimuthal projections touch the earth to a plane at one tangent point; angles from that tangent point are preserved, and distances from that point are computed by a function independent of the angle. Azimuthal equidistant projection is used by amateur radio operators to know the direction to point their antennas toward a point and see the distance to it. Distance from the tangent point on the map is equal to surface distance on the earth.

Azimuthal equal-area projection: distance from the tangent point on the map is equal to straight-line distance through the earth.

Azimuthal conformal projection is the same as stereographic projection.

Azimuthal orthographic projection maps each point on the earth to the closest point on the plane.

### Conformal projections

Conformal map projections preserve angles. Mercator projection wraps a cylinder around the earth; the distance from the equator on the map is being geographical latitude, on a scale where the earth's radius is 1.

Stereographic projection touches a plane to the earth and projects each point in a straight line from the antipode of the tangent.

### *Equal-area projections*

These projections preserve area.

Gall-Peters projection wraps a cylinder around the earth and maps each point on the earth to the nearest point on the cylinder.

### *Azimuthal equal-area: see above.*

Cordiform projection designates a pole and a meridian; distances from the pole are preserved, as are distances from the meridian (which is straight) along the parallels.

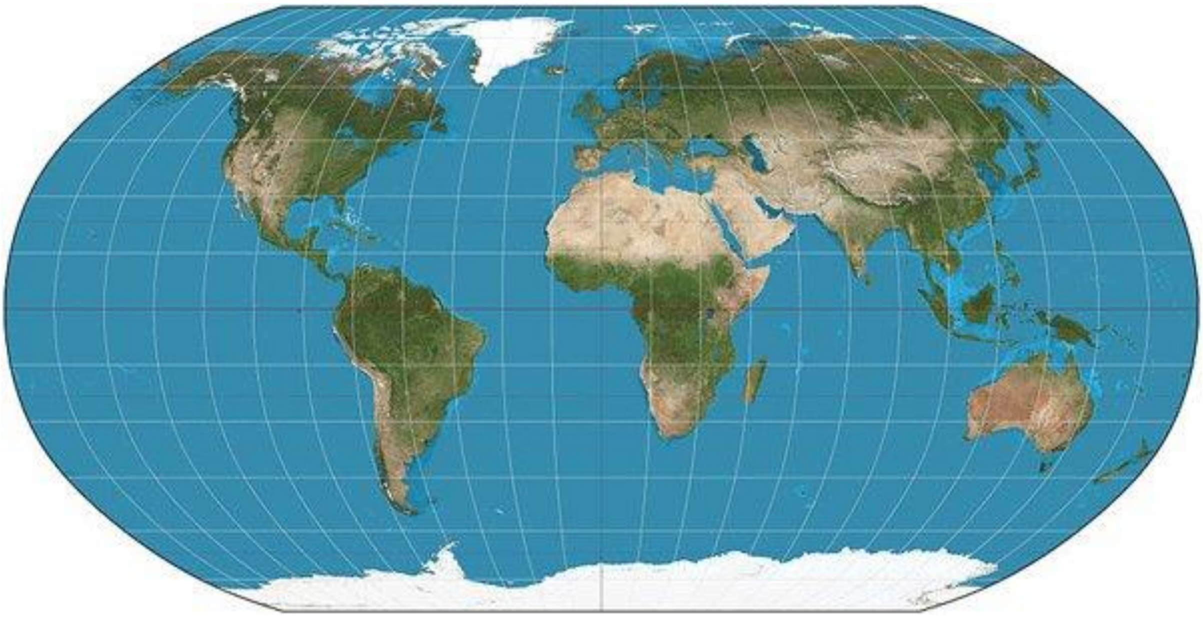
# Common Map Projections

Map projections are one of the fundamental concepts of geography and cartography. Selecting the right map projection is one of the important first considerations for accurate GIS analysis. The problem with projections (and the reason why there are so many types) is that it is very difficult to represent the curved 3D surface of the Earth on a flat 2D surface of a map; some distortion is bound to occur.

Many geographers through the ages have tried to solve the distortion problem through various map projections. A recent example of a pseudocylindrical projection is the Robinson projection which views the entire world at once and one that compromises both area and angles. The longitudinal lines are curved while the latitude lines remain horizontally straight. The Robinson is a compromised view of the Earth's surface with greater amounts of distortion occurring at the poles.

## Robinson Projection

The Robinson Projection was developed by Arthur H. Robinson in 1961 and was indeed to make world maps "look right" rather than measure precisely. This now common projection and has been used in many popular maps such as the Rand McNally series (from the 1960s) and the National Geographic Society (since 1988). In the words of Arthur Robinson: *"…I decided to go about it backwards. I started with a kind of artistic approach. I visualized the best-looking shapes and sizes. I worked with the variables until it got to the point where, if I changed one of them, it didn't get any better. Then I figured out the mathematical formula to produce that effect. Most mapmakers start with the mathematics."*



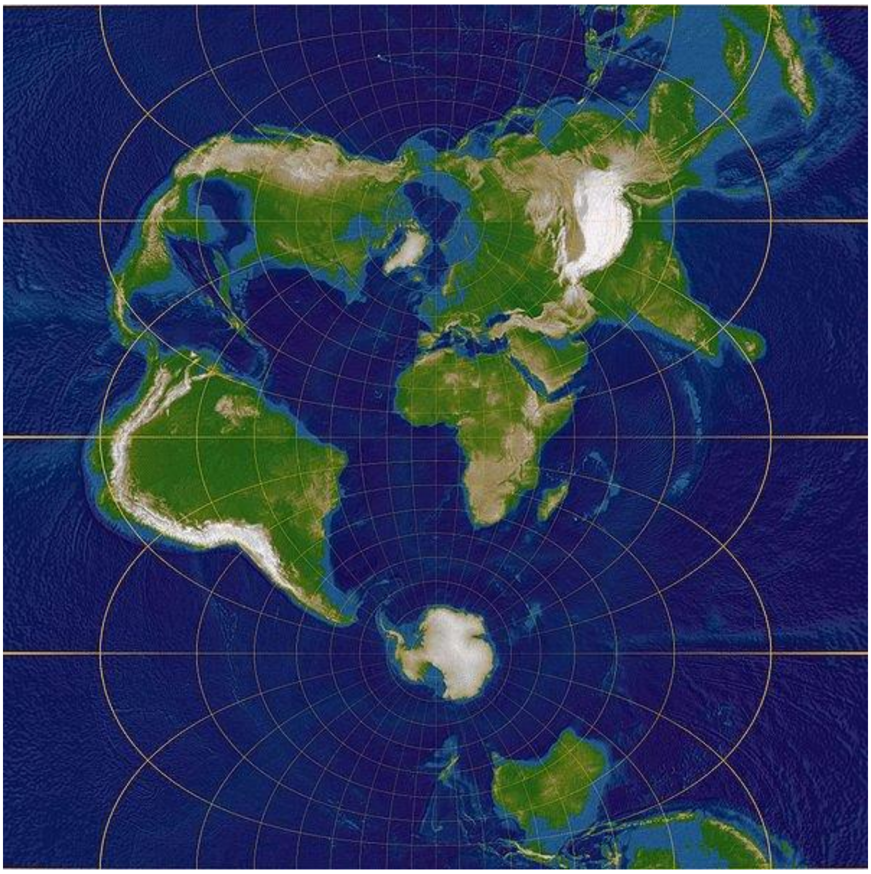
**ROBINSON PROJECTION OF THE WORLD WITH 15° GRATICULES. SOURCE: NASA AND MEDIAWIKI COMMONS.**

A number of map projections have been used throughout history and deciding which projection to use is largely based on what is being mapped. Each projection has its tradeoffs and some are better at depicting the Polar Regions while others are better at depicting mid latitude areas. The scale of the map is also an important consideration as some projections are useful for small areas such as cities and counties while other projects are better for large areas such as continents or world maps.

Further considerations regarding choosing which map projection to use are the complexity of the mathematical functions that transform the coordinates from the curved surface of the earth to a flat plane. With the popularity of GIS software and robust computer hardware, these calculations are now primarily done by computer but without this convenience most mapmakers choose suitable projections with simpler mathematical equations.

# Transverse Mercator Projection

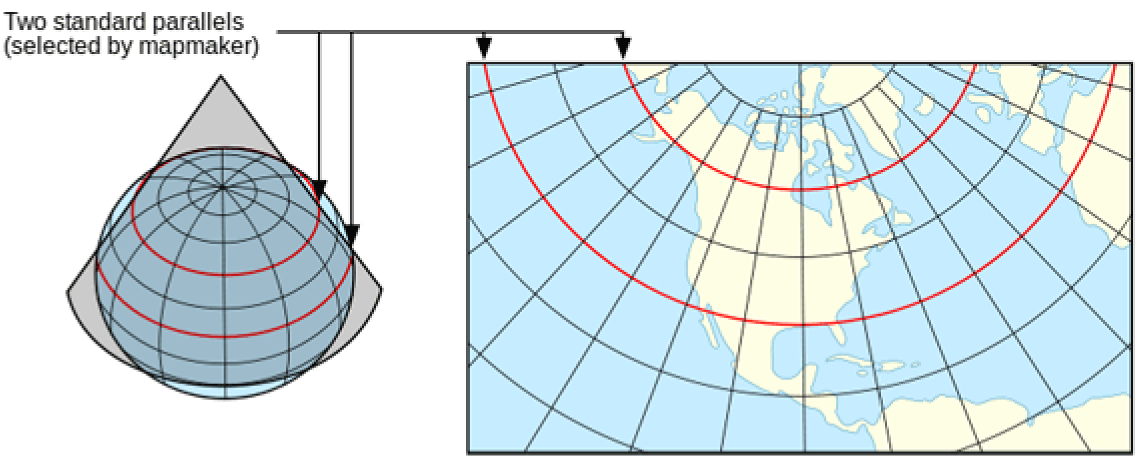
The Transverse Mercator projection is widely used around the world and works especially well for mapping areas smaller than a few degrees longitudinally, such as a state or county. The Transverse Mercator is a revision to the standard Mercator projection in which the cylinder is longitudinally along a meridian instead of the equator. This conformal projection does not maintain true direction (especially evident at large scales) but this distortion can be minimized by placing the central meridian at the region of interest; this is why the Universal Transverse Mercator (UTM) coordinate system uses “zones” that each have their own central meridian. This projection is commonly used on topographic maps, geological maps, and U.S. Geological Survey maps.



