Why use GIS?

Explain using relevant examples how the following government of Kenya ministries would utilize GIS to assist some of their day to day operations?

i. Ministry of devolution and planning
ii. Ministry of education
iii. Ministry of health
iv. Ministry of environment and natural resources
v. Ministry of transport and infrastructure
vi. Ministry of land, housing and urban development
vii. Ministry of energy and petroleum
viii. Ministry of tourism
ix. Ministry of water and irrigation
x. Ministry of mining
xi. Ministry of agriculture, livestock and fisheries.
xii. Ministry of defense

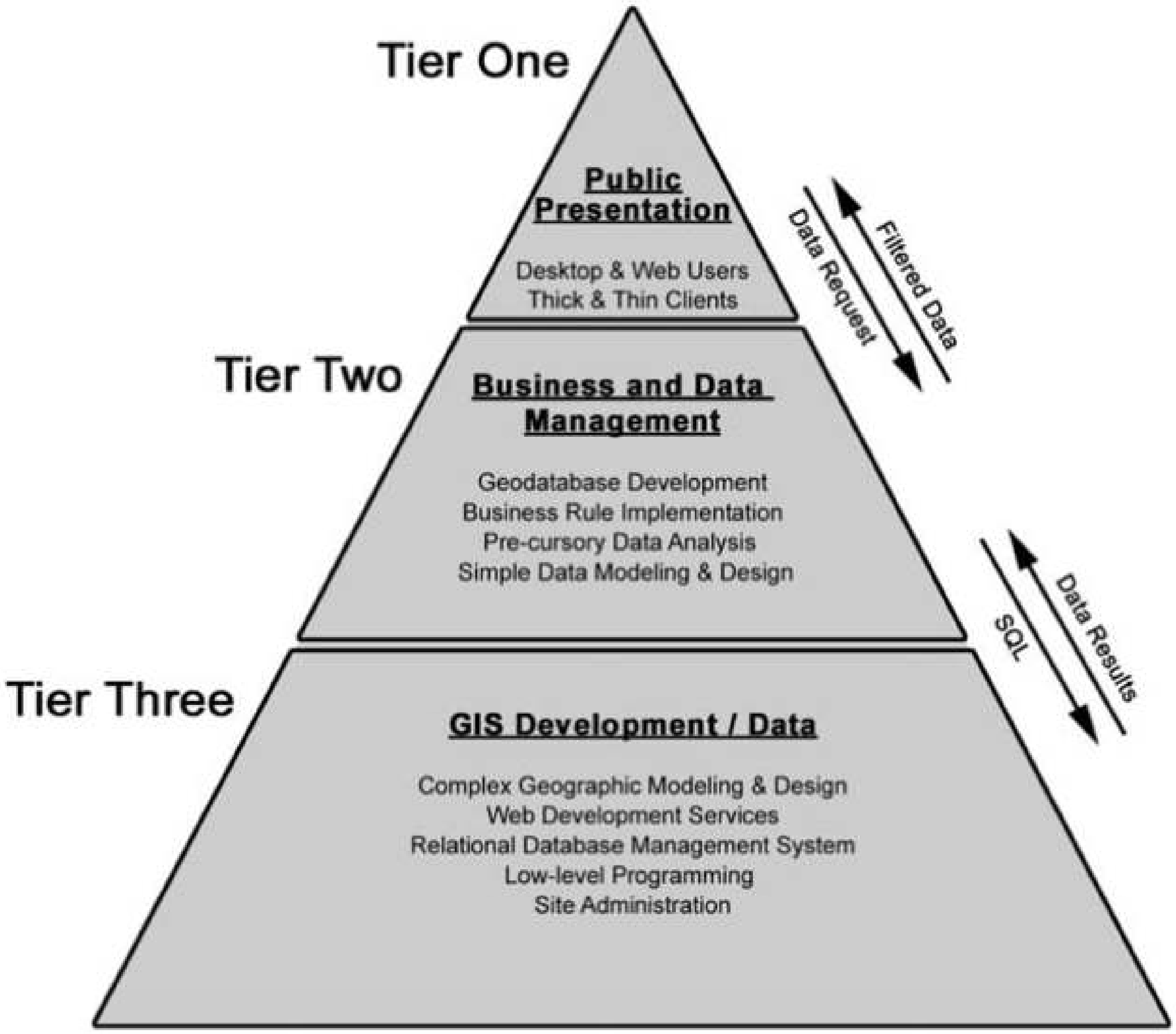
GIS Tiers

Geographic information systems follow the multitier system architectural type in that there are typically three tiers: a public presentation tier, a business and data management tier, and the GIS management tier.

