Chapter 2: Data models

Meng Lu

Juniorprofessor of Geoinformatics, Bayreuth University

1. Introduction
2. Common Spatial Data Models
3. Vector Data Models
4. Raster Data Models
5. Other data models

Introduction

In GIS, the data are:

A simplified view of physical entities (e.g. the roads, mountains, accident locations).
Data include information on the spatial location and extent of the entities, and information on their nonspatial properties.

Each entity is represented by a spatial feature or object. A subset of essential characteristics is recorded -> entities are simplified/abstracted in an object.

For example, the next figure in the left side (original) can be represented by a set of polygons (right side) (Figure 1)

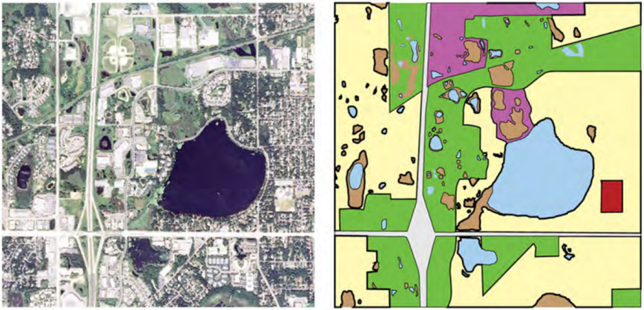
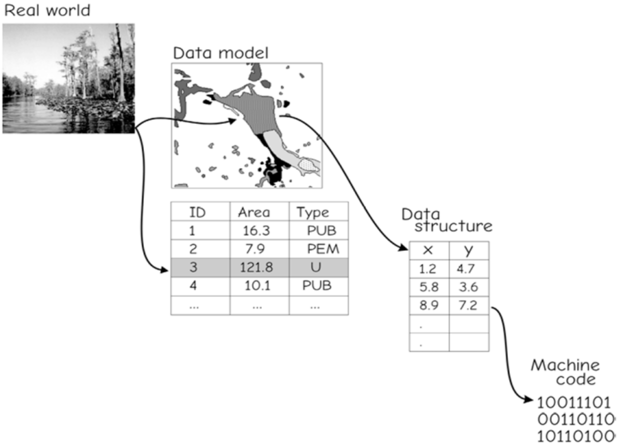


Figure 1: A physical entity is represented by a spatial object in a GIS. Here, lakes (dark areas) and other land cover types are represented by polygons (Bolstad (2016)).

A spatial data model could be defined as the objects in a spatial database plus the relationships among them.

- Providing a formal means of representing and manipulating spatially referenced information (Figure 2).

Levels of abstraction in the representation of spatial entities.



ID	Area	Type
1	16.3	PUB
2	7.9	PEM
3	121.8	U
4	10.1	PUB
...	...	...

x	y
1.2	4.7
5.8	3.6
8.9	7.2
.	
.	

Figure 2: Levels of abstraction in the representation of spatial entities. The real world is represented in successively more machine-compatible but humanly obscure forms (Bolstad (2016)).

GIS commonly store our data as a set of layers (figure 3). Each layer for a kind of cartographic objects. These are often referred to as thematic layers.

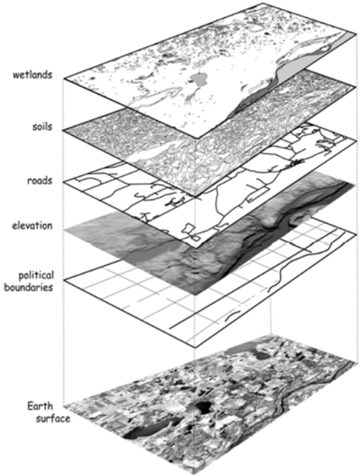


Figure 3: Spatial data are often stored as separate thematic layers, with objects grouped based on a set of properties (e.g. water, roads) (Bolstad (2016)).

Coordinate data

Coordinates define location in two or three-dimensional space.

Spatial data in GIS most often use coordinate pairs (x, y) or triplets (x, y, z) in a Cartesian coordinate system (4).

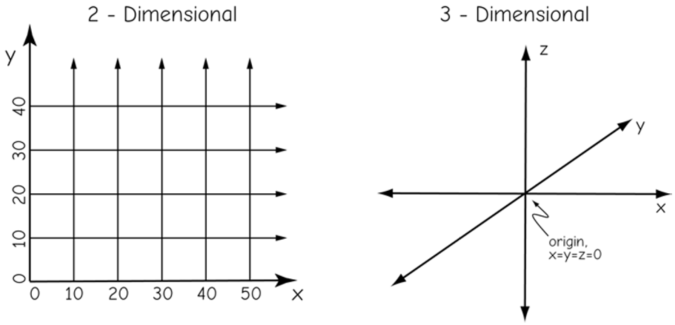


Figure 4: Two dimensional (left) and three dimensional (right) Cartesian coordinate systems (Bolstad (2016)).