**TRANSVERSAL MERCATOR PROJECTION. MAP CENTER IS AT 0°E, 0°N. LEFT BORDER IS NEAR 85°W, RIGHT BORDER IS NEAR 85°E. SOURCE: LARS H. ROHWEDDER.**

# Lambert Conformal Conic

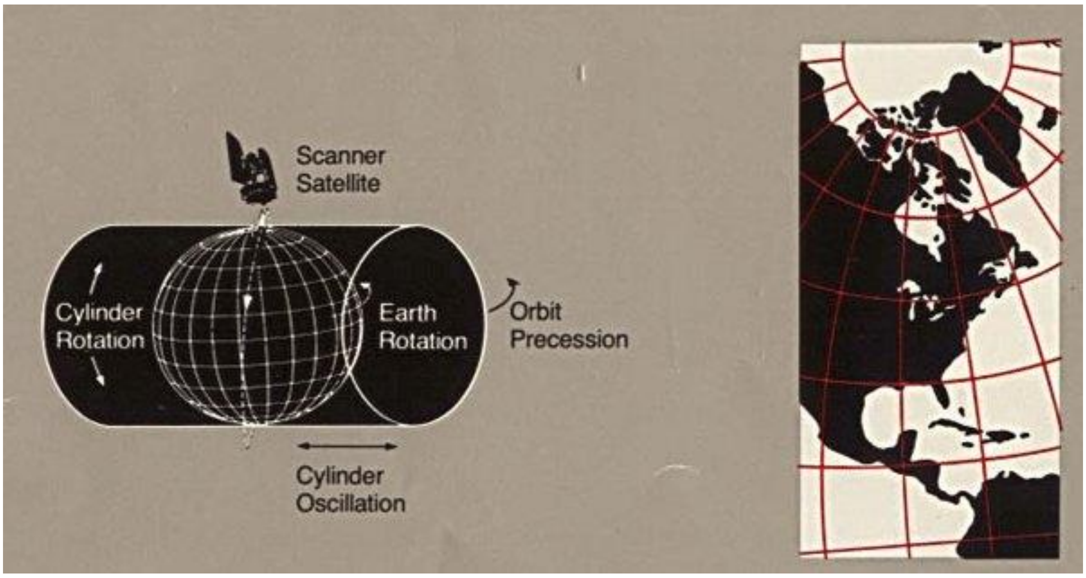
Another common projection currently in use is the Lambert Conformal Conic (LCC). This projection is one of the best to use for middle latitudes and is often used for aeronautical charts, aviators, and maps with wide east-west extents. It is a conformal conical projection with two reference parallels secant lines which help to minimizes distortion; in fact, there is no distortion along the standard parallels but distortion increases further from the chosen parallel.



**LAMBERT CONFORMAL CONIC PROJECTION. SOURCE: USGS.**

# Space Oblique Mercator

A unique and specialized projection is the Space Oblique Mercator (SOM) projection. This map projection was developed fairly recently, in 1976, for the specific purpose of mapping of imagery from an orbiting satellite around the ellipsoidal Earth and is completely free of distortion along the path of the satellite. Originally it was intended for the Landsat satellites but this projection can be used for any satellite in a circular or elliptical orbit around the Earth. This projection has been referred to as, "one of the most complex projections ever devised" by Library of Congress cartographic historian, John W. Hessler.




Many other map projections are currently in use around the globe. The U.S. Geological Survey, charged with mapping the United States, uses more than 18 different map projections with no one particular projection that is used for all applications. Understanding map projections is critical to ensuring accurate and precise mapping.

## 4.3. DATA PREPARATION, CONVERSION AND INTEGRATION

## 4.4. QUALITY ASPECTS OF SPATIAL DATA

When we refer to the concept of geographical information quality is necessary to know that this subject is usually covered by the quality elements. These elements quality should be evaluated in function of the foreseen use and the production cost. Without fear of committing mistakes, we can affirm that every day exist less economic resources and this fact forces the administrators of public resources to find technical solutions that guarantee the optimization of each cent. The direct consequence of this action is the demand of precise and accurate projects. Happily, the manufacturing companies of mensuration equipments accompanied the technological evolution of recent years and have launched new instruments every year with ever more affordable costs. Three decades ago, the acquisition of good angular measurement instruments incorporating an electronic distance measurement, cost around three to four tens of thousands of dollars. Nowadays it is possible to acquire more compact and better quality instruments for one third of the original value. This fact has contributed greatly to the small survey companies and even city halls of small cities, as they can be equipped with instruments that, since they are used in a convenient way, guarantee quality of the collected information. Another interesting example worth mentioning is the great reduction of the acquisition costs of Global Positioning System (GPS) receivers. Less than a decade ago, these instruments were acquired for amounts greater than hundreds of thousands of dollars and nowadays, after the presence of these instruments in the market, the ir prices have great amplitude of values. The user can be given the luxury of choosing the color, model, manufacturer etc. The whole report of the previous paragraph is pure and simply to call the readers attention to always search fast, precise and safe solutions for his projects. Nowadays there are so many norms of procedures that indicate what precision that the project should have or to offer and we have mensuration instruments in the most different levels of precision. It is up to the user to establish the quality that he wants for the project, and from then on take all the necessary precautions given in norms for data collection.

## 4.5. GPS (GLOBAL POSITIONING SYSTEM)

### Global Positioning Systems (GPS)

- a system of earth-orbiting satellites which can provide precise (100 meter to sub-cm.) location on the earth’s surface (in lat/long coordinates or equiv.)

### Remote Sensing (RS)

- use of satellites or aircraft to capture information about the earth’s surface
- Digital ortho images a key product (map accurate digital photos)

### Geographic Information Systems (GIS)

- Software systems with capability for input, storage, manipulation/analysis and output/display of geographic (spatial) information

### WHY GPS AND RS IN GIS?

- GPS and remote sensing imagery are primary GIS data sources, and are very important GIS data sources.
- GPS data creates points (positions), polylines, or polygons
- Remote sensing imagery and airphotos are used as major basis map in GIS
- Information digitized or classified from imagery are GIS layers

GPS and RS are sources of input data for a GIS. A GIS provides for storing and manipulating GPS and RS data.

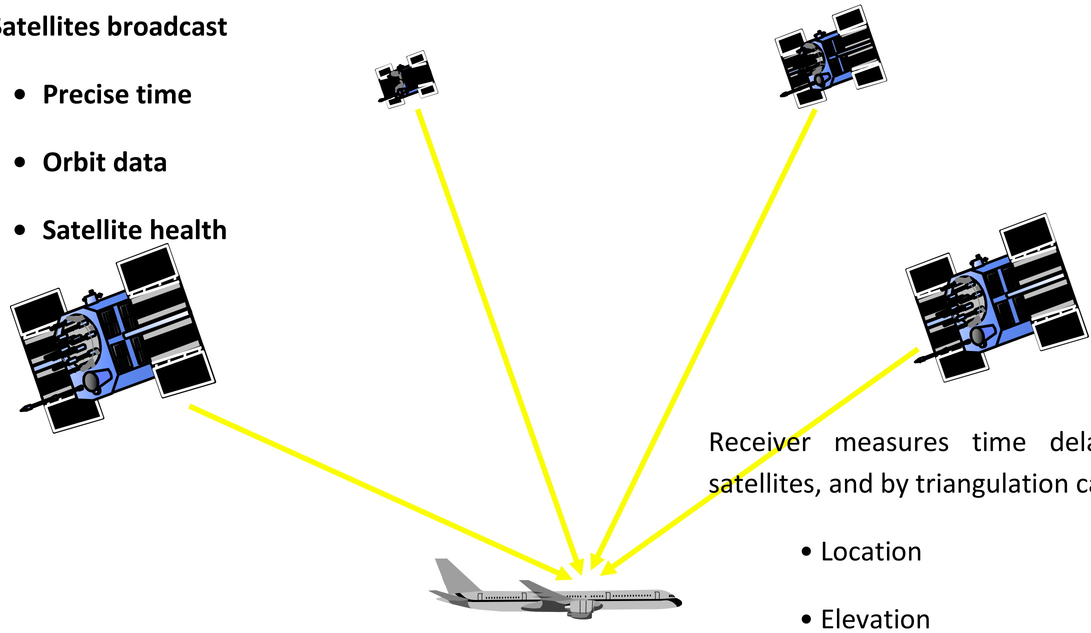
### GPS SYSTEM

- 24 satellites operated by USAF provide 24-hour, all-weather, global coverage
- Satellites are equipped with atomic clocks
- Precise time signals are broadcast on L-band radio frequencies
- Four satellite signals enable receivers to triangulate position

## HOW GPS WORKS?

**Satellites broadcast**

- **Precise time**
- **Orbit data**
- **Satellite health**



Receiver measures time delay from satellites, and by triangulation calculates

- Location
- Elevation
- Velocity

## GPS- HOW IT WORKS

- The GPS receiver compares the time a signal was transmitted by a satellite with the time it was received. The time difference tells the GPS receiver how far away the satellite is.
- With four or more satellites in view, the receiver can determine the user's 3D position (latitude, longitude and altitude).

## GPS IS A DUAL-USE SYSTEM

- Cold War spinoff
  - Developed in 1970s-1980s to support Allied forces
  - Prominent in Gulf War
  - Later civilians gained access to Coarse Acquisition (C/A) signal
- Commercial use now dwarfs military use
- United States Department of Defense (DOD) retains operational control of GPS (for national security?)