The public presentation tier involves general viewing, Web-enabled data access, superficial analysis, and geographic visualization. This first tier is made up of general end users (or public thin clients). The business and data management tier involves users responsible for building the geographic database, site maintenance, critical analysis and visualization, and project management staff.

This second tier acts as a filter to what information can and will be presented to the general public, as well as creates the project-specific geodatabase. The GIS development tier involves GIS professionals and systems responsible for data modeling and design, Web site development, low-level programming, and site administration.

This third tier can also be called a data tier since it primarily functions to supply the entire database information.



The Structure of Geographic Data

Geographic information systems utilize two primary data models to manipulate and structure geographic data: the raster data model and the vector data model. We have already discussed both types briefly in Chapter 1 but now must build upon that initial introduction to gain a better understanding of both the elements and idiosyncratic nature of each type. By examining the nature of each data model you will have a much clearer picture of the important and institutional structure of geographic data, as well as a distinct understanding of the advantages and disadvantages of each.

Characteristics of both vector and raster data

Although distinctly different, these affiliate data structures share two characteristics:

(1) They visually represent real-world features, and

(2) they are subject to orientation within the real world.

By satisfying both of these characteristics, geographic data are born and made interoperable with other geographic data sources within GIS.

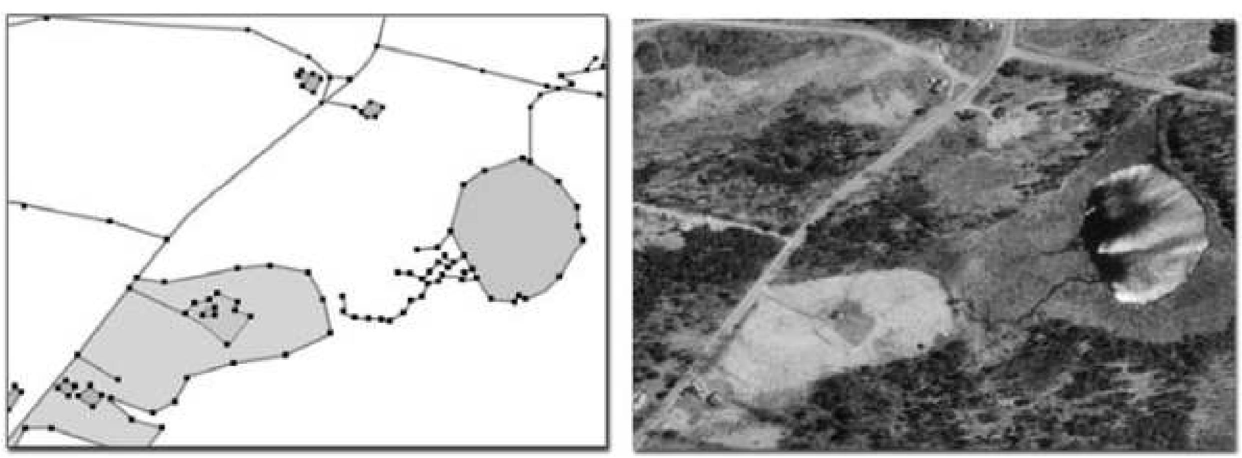
Raster data

Raster data structures characterize continuous data (such as imagery) and are exceptionally strong where boundaries and point information are not well defined. Raster data provide data as a pixel grid, whereby each pixel or cell is a feature capable of retaining properties and attributes. These pixels approximate pictures and images in an impressionistic way, with all of the smallish, monothematic cells contributing to a greater whole. Adding further identity, a raster image can vary in file format, color representation, resolution (size of pixels/number of pixels per set area), and potential properties.

Vector data

Vector data are a bit different. Vector data structures characterize discrete data (such as roads, pipelines and topographic features) and are exceptionally strong where distinct boundaries and point information are well defined. This data structure is constructed on ordered two- and three-dimensional coordinates ([*x*,*y*] and [*x*,*y*,*z*], respectively). Features are represented as geometric shapes defined through single or grouped coordinates on a set grid.