Coordinates on a Sphere

The most common spherical system uses two angles of rotation, longitude (ϕ) and latitude (λ), with a radius, r, to specify locations on a modeled earth surface.

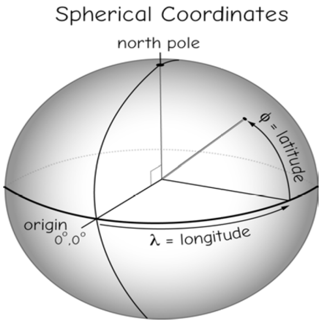


Figure 5: Three dimensional, spherical coordinates may define location by two angles of rotation, λ and ϕ, and a radius vector, r, to a point on a sphere (Bolstad (2016)).

Spherical system for geographic coordinates is non-Cartesian, so we need to project between two systems (will be introduced in the following chapter).

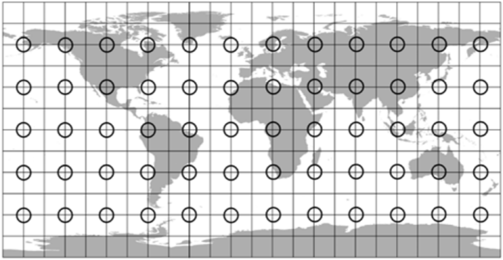


Figure 6: Left: geographic coordinates on a spherical representation, notice that circles defined by a 5° radius do not form circles on the Earth's surface near the poles. Right: a distorted Cartesian representation (Bolstad (2016)).

- The distance spanned by a degree of longitude varies from south to north. A degree of longitude spans approximately 111.3 kilometers at the equator, but 0 kilometers at the poles. In contrast, the ground distance for a degree of latitude varies only slightly, from 110.6 kilometers at the equator to 111.7 kilometers at the poles.

There are two primary conventions used for specifying latitude and longitude, the first one is:

- Using letters N, S, E, or W to indicate direction, followed by a number to indicate location. Northern latitudes are preceded by an N (N90°) and southern latitudes by an S (S10°). Longitude values are preceded by an E or W, respectively; for example W110°

The second one is signed coordinates convention for specifying latitude and longitude in a spherical system.

- Northern latitudes are positive and southern latitudes are negative, and eastern longitudes positive and western longitudes negative. Latitudes vary from -90 degrees to 90 degrees, and longitudes vary from -180 degrees to 180 degrees.

DMS (Degrees-minutes-seconds) to DD (decimal degrees) conversion Spher-

ical coordinates are commonly recorded in a degrees-minutes-seconds (DMS) notation; N43° 35$'$ 20$''$ for 43 degrees, 35 minutes, and 20 seconds of latitude.

In DMS each degree is made up of 60 minutes of arc, and each minute is in turn divided into 60 seconds of arc (Figure 7).

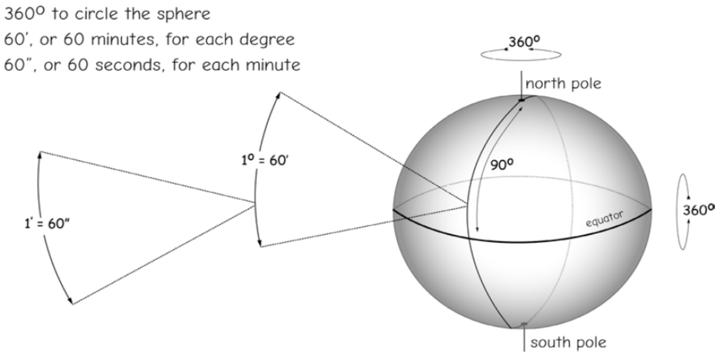


Figure 7: There are 360 degrees in a complete circle, with each degree composed of 60 minutes, and each minute composed of 60 seconds (Bolstad (2016)).

Spherical coordinates also can be converted to decimal degrees (DD). So, for DD the degrees take the usual -180 to 180 (longitude) and -90 to 90 (latitude) limits, but minutes and seconds are reported as a decimal portion of a degree (from 0 to 0.99999...). So, from DMS to DD:

$$\mathrm{DD} = \mathrm{DEG} + \mathrm{MIN}/60 + \mathrm{SEC}/3600 \tag{1}$$

For example, for N43° 35′ 20″, the conversion is 43.5888.

The next scheme shows the conversions between DD from DMS and vice-versa (Figure 8).

DD from DMS	DMS from DD
DD = D + M/60 + S/3600 e.g. DMS = 32° 45' 28" DD = 32 + 45/60 + 28/3600 = 32 + 0.75 + 0.0077778 = 32.7577778	D = integer part M = integer of decimal part x 60 S = 2nd decimal x 60 e.g. DD = 24.93547 D = 24 M = integer of 0.93547 x 60 = integer of 56.1282 = 56 S = 2nd decimal x 60 = 0.1282 * 60 = 7.692 so DMS is 24° 56' 7.692"

Figure 8: Examples for converting between DMS and DD expressions of spherical coordinates (Bolstad (2016)).