## CIVILIAN VS. MILITARY GPS

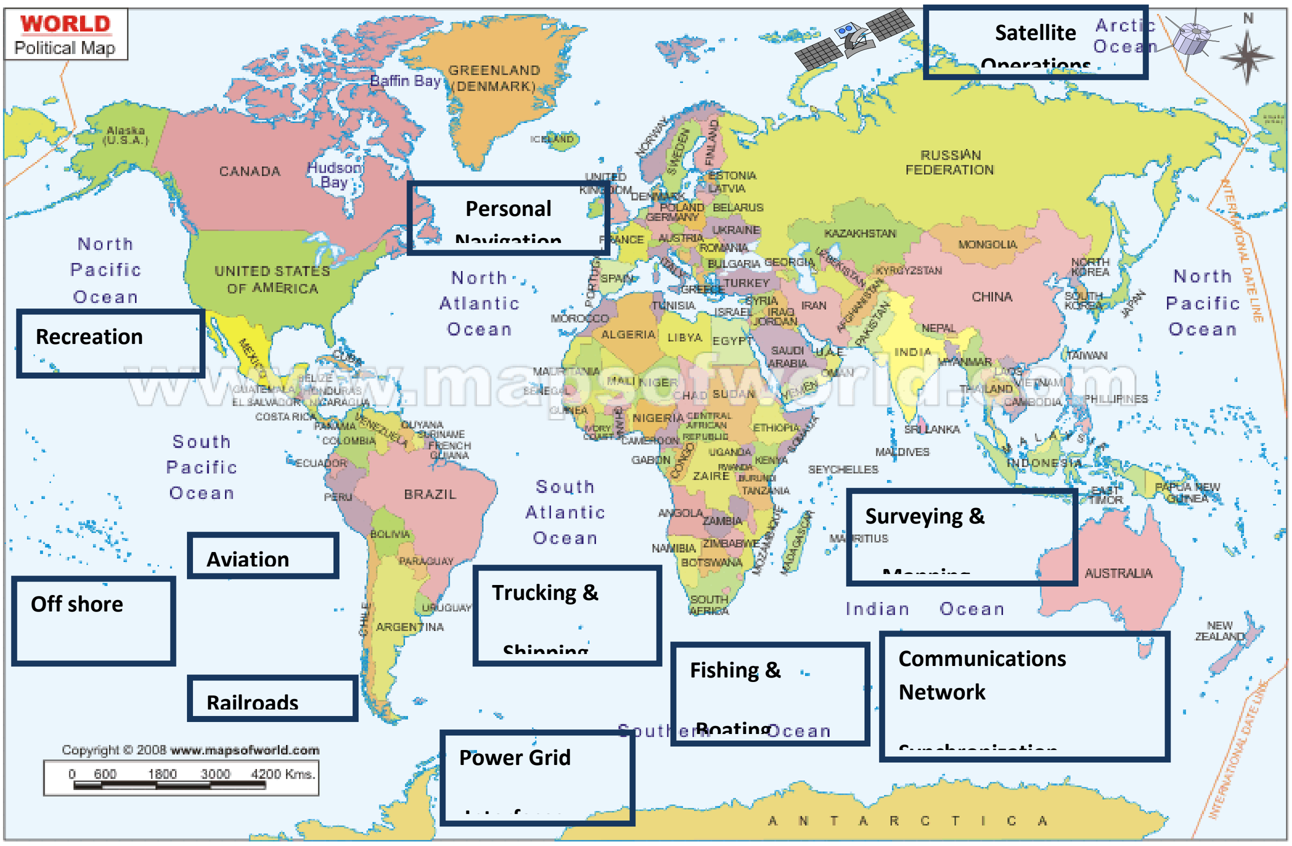
- Civilians: Standard Positioning Service (SPS)
  - Coarse Acquisition (C/A) signal at L1 frequency
  - Selective Availability (SA) limits accuracy to 100 m
  - Augmentation systems can improve accuracy to 1-3 m
- Military: Precise Positioning Service (PPS)
  - SPS plus two encrypted (P(Y)) signals at L1 and L2
  - 20 m accuracy

## GPS IS A GLOBAL INFORMATION UTILITY

- Time and spatial data are critical elements of the global information infrastructure
- Use of GPS increases productivity and is changing the way we live and work
- GPS data is embedded into information systems and is often transparent to end users
- Offered freely to the world as a public good

## IMPORTANT FOR GIS ?

### Major GPS Markets




### Precision Agriculture

- Precise plowing, seeding, watering, spraying
- Localized identification and treatment of distressed crops reduces chemical use
- Precise leveling of fields prevents fluid runoff
- Machinery, asset, and personnel management
- Automated tractor control

### Open Pit Mining

- Enhanced management of assets, equipment
- Work progress tracked in real-time, remotely

- ❖ Improved machine control saves time, lowers maintenance and fuel consumption, prevents accidents
- ❖ Rapid surveying for drilling blast holes
- ❖ Smaller, more empowered workforce

### Timing Applications

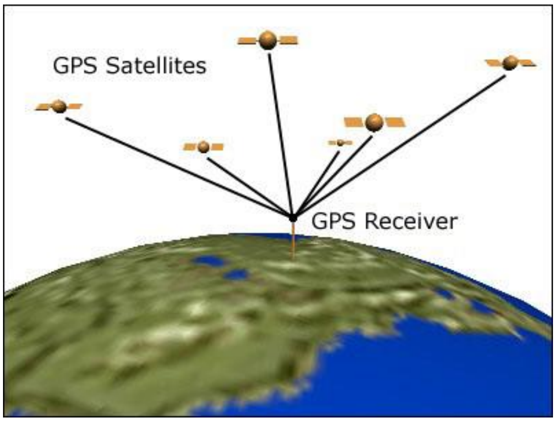
- ❖ Communications network synchronization and management
    - Phone, wireless systems
    - LANs, WANs, Internet
- ❖ Power grid management and fault location
- ❖ Financial transactions
- ❖ E-commerce signatures

### Other Civilian Applications

- ❖ Public Safety
- ❖ Scientific Research
- ❖ Environmental Management
- ❖ Telematics

### GPS IN DETAIL

- ❖ GPS is a Satellite Navigation System
- ❖ GPS is funded and controlled by the U. S. Department of Defense (DOD). While there are many thousands of civil users of GPS world-wide, the system was designed for and is operated by the U. S. military.
- ❖ GPS provides specially coded satellite signals that can be processed in a GPS receiver, enabling the receiver to compute position, velocity and time.
- ❖ At least 4 satellites are used to estimate 4 quantities: position in 3-D (X, Y, Z) and GPSing time (T)




### SATELLITE FACTS

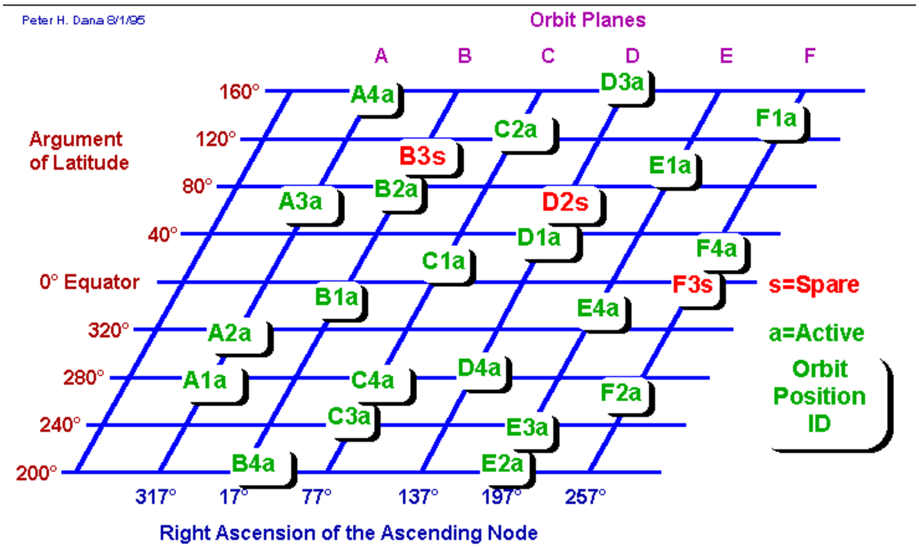
- ❖ Revolve around the Earth in two rotations per day
- ❖ Located 12,000 miles above Earth
- ❖ 7,000 miles per hour
- ❖ 24 of the 27 are working (3 for backup)
- ❖ 3,000 – 4,000 lbs. each
- ❖ Solar powered

### SPACE SEGMENT

The nominal GPS Operational Constellation consists of 24 satellites that orbit the earth in 12 hours. There are often more than 24 operational satellites as new ones are launched to replace older satellites. The satellite orbits repeat almost the same ground track (as the earth turns beneath them) once each day. The orbit altitude is such that the

GPS Nominal Constellation
24 Satellites in 6 Orbital Planes
4 Satellites in each Plane
20,200 km Altitudes, 55 Degree Inclination

satellites repeat the same track and configuration over any point approximately each 24 hours (4 minutes earlier each day). There are six orbital planes, with nominally four SVs (Satellite Vehicles) in each, equally spaced (60 degrees apart), and inclined at about fifty-five degrees with respect to the equatorial plane. This constellation provides the user with between five and eight SVs visible from any point on the earth.



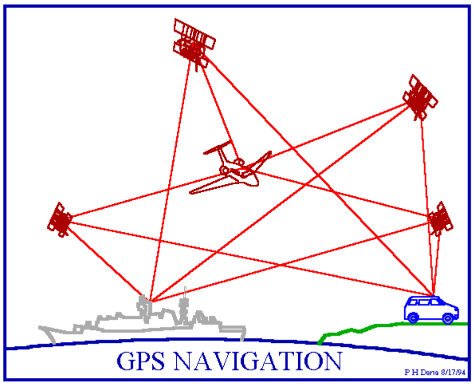

**Simplified Representation of Nominal GPS Constellation**

## CONTROL SEGMENT

The Master Control facility is located at Schriever Air Force Base (formerly Falcon AFB) in Colorado. These monitor stations measure signals from the SVs which are incorporated into orbital models for each satellites. The models compute precise orbital data (ephemeris) and SV clock corrections for each satellite. The Master Control station uploads ephemeris and clock data to the SVs. The SVs then send subsets of the orbital ephemeris data to GPS receivers over radio signals.

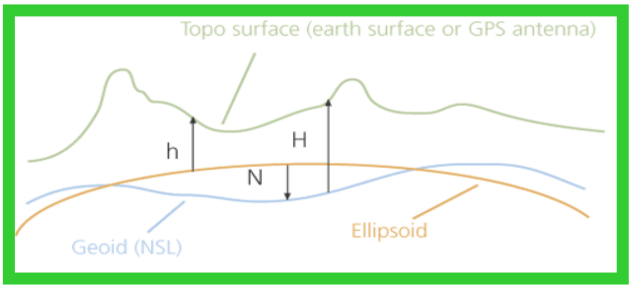

## USER SEGMENT

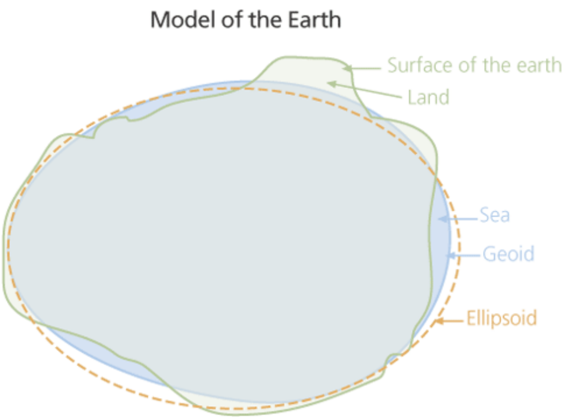
- The GPS User Segment consists of the GPS receivers and the user community. GPS receivers convert SV signals into position, velocity, and time estimates.
- GPS receivers are used for navigation, positioning, time dissemination, and other research.