For a clearer understanding of both models, refer to Figure 1, which illustrates the visible differences between raster and vector data, as well as how each data set represents Earth features. The original graphic is a raster, high resolution satellite image of a land parcel with infrastructure and a pond. Notice the distinct feature variance between its continuous data image and vector's point representation. It is clear that the raster image serves as a good overall representation, while the vector representation serves as a good individual feature representation.



Vector representation Raster Representation

Figure I: Data Models

Both the raster and vector data structures have inherent advantages and disadvantages that allow GIS users a certain degree of choice. A full understanding of the native characteristics of each data model is a prerequisite for GIS success. Often certain requirements force the use of one data model rather than the other, such as the need for a better output resolution, easy image analysis, or enhanced spatial accuracy.

Advantages of Raster

Raster data, for example, offer a truly simple data structure that involves a grid of row and column data. This simple grid structure allows for easy raster image analysis, as well as analysis among multiple images. Raster modeling is also much easier to implement due to the single-value cell structure and relatively simple software programming.

Disadvantages of Raster

Disadvantages to raster data include general spatial inaccuracies and misrepresentations, low resolution, and massive data sets that require significant processing capability. The lack of accurate topology is also a major raster-based limitation.

Advantages of Vector

Vector data, are spatially accurate and support a better, higher resolution than the raster data model. The ability to provide topology or feature relationships is a definite advantage, as well as the minimal data storage requirements.

Disadvantages of vector

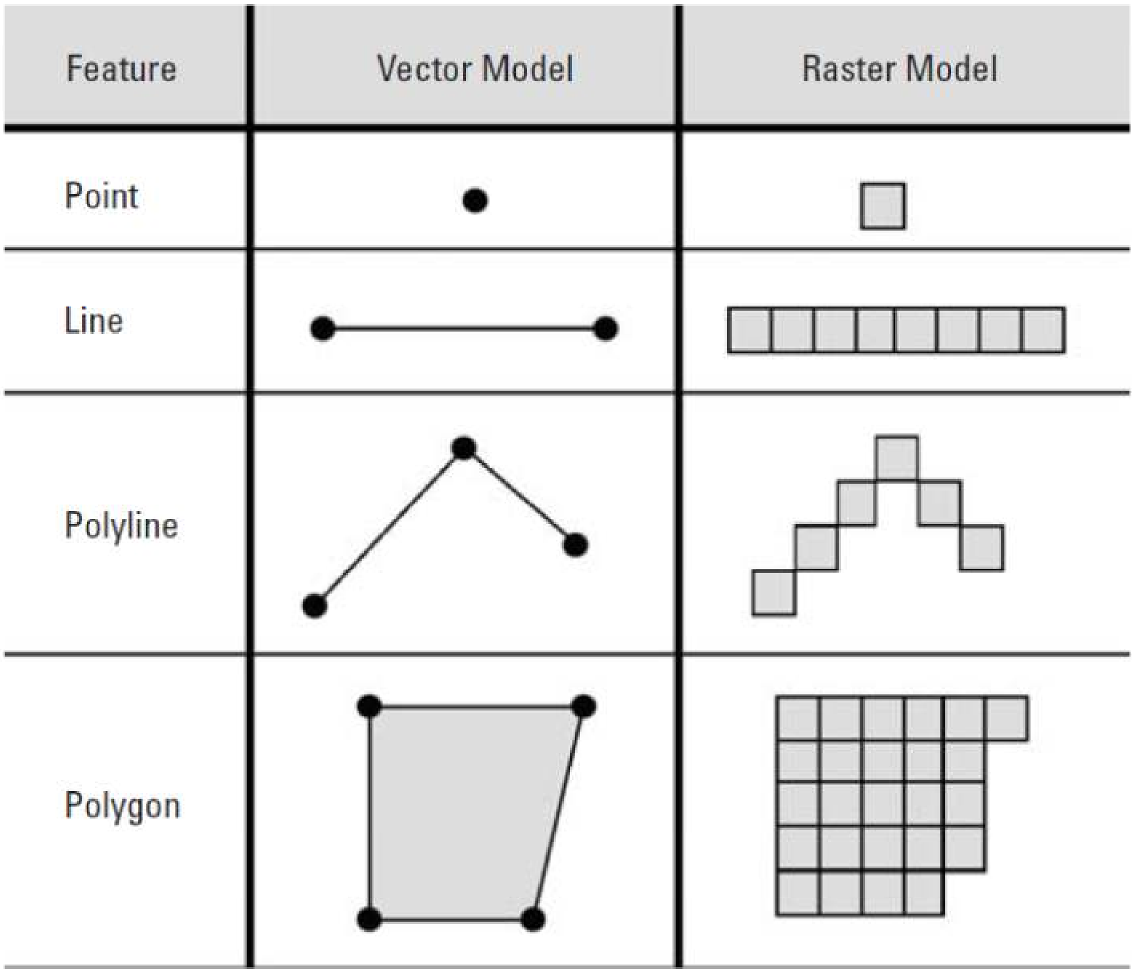
Due to the complex data structure, vector data require a greater and more powerful processing capability. With this comes the need for better, faster workstations to minimize the data processing times that typical computers face. Inevitably, the costs to run a vector GIS and expeditiously process complex geographic data sets can become highly expensive.