## COORDINATE SYSTEM AND HEIGHT

- GPS use the WGS 84 as datum
- Various coordinate systems are available for chosen
- GPS height (h) refers to ellipsoid surface, so it is a little difference from the real topographic height (H). the difference is the geoid height (N), the approximate Mean Sea Level. Some newer GPS units now provide the H by using the equation H=h-N (N from a globally defined geoid – Geoid99)




H: topographic height or

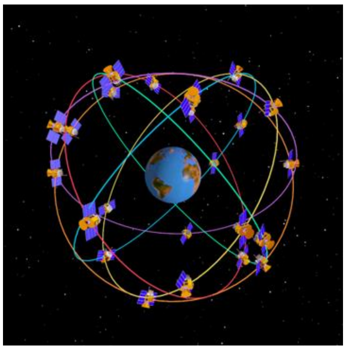
orthometric height

h: ellipsoid height

N: geoid height

## INTERRUPTIONS TO THE SATELLITE

- There are some factors that can affect the satellites performance and job to relay the data to the receivers such as:
  - Ionosphere and Troposphere Delays
  - Signal Multipath
  - Receiver Clock Errors
  - Orbital Errors
  - Number of Satellites Visible
  - Satellite Geometry/Shading
  - Intentional Degradation of the Satellite Signal



- Ionosphere and Troposphere:
  - Signal that slows down in transition through the atmosphere
- Signal Multi-path:
  - Affected by things in the surrounding area (such as rocky mountains or buildings)
- Receiver Clock:
  - Clocks do not match up on the receiver and the satellite
- Number of Visible Satellites:
  - The more the better
- Satellite Geometry/Shading:
  - Need to be spaced properly; line/tight group = bad signals
- Intentional Degradation of Satellite Signal:
  - Specific for the Military use but affects the civilian populations use of GPS

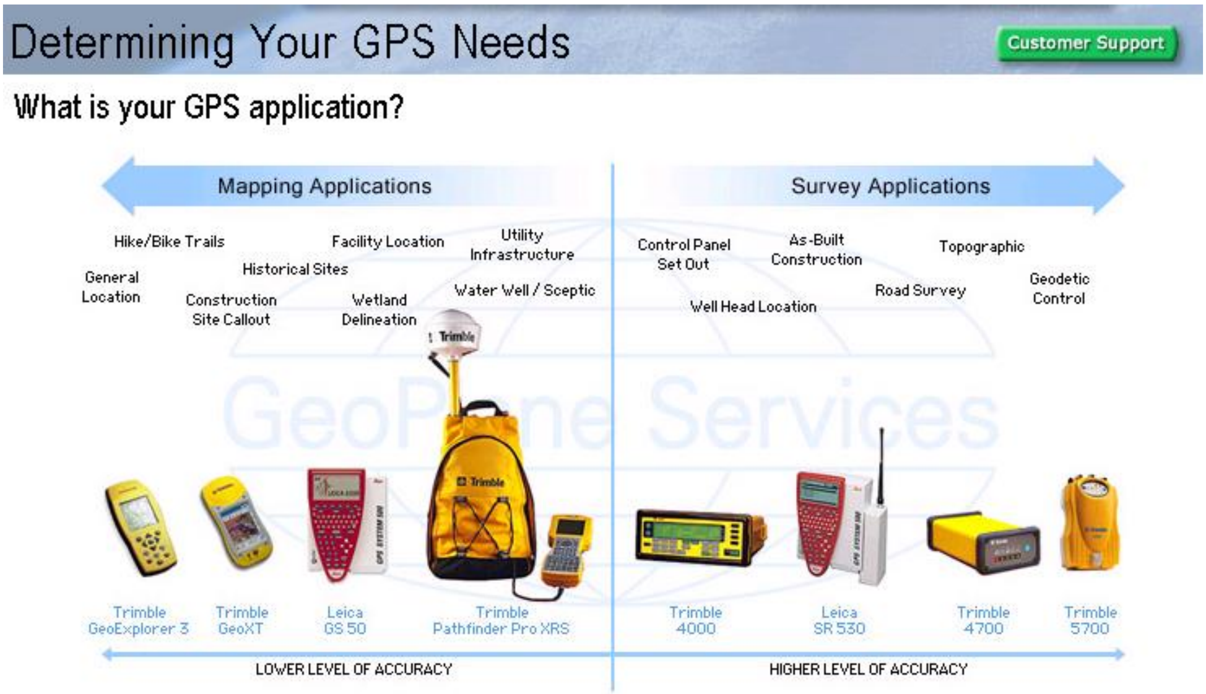
## GPS POSITIONING SERVICES SPECIFIED IN THE FEDERAL RADIO NAVIGATION PLAN

- PPS (precise positioning service) for US and Allied military, US government and civil users. Accuracy:
  - 22 m Horizontal accuracy
  - 27.7 m vertical accuracy
  - 200 nanosecond time (UTC) accuracy
- SPS (standard positioning service) for civil users worldwide without charge or restrictions:
  - 100 m Horizontal accuracy
  - 156 m vertical accuracy
  - 340 nanosecond time (UTC) accuracy
- DGPS (differential GPS techniques) correct bias errors at one location with measured bias errors at a known position. A reference receiver, or base station, computes corrections for each satellite signal.
  - Differential Code GPS (navigation): 1-10 m accuracy
  - Differential Carrier GPS (survey):1 mm to 1 cm accuracy

## DIFFERENTIAL GPS

- **The idea behind differential GPS**: We have one receiver measure the timing errors and then provide correction information to the other receivers that are roving around. That way virtually all errors can be eliminated from the system.
- If two receivers are fairly close to each other, say within a few hundred kilometers, the signals that reach both of them will have traveled through virtually the same slice of atmosphere, and so will have virtually the same errors
- **Real Time Transmission DGPS or Post-processing DGPS**
- Reference stations established by The United States Coast Guard and other international agencies often transmit error correction information on the radio beacons that are already in place for radio direction finding (usually in the 300kHz range). Anyone in the area can receive these corrections and radically improve the accuracy of their GPS measurements. Many new GPS receivers are being designed to accept corrections, and some are even equipped with built-in radio receivers.
- If precise positioning immediately (real time) is not needed, recorded data can be merged with corrections recorded at a reference receiver (through internet) for a later clean-up.

## GPS NEEDS



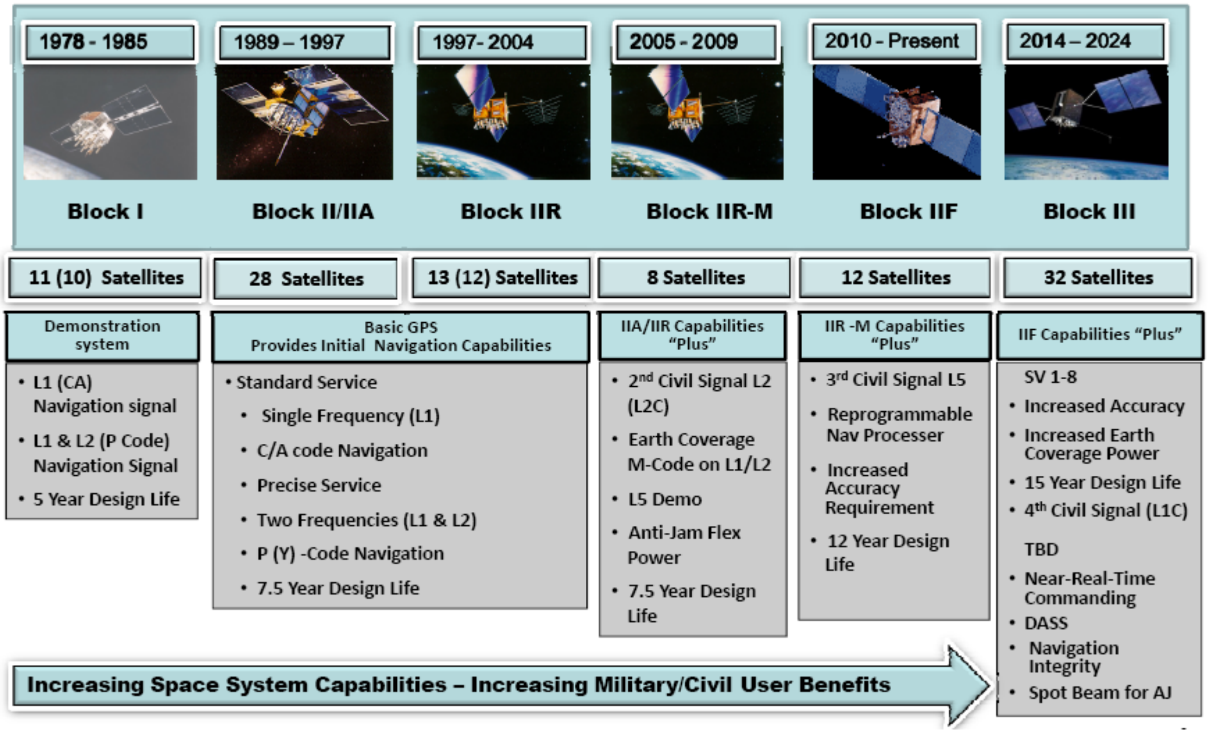

Project tasks can often be categorized by required accuracies which will determine equipment cost.

## GPS ACCURACIES, COSTS, AND SIGNALS

| GPS APPROACH | ACCURACY ESTIMATE | RECEIVER COST ESTIMATE | GPS SIGNALS | | | | |
|---|---|---|---|---|---|---|---|
| | | | L1 C/A CODE | L1 P-CODE | L1 CARRIER | L2 P-CODE | L2 Y-CODE |
| SPS NAVIGATION | 100 M | $1,000 | X | | | | |
| SPS DIFFERENTIAL >30KM | 10 M | $5,000 | X | | | | |
| SPS DIFFERENTIAL <30KM | 1 M | $5,000 | X | | | | |
| PPS NAVIGATION | 10 M | $10,000 | X | X | | X | |
| ANTI-SPOOFING NAVIGATION | 10 M | $20,000? | X | X | X | X | X |
| L1 CARRIER PHASE SURVEY | 0.1 M | $10,000 | X | | X | | |
| L1 L2 CARRIER PHASE SURVEY | 0.01 M | $15,000 | X | X | X | X | |

Peter H. Dana 8/28/94

### GPS PROGRAM EVOLUTION



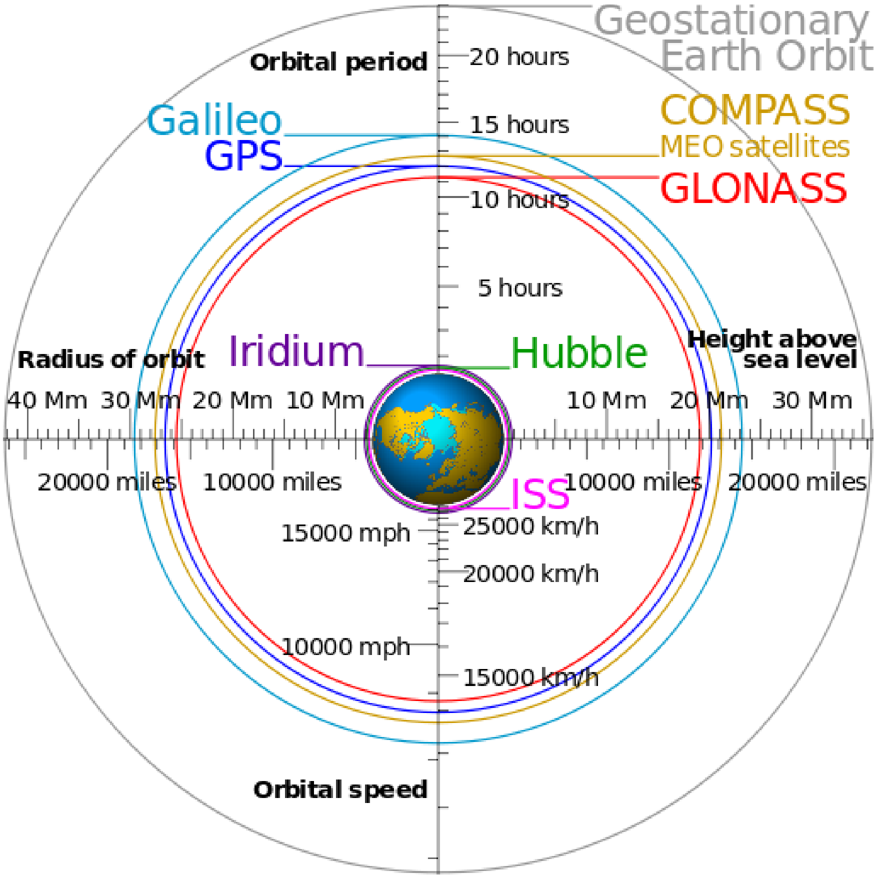

## OTHER GLOBAL NAVIGATION SATELLITE SYSTEMS (GNSS)

- GLONASS
  - Russian Federation
  - (24) Satellites
- Galileo
  - European Union
  - (27+3) Satellites
- Compass
  - China
  - (27 MEO+3IGSO+5GEO) Satellites
- Regional Constellation
  - Indian Regional Navigational Satellite System (IRNSS) (7)
  - Quasi-Zenith Satellite System (QZSS) (Japan) (4)

### 4.5.1. Remote Sensing

- Remote Sensing is the science and art of acquiring information (spectral, spatial, temporal) about material objects, area, or phenomenon, without coming into physical contact with the objects, or area, or phenomenon under investigation.
- In remote sensing, information transfer is accomplished by use of electromagnetic radiation (EMR).
- EMR is a form of energy that reveals its presence by the observable effects it produces when it strikes the matter.
- So, one is looking at the physical nature of spatially distributed features.

### SATELLITE NAVIGATION ORBITS COMPARISON



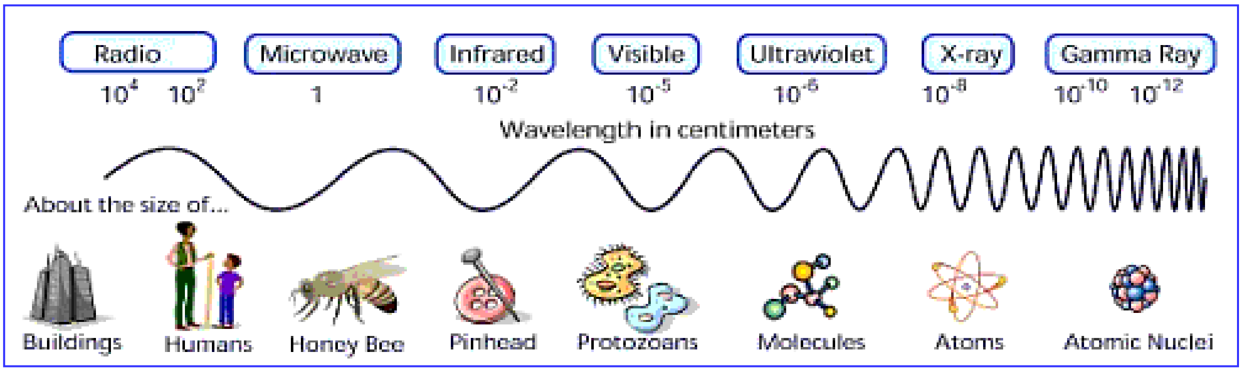

### ISSUES IN GNSS

- **Compatibility**
  - Ability of all PNT services to be used separately or together without interfering with each individual service or signal
  - Radio frequency compatibility
- **Interoperability**
  - ability of all PNT services to be used together to provide the user better capabilities than would be achieved by relying solely on one service or Signal
  - Promote fair competition in the global marketplace

### 4.5.2. Remote Sensing Basics

- Using electromagnetic spectrum to image the land, ocean, and atmosphere.

- EMR is considered to span the spectrum of wavelengths from 10-10 µm to cosmic rays up to 1010 µm



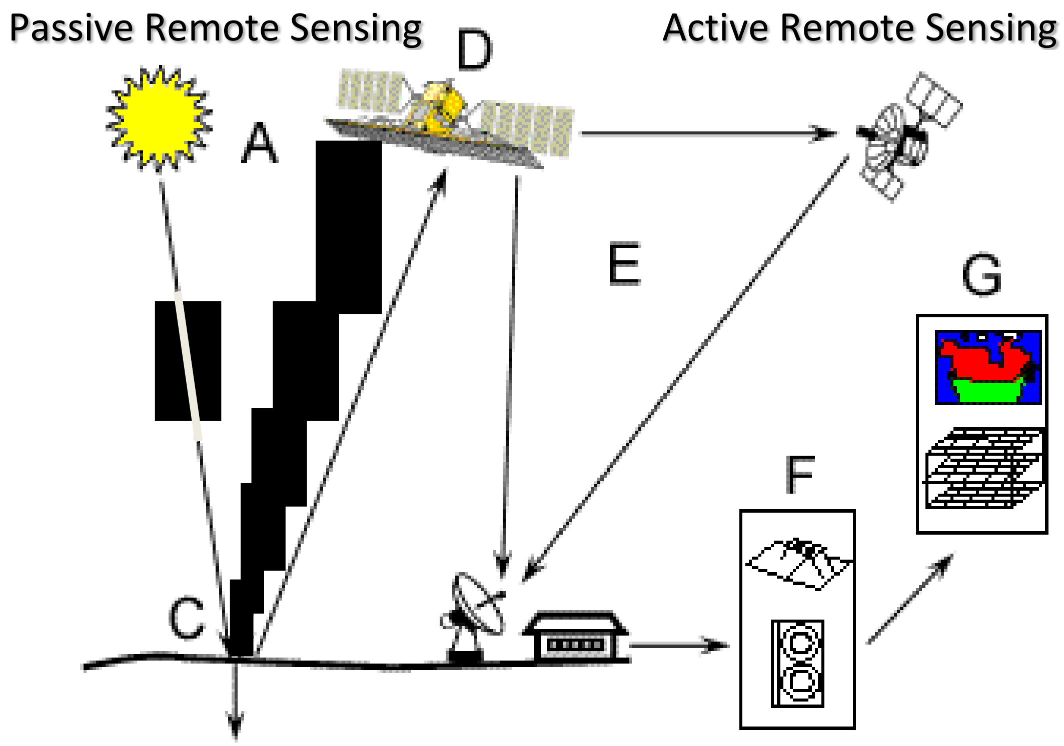

When you listen to the radio, or cook use microwave oven, you are using EM waves.

When you take a photo, you are actually doing remote sensing.

### 4.5.3. Types of Remote Sensing

- **In respect to the type of Energy Resources:**
  - **Passive Remote Sensing:** Makes use of sensors that detect the reflected or emitted electro-magnetic radiation from natural sources.
  - **Active remote Sensing:** Makes use of sensors that detect reflected responses from objects that are irradiated from artificially-generated energy sources, such as radar.

- **In respect to Wavelength Regions:**
  Remote Sensing is classified into three types in respect to the wavelength regions
  - Visible and Reflective Infrared Remote Sensing.
  - Thermal Infrared Remote Sensing.
  - Microwave Remote Sensing.




A. the Sun: energy source

C. target

D. sensor: receiving and/or energy source

E. transmission, reception, and pre-processing

F. processing, interpretation and analysis

G. analysis and application

### 4.5.4. Bands Used in RS

- Emission of EMR (Electo-Magnetic Radiation) from gases is due to atoms and molecules in the gas.
- Emission from solids and liquids occurs when they are heated and results in a continuous spectrum. This is called thermal emission and it is an important source of EMR from the viewpoint of remote sensing.
- The Electro-Magnetic Radiation (EMR), which is reflected or emitted from an object, is the usual source of Remote Sensing data. However, any medium, such as gravity or magnetic fields, can be used in remote sensing.