Needless to say, both the raster and vector data models have their own brand of geographic data expertise and capability embedded within a GIS. Nowadays, GISs are offered as raster-based systems, vector-based systems, or raster- vector capable systems. When necessary, raster-to-vector/vector-to-raster conversions can be easily accomplished through GIS or third-party conversion software. Undoubtedly, the choice of data model lies with the user and the available software/hardware.

Vector feature geometry

Real-world objects can be represented as individual or a group of geometric shapes called feature geometries.

In any geospatial platform, there are three primary types of feature geometries: points, lines, and polygons. As a subset of these three primary types there exists a fourth geometric feature called a polyline. Figure 2 illustrates these major feature geometries in both the raster and vector data models. The differences between and usefulness of the models are obvious as they relate to depicting feature geometry.

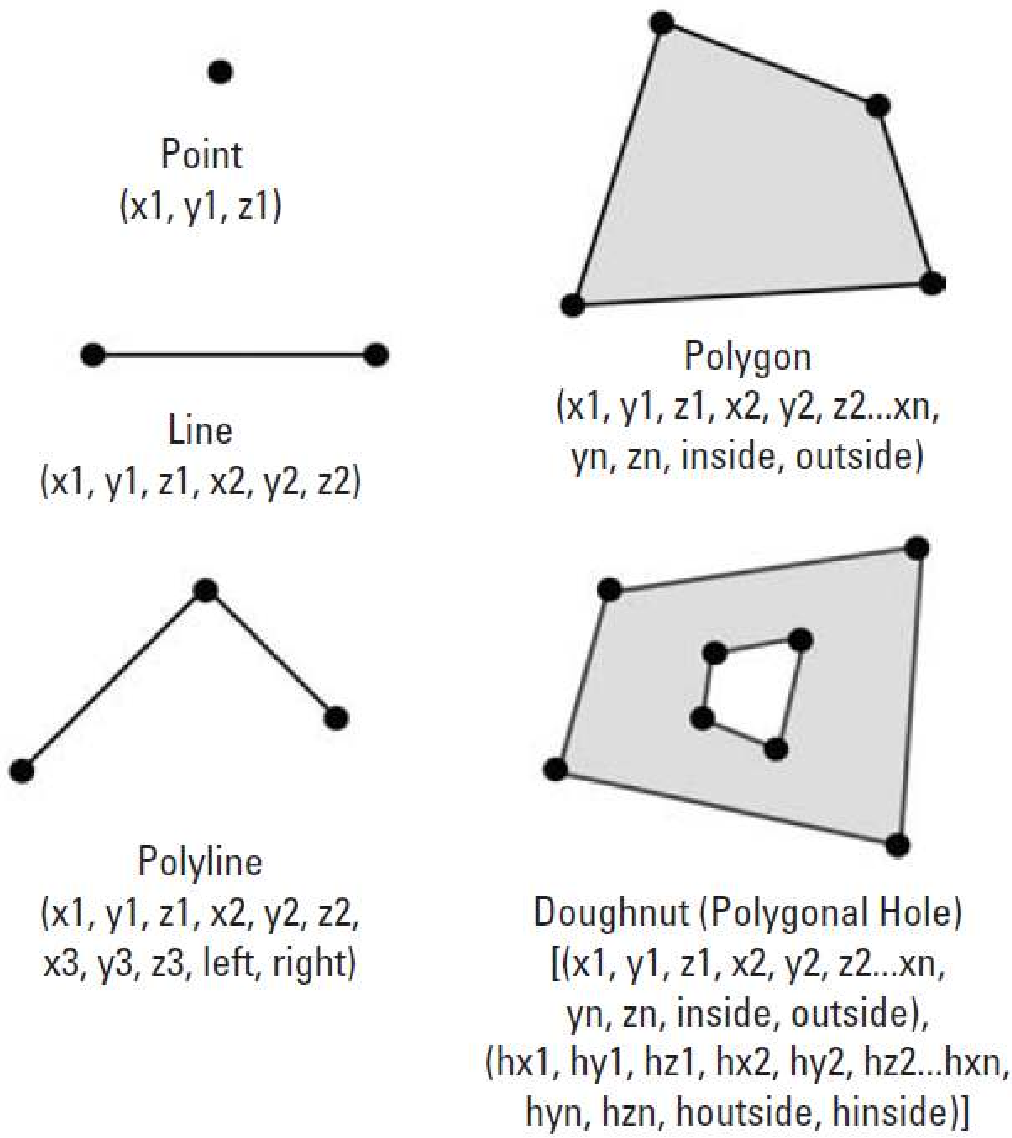


Feature geometries in raster and vector.

A point is an individual position defined as a vector *x-y-z* coordinate or as one raster pixel. Lines are two connecting points with two distinct coordinates in vector or a linear block of pixels in the raster model. Polygons are a grouping of vector coordinates connected in a sequential fashion or a group of pixels forming the object's general shape. Lastly, the subset polyline (known also as an arc) is a connected string of vector points or raster pixels. Often a polyline involves two lines sharing a same point or pixel (vertex).

Vector features are constructed on ordered pairs of vertices*,* whereby these ordered pairs reside in the design plane having known *x*, *y*, *z* locations. The positions of these ordered pairs are recorded in the vector file and often encoded as binary. These vertices are objects and can be grouped to form features that exist as independent entities. This is in contrast to the raster model, where the entire image is an object. In raster, features must first be sampled and then represented as image pixels, which, as we already know, visually approximate the features. Raster images have no independent vector ordered pairs, the actual building blocks of feature-driven spatial information.

Vector data for feature geometry comprise positional coordinates and, for some, other position-defining data, such as inside/outside and left/right. Figure 3 depicts the three primary types of geospatial feature geometries, as well as the polyline and the more complex polygon within a polygon (i.e., the "doughnut").