### 4.5.5. RS Sensors

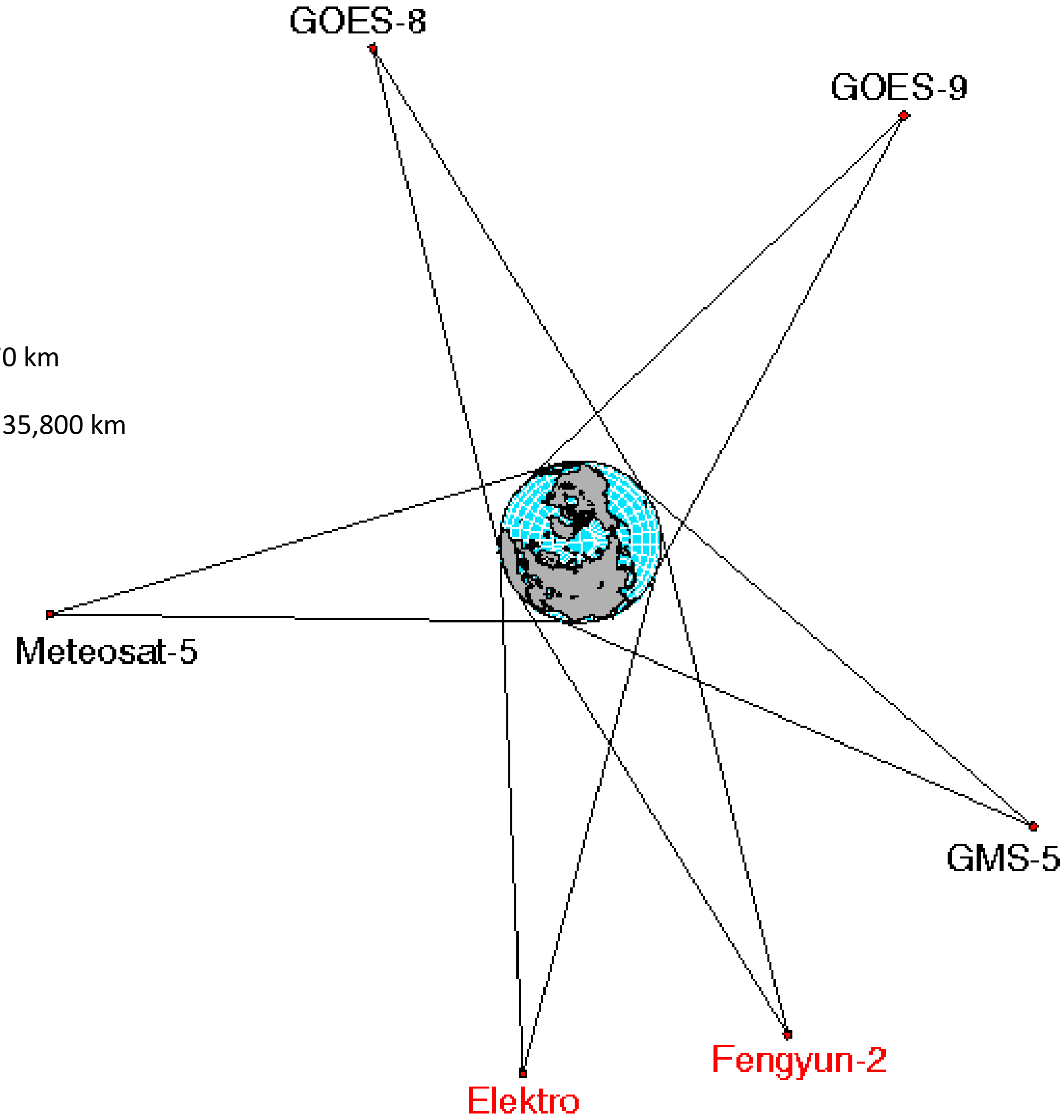
- One cannot select the sensors to be used in any given remote-sensing task arbitrarily; one must instead consider

- The available spectral sensitivity of the sensors,
- The source, magnitude, and spectral composition of the energy available in these ranges.
- Ultimately, however, the choice of spectral range of the sensor must be based on the manner in which the energy interacts with the features under investigation.

### 4.5.6. Global Geostationary Satellites

Earth radius 6,370 km

Satellite altitude 35,800 km




### 4.5.7. Energy Interactions and Colour Readability

- All matter is composed of atoms and molecules with particular compositions. Therefore, matter will emit or absorb electro-magnetic radiation on a particular wavelength with respect to the inner state. All matter reflects, absorbs, penetrates and emits Electro-magnetic radiation in a unique way.
- Electro-magnetic radiation through the atmosphere to and from matters on the earth's surface are reflected, scattered, diffracted, refracted, absorbed, transmitted and dispersed.

- For example, the reason why a leaf looks green is that the chlorophyll absorbs blue and red spectra and reflects the green. The unique characteristics of matter are called spectral characteristics.

### 4.5.8. Energy Interactions

- The proportions of energy reflected, absorbed, and transmitted will vary for different earth features, depending upon their material type and conditions.
- These differences permit us to distinguish different features on an image.
- Even within a given feature type, the proportion of reflected, absorbed, and transmitted energy will vary at different wavelengths.

### 4.5.9. Remote Sensing of Earth's Resources -Process & Elements

Following are major components of Remote sensing System:

1. Energy Source
   - Passive System: sun, irradiance from earth's materials;
   - Active System: irradiance from artificially generated energy sources such as radar.
2. Platforms:(Vehicle to carry the sensor) (truck, aircraft, space shuttle, satellite, etc.)
3. Sensors:(Device to detect electro-magnetic radiation) (camera, scanner, etc.)
4. Detectors: (Handling signal data) (photographic, digital, etc.)
5. Processing:(Handling Signal data) (photographic, digital etc.)
6. Institutionalisation: (Organisation for execution at all stages of remote-sensing technology: international and national orrganisations, centres, universities, etc.).

### 4.5.10. RS Platform

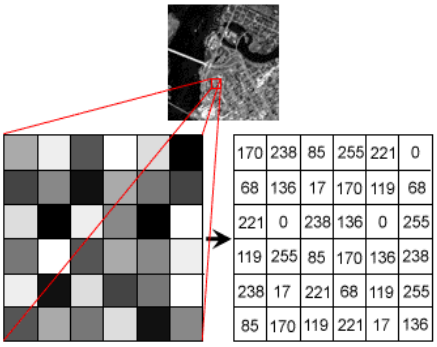
- The vehicles or carriers for remote sensors are called the platforms.
- Typical platforms are satellites and aircraft, but they can also include radio-controlled aeroplanes, balloons kits for low altitude remote sensing, as well as ladder trucks for ground investigations.
- The key factor for the selection of a platform is the altitude that determines the ground resolution and which is also dependent on the instantaneous field of view (IFOV) of the sensor on board the platform.

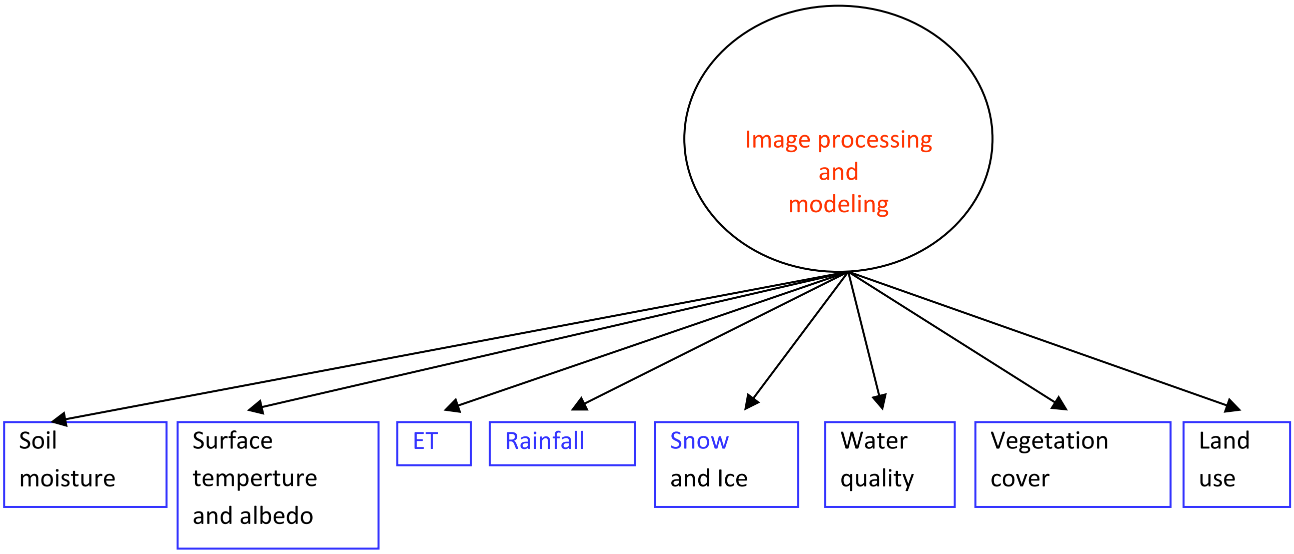
### 4.5.11. Resolution

- Resolution is defined as the ability of an entire remote-sensing system, including lens antennae, display, exposure, processing, and other factors, to render a sharply defined image. Resolution of a remote-sensing is of different types.
  1. Spectral Resolution: of a remote sensing instrument (sensor) is determined by the band-widths of the Electro-magnetic radiation of the channels used. High spectral resolution, thus, is achieved by narrow bandwidths width, collectively, are likely to provide a more accurate spectral signature for discrete objects than broad band width.
  2. Radiometric Resolution: is determined by the number of discrete levels into which signals may be divided.
  3. Spatial Resolution: in terms of the geometric properties of the imaging system, is usually described as the instantaneous field of view (IFOV). The IFOV is defined as the maximum angle of view in which a sensor can effectively detect electro-magnetic energy.
  4. Temporal Resolution: is related to the repetitive coverage of the ground by the remote-sensing system. The temporal resolution of Landsat 4/5 is sixteen days.

### 4.5.12. Image processing and modeling

The size of a cell we call image resolution, depending on… Such as 1 m, 30 m, 1 km, or 4 km




### 4.5.13. An Ideal Remote Sensing System-I

- Ideal RS system have following characters:
  - Uniform energy source
    - This source will provide energy all over wavelengths, at a constant, known, high level of output, irrespective of time and place.
  - Non-interfering atmosphere
    - This will be an atmosphere that will not modify the energy from the source in any manner, whether that energy is on its way to earth's surface or coming from it. Again, ideally this will hold irrespective of wavelength, time, place, and sensing altitude involved.

### 4.5.14. An Ideal Remote Sensing System-II

- Distinct and unique spectral response patters for every feature
  - These interactions will generate reflected and or/or emitted signals that are not only selective in respect to wavelengths, but also are known, invariant, and unique to each and every earth surface feature type and subtype of interest.
- A super-sensor
  - This will be a sensor, highly sensitive to all wavelengths, yielding spatially detailed data on the absolute brightness (or radiance) from a scene (a function of wavelength), throughout the spectrum. This super sensor will be simple and reliable, require, virtually no power or space , and be accurate and economical to operate.
- Real time data acquisition
  - In this system, the instant the radiance versus wavelength response over a terrain element is generated, it will be processed into an interpretable format and recognized as being unique to the particular terrain element from which it comes. This processing will be performed nearly instantaneously (real time), providing timely information.
- Multiple users
  - These people will have comprehensive knowledge of both their respective disciplines and of remote-sensing data acquisition and analysis techniques. The same set of data will become various forms of information for different users, because of their vast knowledge about the particular earth resources being used.

### 4.5.15. Real Life RS

- Unfortunately, an ideal remote-sensing system does not exist.
- Real remote- sensing systems fall short of the ideal at virtually every point in the sequence outlined.

### 4.5.16. “3P”

- GPS data creates points (positions), polylines, or polygons as GIS layers
- GPS is always together with remote sensors while acquiring imagery
- GPS points used as ground control points for remote sensing image registration
- Remote sensing imagery and air photos are used as major basis map in GIS
- Information digitized or classified from imagery are GIS layers
- GIS layers help field data collection including GPS measurements
- GIS layers help image classification and field truth

# 5. SPATIAL ANALYSIS AND VISUALIZATION

## 5.1. SPATIAL ANALYSIS

## 5.2. MAP OUTPUTS AND ITS BASIC ELEMENTS

# 6. 6. INTRODUCTION TO SPATIAL DATA INFRASTRUCTURE

## 6.1. SDI CONCEPTS AND ITS CURRENT TREND

Making decisions based on geography is basic to human thinking. Where shall we go, what will it be like, and what shall we do when we get there are applied to the simple event of going to the store or to the major event of launching a bathysphere into the ocean's depths. By understanding geography and people's relationship to location, we can make informed decisions about the way we live on our planet. A geographic information system (GIS) is a technological tool for comprehending geography and making intelligent decisions. GIS organizes geographic data so that a person reading a map can select data necessary for a specifi c project or task. A thematic map has a table of contents that allows the reader to add layers of information to a basemap of real-world locations. For example, a social analyst might use the basemap of Eugene, Oregon, and select datasets from the U.S. Census Bureau to add data layers to a map that shows residents' education levels, ages, and employment status. With an ability to combine a variety of datasets in an infinite number of ways, GIS is a useful tool for nearly every field of knowledge from archaeology to zoology. A good GIS program is able to process geographic data from a variety of sources and integrate it into a map project. Many countries have an abundance of geographic data for analysis, and governments often make GIS datasets publicly available. Map file databases often come included with GIS packages; others can be obtained from both commercial vendors and government agencies. Some data is gathered in the field by global positioning units that attach a location coordinate (latitude and longitude) to a feature such as a pump station. GIS maps are interactive. On the computer screen, map users can scan a GIS map in any direction, zoom in or out, and change the nature of the information contained in the map. They can choose whether to see the roads, how many roads to see, and how roads should be depicted. Then they can select what other items they wish to view alongside these roads such as storm drains, gas lines, rare plants, or hospitals. Some GIS programs are designed to perform sophisticated calculations for tracking storms or predicting erosion patterns. GIS applications can be embedded into common activities such as verifying an address. From routinely performing work-related tasks to scientifically exploring the complexities of our world, GIS gives people the geographic advantage to become more productive, more aware, and more responsive citizens of planet Earth.

### *What Is SDI?*

Social challenges, environmental issues, and economic downturns all take cooperation to solve. Working together to map and document the earth helps create a structure for managing knowledge. From large countries to small nations, everyone benefits from documented public works and utilities, protected environments and biodiversity, correctly assessed resources, and completed strategic planning. The term *spatial data infrastructure* was coined in 1993 by the U.S. National Research Council to denote a framework of technologies, policies, and institutional arrangements that together facilitate the creation, exchange, and use of geospatial data and related information resources across an information-sharing community. Such a framework can be implemented narrowly to enable the sharing of geospatial information within an organization or more broadly for use at a national, regional, or global level. In all cases, an SDI will provide an institutionally sanctioned, automated means for posting, discovering, evaluating, and exchanging geospatial information by participating information producers and users. SDI extends a GIS by ensuring that geospatial data and standards are used to create authoritative datasets and polices that support it.

### *What is a SDI?*

"The Global Spatial Data Infrastructure supports ready global access to geographic information. This is achieved through the coordinated actions of nations and organisations that promote awareness and implementation of complimentary policies, common standards and effective mechanisms for the development and availability of interoperable digital geographic data and technologies to support decision making at all scales for multiple purposes."

### *Four main components*

- Overriding objective to maximise the use of national geographic information assets
- This requires some form of coordinated action on the part of government
- It must be user driven 'to support decision making at all scales for multiple purposes'
- This involves a wide range of activities including technical and institutional matters and human resource development

## 6.2. THE CONCEPT OF METADATA AND CLEARING HOUSE

### *What are Metadata?*

A metadata record is a file of information, usually presented as an XML document, which captures the basic characteristics of a data or information resource. It represents the *who, what, when, where, why* and how of the resource. Geospatial metadata commonly document geographic digital data such as Geographic Information System (GIS) files, geospatial databases, and earth imagery but can also be used to document geospatial resources including data catalogues, mapping applications, data models and related websites. Metadata records include core library catalogue elements such as Title, Abstract, and Publication Data; geographic elements such as Geographic Extent and Projection Information; and database elements such as Attribute Label Definitions and Attribute Domain Values.

The FGDC is tasked by an Executive Order to develop procedures and assist in the implementation of a distributed discovery mechanism for national digital geospatial data. Geospatial metadata are critical to data discovery and serves as the fuel for the NSDI Clearinghouse.

Most NSDI stakeholders have long utilized the Content Standard for Digital Geospatial Metadata (CSDGM), which will continue to have a legacy for many years. International geospatial metadata standards are emerging in the community. FGDC policy states that non-Federally authored standards that are endorsed by the FGDC have the same status as FGDC developed standards. Since ISO 19115 and the associated standards are endorsed by the FGDC, federal agencies are encouraged to transition to ISO metadata as their agencies are able to do so. While the selection of appropriate standards is dependent on the nature of your metadata collection and publication process, ISO metadata should be considered an option now. It's recognized that the transition to ISO metadata will be occurring over the next few years.

### *Why bother with Metadata?*

Metadata helps people who use geospatial data find the data they need and determine how best to use it. Metadata supports producers in locating and using their own data resources and data consumers in locating and using data resources produced by others. Metadata also supports:

### *Data Management requirements to:*

- Preserve the data history so that it can be re-used or adapted,
- Assess the age and character of data holdings to determine which data should be maintained, updated, or deleted,
- instill data accountability by requiring you to state what you know about the data and realizing what you don’t, but should, know about your data
- Limit data liability by explicitly designating the effective and administrative limits of use of the data.

### *Project Management requirements to:*

- Plan and document the data types and content needed to support the project
- Monitor data development by regular review of the process steps completed and recorded within the metadata
- Provide all project participants a common language of attributes and process methods and a place to record and share their progress
- Access the lineage and content of outsourced data production by requiring robust metadata as a contract deliverable.

As personnel change in an organization, institutional knowledge leaves the organization. Undocumented data can lose their value. Subsequent workers may have little understanding of the contents and uses for a digital data base and may find they can't trust results generated from these data. Also, lack of knowledge about other organizations' data can lead to duplication of effort. It may seem burdensome to add the cost of generating metadata to the cost of data collection, but in the long run metadata are worth it.

### *Metadata:*

Metadata is defined by the New Merriam-Webster Dictionary as “data that provides information about other data”. Geographic metadata is used to document the attributes of geographic data, e.g. database files and data develop within a Geographic Information System (GIS), in the same way that the nutrition label to the right documents the attributes of a food product. Geographic metadata seeks to answer questions such as: Who developed the data? When was the data collected? How were the data processed? How are the data attributes defined? In what formats are the data available? How does one obtain the data? The information in the metadata provides context for the data and supports the effective application of the data.

### *What are the core components of a metadata record?*

Metadata Record Information - information about the metadata record including the language in which the record is written, a unique file identifier for the metadata record, the metadata standard used to organize the record, a point of contact for the metadata record, and the date that the metadata record written.

Identification Information – citation-level information about the data including the title, abstract, purpose for creation, status, keywords (theme and place), and extent (temporal, vertical and horizontal).

Constraints Information – information about legal and security limitations to data access and use. Data Quality Information – information about the processes and sources used to develop the data and positional and/or accuracy assessments performed.

Maintenance Information – information about the scope and frequency of data updates. Spatial Representation – information about the mechanism used to represent spatial data (grid, point, vector).

Reference System Information – information about the reference systems used to represent geographic position and time.

Content Information – information about the data set entities and attributes. Symbology Information – information about the symbols used to represent spatial features.

Distribution Information – information about the data distributors and methods for obtaining the data.

Metadata Extension Information – information about custom, user-based, changes to the elements, domains or conditionality of the standard.

Application Schema Information – information about the schema or data models used to structure the data.

FGDC (Federal Geographic Data Committee) member agencies are actively exploring ISO metadata implementation.

## CLEARINGHOUSE CONCEPTS

This document describes the context of the National Geospatial Data Clearinghouse Network and details of its construction and operation.



### What is Clearinghouse?

The NSDI Clearinghouse Network, sponsored by the FGDC, is a distributed system of agency servers located on the Internet that contain field-level descriptions of available and planned digital spatial data, applications, and services. This descriptive information, known as metadata, is collected in a standard format to facilitate query and consistent presentation across multiple participating sites. Clearinghouse uses standards-based Web technology for the publication and discovery of available geospatial resources through the Geospatial Platform portal.

The fundamental goal of Clearinghouse is to provide access to digital spatial data and related online services for data access, visualization, or order. The Clearinghouse Network functions as a detailed catalog service with support for links to spatial data and browse graphics. Clearinghouse metadata are expected to include hyperlinks to online resources (e.g. map services, data download locations, data access services, applications) within their metadata entries to enable access to all facets of the described resource. Where digital data are too large to be made available through the Internet or the data products are made available for sale, linkage to an order form can be provided in lieu of a data set. Through this model, Clearinghouse metadata provides low-cost advertising for providers of spatial data, both non-commerical and commercial, to potential customers via the Internet.

Clearinghouse allows individual agencies, consortia, or geographically-defined communities to band together and promote their available digital spatial data through a federated metadata service. These servers may be installed at local, regional, or central offices, as dictated by the organizational and logistical efficiencies of each organization. All Clearinghouse servers are considered "peers" within the Clearinghouse activity -- there is no hierarchy among the servers -- permitting query by any user on the Internet with minimum transactional processing. When these Clearinghouse services are registered with the Platform portal, the system will harvest and cache a copy of the metadata for rapid retrieval, enabling search through a single interface to all registered assets in the U.S.