Major vector feature geometries.

As illustrated in the figure, the point depicts a discreet position in space, defined by an individual geospatial coordinate. The line is defined by two geospatial coordinates. The open-ended third feature is a polyline, which comprises two vector lines sharing a common point (vertex). Interesting enough, polylines are defined not only by geospatial coordinates, but also by left and right characteristics that indicate whether other features are directly left or right of the polyline.

Growing in complexity, the remaining two geospatial feature geometries in Figure 3.3 are closed features and depict specific areas. The polygon involves numerous vector points that are connected in sequence. Polygons are defined with inside and outside characteristics that delineate whether any overlapping geospatial feature geometries exist inside the polygon's boundaries.

The most complex of the geospatial feature geometries presented in Figure 3 is the doughnut or polygonal hole. A doughnut comprises a polygon within another polygon. Polygon 1 forms the boundary area and polygon 2 serves as the "cutout" within polygon 1. Think of a doughnut, whereby polygon 1 is the outermost doughnut boundary and polygon 2 is the cutout in the doughnut center (doughnut hole). Each polygon enables coordinate characteristics for inside and outside elements.

To uniformly structure, manage, and manipulate feature geometries like the ones just discussed, a GIS stores vector data in geospatial feature file formats. Common feature file formats, such as the *shapefile* developed by Environmental Systems Research Institute (ESRI) or the *TAB file* developed by MapInfo, are basic geometric containers compatible with an overwhelming majority of GISs.

These geographic dataset types store non-topological vector (or coordinate) geometry in the form of real-world spatial features, as well as links to attribute information for these respective objects.

Raster Image Structures

The nature of images, such as aerial photographs and base maps, involves a continuous array of data. As discussed, the raster model provides the best solution for continuous data. Due to the lack of specified feature boundaries, point locations, and multiple objects, vector data are not intrinsically fit to handle the image information. A vector data representation of such an image produces a complex and often massive image structure.