### Why promote the Clearinghouse Activity?

The development of the Clearinghouse was motivated by a desire to minimize duplication of effort in the collection of expensive digital spatial data and foster cooperative digital data collection activities. By promoting the availability, quality, and requirements for digital data through a searchable on-line system a Clearinghouse facility would greatly assist in coordination of data collection and research activities. Clearinghouse also provides a primary data dissemination mechanism to traditional and non-traditional spatial data users.

### Why not just use Internet search engines?

Digital spatial data and metadata are stored in many forms and systems which make their discovery on the Internet difficult. Structured metadata is typically exchanged in XML format with significant meaning stored in 'fields' or XML elements rather than the HTML documents typically indexed in search engines. Use of current web indexing technology offers literal text search and matching for metadata which happen to be stored in HTML, but do not generally provide the indexing required for search of coordinates, dates and times, and other numeric values. In addition, some entire collections of metadata are being managed within dynamic databases whose content is not accessible to search engines. The Clearinghouse functionality as implemented in the geodata.gov portal goes beyond existing search engine technology to include spatial query and permit simple search of metadata based on location and full-text search. Field-level search is also available to refine searches based on topical classification, geography, time, and other key fields in ways not possible with off-the-shelf search engine technology.

The general trend toward connectivity of spatial data producers, vendors, and users on the Internet coupled with the provision of online data via web services indicate a long-term public commitment to not only on-line data discovery but direct data access by client processes across internal and public networks. Clearinghouse provides a standards-based solution to catalog interoperability on the Internet today.

### Who should participate in Clearinghouse?

Although initially targeted at federal agencies, the NSDI Clearinghouse Network includes numerous federal, state, university, and tribal metadata collections. Hundreds of metadata servers are also in operation outside the United States supporting the same interoperability standards. In short, any group regardless of size may publish their metadata to the Clearinghouse and make it visible in geodata.gov. Similar publishing portals exist in other countries for the coordination and publication of geographic resources outside the U.S. The federated catalog behind the NSDI Clearinghouse Network is also registered with the Group on Earth Observation (GEO) and its Global Earth Observation System of Systems (GEOSS). Thus U.S. content is now also visible via the GEO Web Portal.

The role of the FGDC in Clearinghouse is to collect stakeholder requirements, design and deploy federated search, discovery, and access solutions for the U.S geospatial community. The Geospatial Platform, in concert with the data.gov initiative, provide community coordination of the Clearinghouse, catalog, and its contributions to visualization, analysis, and application development in the emerging Platform environment. It is not the intent of the FGDC to create a centralized data system but to facilitate access to agency-operated distributed stores of spatial metadata, data, and services on the Internet.

### What are the requirements for being a Clearinghouse provider and user?

A prospective spatial data publisher must have a public-facing web server with online access to metadata, catalogs, and spatial data. It is recommended that metadata services be co-located on hosts with spatial data collections to encourage synchronization between the spatial data, services, and the metadata being served. A publisher can share metadata through either 1) a Z39.50 server, 2) an OGC Catalog Server (CSW), or 3) a Web Accessible Folder (WAF) -- a browse-enabled directory on a host organization's web server that holds the XML metadata for direct harvest by the portal. An online registry is operated by the FGDC to track the operating details of existing Clearinghouse metadata services. Prospective users of Clearinghouse must have access to a current Web browser with a broadband connection to the Internet. Search and visualization interfaces exist at geo.data.gov and GeoPlatform.gov to provide custom levels of search access.

### What information is accessible through Clearinghouse?

A "digital geospatial data set" is the primary item being described with metadata in the Clearinghouse activity. The definition of a data set can be adjusted to meet a given agency's requirements but it generally corresponds to individual identifiable data products (e.g. file, layer, service) for which metadata are customarily collected. This may equate to a specific satellite image, a shapefile, or a national vector data set, as managed by a data producer or distributor. Collections of data sets (e.g. flight lines, satellite "paths", map or data series) may also have generalized metadata that could be inherited by individual data sets.

Other geospatial resources including online services (Web Map Service, Web Feature Service), data download locations, interactive web applications, documents, and other web-accessible resources. The Geospatial Data Presentation Form field in the metadata record can store this information, though other context can be inferred from the style of the URL.

### How does Clearinghouse work?

To provide search interoperability among different servers of geospatial metadata, the search and retrieve protocol known as ANSI Z39.50-1995 (ISO 23950) was initially selected by the FGDC Clearinghouse activity. Although in use by a few organizations today, it has been effectively replaced by the Open Geospatial Consortium (OGC) Catalog Services specification, more specifically the HTTP version known as Catalog Service for the Web (CSW). Multiple catalog services and metadata collections (WAF) are registered with the GeoPlatform.gov site. A periodic harvest of all metadata is performed, and all metadata are indexed for search, as if all the metadata and data resources were consolidated in one location, though they are actually distributed among the agencies. This federated model preserves the notion of 'data closest to source' allowing agencies full control of the content, metadata, and update frequency.

## 6.3. CRITICAL FACTORS AROUND SDIS

# 7. OPEN GIS

## 7.1. INTRODUCTION OF OPEN CONCEPT IN GIS

## 7.2. OPEN SOURCE SOFTWARE FOR SPATIAL DATA ANALYSIS

## 7.3. WEB BASED GIS SYSTEM

## 7.4. SYSTEM ANALYSIS AND DESIGN WITH GIS

### *What is Design?*

### *GIS Design Essentials*

The practice of design focuses on planning the structure and features associated with any system. Design activities can focus on something as simple as a coffee maker or as complex as an aircraft carrier. The product of the design practice is a plan that can be used for implementation. This is what is referred to as "the design." So there are really two things happening - there is a process associated with designing something, and then the product of that process is often called the *design*.

Additionally, there is a very wide range of depth and methodology associated with the practice of design. For some, design is a rapid activity that takes place as a solitary activity moments before implementation begins (a GIS application developer in a small consulting firm, for example). For others, design requires months or years of planning and iterates through multiple stages (a GIS project manager for a national emergency management organization, for example).

We'll be focusing on a range of topics in the next several lessons that reflect this range of design activities, so that, in the end, you have a broad range of tools available to you to deal with a diverse set of design situations.

In this course, our application focus is on designing GIS tools. All of you have substantial background knowledge about GIS systems, how they are applied, and what their capabilities include. To help frame things a bit, I'd like you to consider my argument that, in the GIS world, design centers on a few key components:

- **Users -** Who will be the target audience for your system?
- **Tasks -** What do the users expect to accomplish with the system?
- **Data -** Which source materials will be required to complete system tasks?
- **Products -** What outputs from the system are needed?

Those are a few of the key bits, anyway. I am sure you can think of others (or elaborate on the short definitions I've given here).

### *The Process of GIS Design*

The basic design process we'll focus on in this course is outlined in the graphic below. This depiction of the design process has four key stages, each of which is influenced in part by some sort of evaluation activity.

- **Needs Assessment** refers to the user and task requirement analysis stage where the goal is to identify the key components I listed above (users, tasks, data, and products).
- **Concept Development** takes the results from needs assessment and formalizes those results into specific design plans, which in the GIS case could include interface mockups and system programming architectures.
- **Prototyping** is an activity that demonstrates the design concepts in a form that can be readily evaluated.
- **Implementation** is the final phase of work that combines results from the previous design phases and results in execution of the complete, refined project.
- **Evaluation** is any activity intended to measure the success of a particular design phase. It is presented in the graphic below as something that occurs throughout each of the other discrete activities.

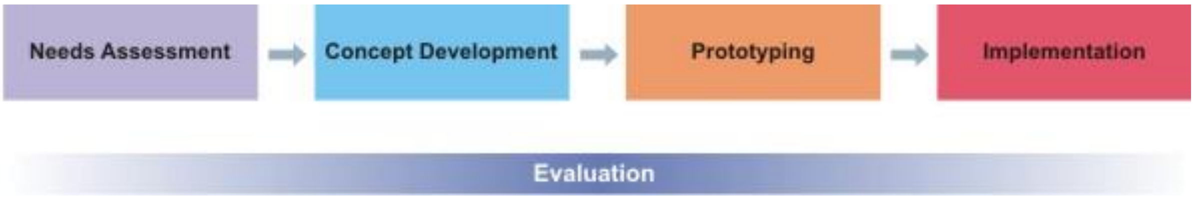



**Figure 1.03: A basic view of the system design and evaluation process.**
*Image credit: Anthony C. Robinson*

This process reflects my view of the design process based on my past research on the design and evaluation of tools for GIS and Geographic Visualization. There are, no doubt, plenty of other conceptualizations of design that are worth checking out, too - so fire up Google and do a bit of looking around if you're curious.

One **important caveat** here that I want to mention is that this process I have outlined is (and should be) an iterative process. Sometimes the results of concept development warrant new attention on needs assessment (for example, you find out that users didn't elaborate very well which tasks they need to complete). I see **evaluation** as the key driving force behind each iterative step - measuring your success in some way after each stage will help you decide where to go next, and in most real world design activities there is a lot of jumping around from stage to stage.

## When Design Goes Awry

So now that we've covered a few basics on good design, here are a few situations that can cause designs to fail:

**Little/No Design Effort -** This is probably the most common issue with respect to GIS design. Sometimes there just isn't any money in the budget to really spend time thinking out and evaluating what should be implemented. Some customers don't see the immediate value in spending money on what may be perceived as an intellectual effort, when, in fact, it is essential for success to have spent some serious attention on design issues.

**Design After The Fact -** Another common problem is the "Tool In Search Of An Application" that I'm sure all of you have encountered from time to time. Someone starts with a simple idea (e.g. a web mapping tool to disseminate emergency management information); a consulting group takes on the task and delivers what they think will work fine. Eventually, a real person uses the tools, and it becomes clear that the tools do something new and exciting, but not something terribly useful.

**Scope Creep -** Taking some time to design a new system can reveal all sorts of opportunities for new tools, data sources, output formats, etc... A common problem is managing all of the possibilities adequately so that the scope of the project does not continuously increase over time. The design focus may start with a relatively small problem area, and as momentum on the project builds, decision makers and stakeholders all chime in until eventually you are responsible for designing One System To Rule Them All that is all things to all people.

You have probably experienced problems like these (or others) on GIS efforts in your work experience. I encourage you to share your thoughts on these (and other) issues in our course discussions on this and other pages. We learn best when we learn with others!