Unlike a vector format, the raster palette is much more desirable for applications requiring the use of photographs or digital scans. Raster imagery instantiates outrageously complex geometries quite easily with the use of attributed image pixels. The essence and value of raster image structures are that raster represents real-world information with native visual properties—a feat vector imagery cannot reproduce.

Because raster images are physically continuous by nature, a reasonably accurate transformation process can be applied to fit raster data onto the real world. In brief, images are "rasterized" or digitally transformed to raster data through a matrix of pixels. Photographs are typically scanned with a set imageresolution defined by pixels per inch (ppi), more commonly known as dots per

inch (dpi).

With the use of a GIS, the organic transformation into a raster image can enable users to characterize previously nonattributed existing geographic data material. In fact, solutions produced from this process result in alternative, often unexplored, geographic data sets.

There are times when an existing raster image is not at the desired resolution for the present application and an image adjustment is necessary. Manipulating resolution within raster images is a technique to adequately control and manage the amount of raster data to be processed and the overall image quality.

As mentioned earlier, raster data resolution is the result of the amount of row and column pixels used to define an image. Raster image dataset size is directly proportional to the amount of raster pixels used to define an image. The rule of thumb is: The more pixels in the grid, the higher the image resolution, quality, and dataset size; the fewer pixels in the grid area, the lower the image resolution, quality, and dataset size.

Modifying the image's resolution presents varying results within the image. You can easily transform a high resolution raster into a low resolution raster without desecrating image quality. The image size often remains unchanged through an intelligent interpolation that transforms a grouping of pixels into one pixel. However, it is not always easy to reverse this transformation.

Taking a low resolution into a high resolution form is not always feasible. To retain moderate image quality, a high resolution raster will take on a reduced, fractional size to that of the original low resolution raster. If image size remains unchanged, the new high resolution image will be relatively unusable and indistinguishable in content. Care should always be taken when modifying the resolution of raster images.

In actuality, geographic information obtained from these types of organic processes may be of critical relevance to future projects and unwittingly provide solutions to problems not yet known. Many times these unforeseen criticalities involve complications put forth by the presence of precious natural resources on a project. The manifestation (or realization) of these transformations is the fundamental building block for geospatial systems and makes available opportunity for other advanced GIS techniques.

Topology

It has already been established that in a vector-based GIS, primary geometries (i.e., points, lines, and polygons) represent real-world features. Topology is the set of rules through which a GIS represents features with the primary geometric shapes (i.e., point, line, and polygon). The vector data model utilizes topology to organize spatial relationships between discrete features. In essence, the main functions of topology are to define:

(1) feature-to-feature locality or, simply, where a feature is in relation to another feature,

(2) what is shared between different features, and

(3) how features are grouped or connected within a set.

In a GIS, topology establishes geometric harmony within a geographic data set. Illustrating this precept is ESRI's *shapefile*, in which the vector feature file employs a set of natural numbers plotted and structured in binary format.

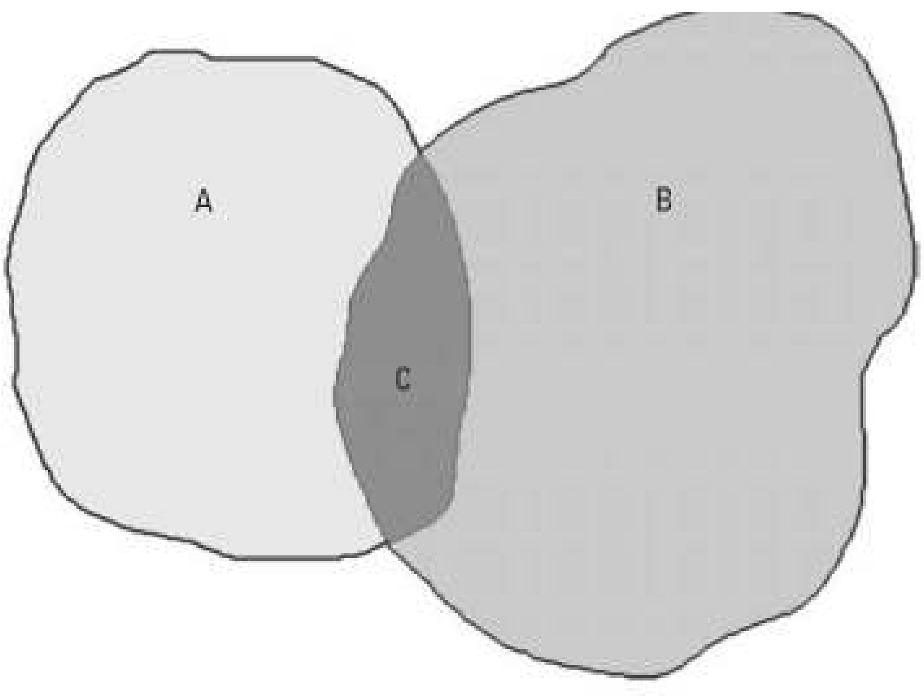
This set of natural numbers delineates the boundaries and coverage of a particular feature. The simple binary structure defines a topological space for the feature and establishes continuity with other features, forming topological relationships.

Topological relationships are defined in all types of feature files and are generally categorized into the three primary functions of topology (previously mentioned):

1. Feature-to-feature locality, called a *complement;*
2. What different features share, called an *intersection;*
3. How features are grouped, called a *union.*

In Figure 4, the simplistic Venn diagram depicting two overlapping, amorphous shapes highlights the base concepts of the three feature topology functions.

Within this figure, the lighter areas are considered *complementary* shapes (*A* and *B*), while the darkest area depicts the *intersection* of the two shapes (*C*).



Venn diagram.

All three areas together (*A*, *B*, and *C*) form the *union* of the set. Applying these concepts to a vector GIS, shapes *A* and *B* can be considered two separate features that share topologic space (*C*).

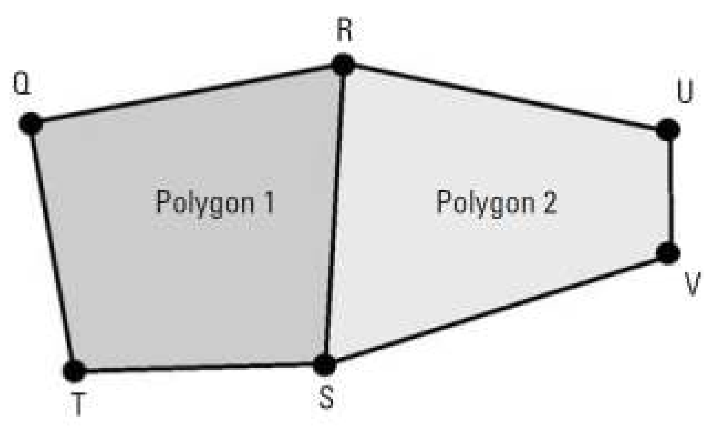
Feature topology is also very sensitive to retaining the original shape of different features and the demarcation of each individual feature. In the context of traditional mathematics, topology is the examination of objects and groupings of objects that exhibit a predictable structure. Put simply, topology involves object aspects that remain unchanged even when the object itself is under some form of physical transformation (i.e., intersection). Similarly within the figure, shape *A* maintains its form even when sharing topological space with shape *B*.

All types of geographic data sets can utilize topology within geospatial software environments. These predictable structures (such as shapes *A* and *B*) constitute a set of features (*N*), which represents whole geospatial features.

Topology implements a successor function: $s(x)\ x + 1$, which succinctly indexes each separate feature for identification (FID). The successor function enables unique identification of each feature within the set. In this way topology serves as a geographic data quality overseer by ensuring the geometric integrity of geographic features between the real world and GIS, as well as retaining responsibility for producing truly clean and representative geographic data products.

Topology also helps to avoid repeating feature data, such as shared boundaries and shared nodes (points). The data model stores a single line to represent a boundary, as opposed to two lines with the same coordinates. This topological quality control helps maintain a smaller data set and vector feature file.

In Figure 5, polygon 1 (points *QRST*) and polygon 2 (points *RUVS* are side by side, connected by a shared border (*RS*). Topology dictates that line *RS* for polygon 1 and line *RS* for polygon 2 are identical and the data model only accounts for *RS* once. Feature topology proves diligent in avoiding topological overlap.



Topology of two polygons.

These geometric monitoring techniques allow GIS to control, query, and edit the topological coincidence between geospatial features (objects). Topology introduces the notion of absolute feature continuity, opening the door for numerous potential software compatibilities, including complex mathematics and engineering programs.