An Ellipsoidal Earth

Earth is better approximated as an ellipsoid than a sphere (Figure 9).

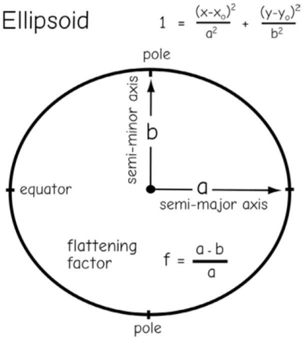


Figure 9: Ellipsoidal shape, a spherical figure with unequal semi-major and semi-minor axes (Bolstad (2016)).

The best estimate for a is 6,378,137.0 meters (m), and for b 6,356,752.3 m. When a simple spheroidal shape is assumed where a equals b, some number between these two is usually applied, often the mean value of 6,367,444.7 m.

The "flattening" is quite small, however, is significant enough to need modeling for the most precise measurements and navigation on the surface of the Earth. Many navigation and measurement estimates have two sets of formulas, one an approximation based on a purely spherical globe, and a more precise, and much more complicated set based on an ellipsoidal shape.

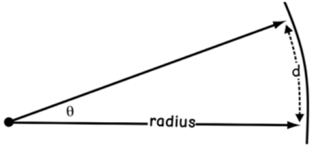
Converting Arc to Surface Distances

If we have two sets of coordinates that differ by 10 seconds of arc, and wish to estimate the distance between them, then we can approximate the surface distance on a circle or sphere by:

$$d = r * \theta, \tag{2}$$

where d is the approximate ground distance, r is the radius of the circle or sphere, and θ is the angle of the arc (defined as 2π radians per the 360 degrees in a complete circle).

Example calculation of arc length, using the average earth radius:



d = radius • θ

where θ is measured in radians,
with
1 radian = 57.2957°

Given an Earth radius of 6,378,137m, how much distance is spanned by 10" of arc?

Arc= 10"/3600"/1° = 0.00277778°

=0.00277778°/57.2957 degrees/radian
= 0.000048481435 radians

d = 6378137m • 0.000048481435
= 309.2 meters

Figure 10: Calculation of the approximate surface distance spanned by an arc (Bolstad (2016)).

The great circle distance should be used to estimate the surface distance between two points with latitude/longitude. Here, *great circle* is defined as any line resulting from the intersection of a plane passing through the center of a globe.

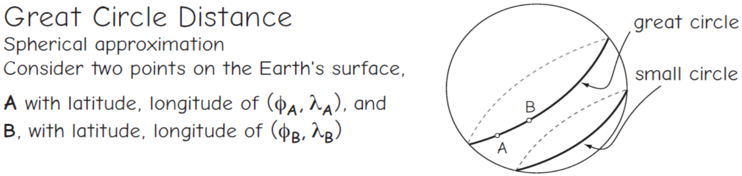


Figure 11: Calculation of the approximate surface distance spanned by an arc (Bolstad (2016)).

Given by:

$$d = r * \cos^{-1}[\sin(\phi_A) * \sin(\phi_B) + \cos(\phi_A)\cos(\phi_B) * \cos(\lambda_A - \lambda_B)], \quad (3)$$

where d is the shortest distance on the surface of the Earth from A to B, and r is the Earth radius, approximately 6378 km.

Example: Distance between Paris (France) and Seattle (USA):

- Latitude - Longitude France: 48.864716°, 2.349014°
- Latitude - Longitude Seattle: 47.655548°, -122.30320°

Solution

$$\begin{aligned} d &= 6378 * \cos^{-1}[(\sin(48.864716) * \sin(47.655548) + \cos(48.864716) * cos(47.655548) * cos(2.349014 - (-122.30320))] \\ &= 8,043.6558km \end{aligned}$$

Given by:

$$d = r * \cos^{-1}[\sin(\phi_A) * \sin(\phi_B) + \cos(\phi_A)\cos(\phi_B) * \cos(\lambda_A - \lambda_B)], \quad (3)$$

where d is the shortest distance on the surface of the Earth from A to B, and r is the Earth radius, approximately 6378 km.

Example: Distance between Paris (France) and Seattle (USA):

- Latitude - Longitude France: 48.864716°, 2.349014°
- Latitude - Longitude Seattle: 47.655548°, -122.30320°

Solution

$$\begin{aligned} d &= 6378 * \cos^{-1}[(\sin(48.864716) * \sin(47.655548) + \cos(48.864716) * cos(47.655548) * cos(2.349014 - (-122.30320))] \\ &= 8,043.6558\text{km} \end{aligned}$$

Conversion from Geographic to 3-Dimensional Cartesian Coordinates

Sometimes we want to convert between latitude/longitude coordinates and a 3-D Cartesian coordinate system. In this case the Cartesian system is aligned with the Z axis through the geographic North Pole, and the X and Y axes forming a plane on the equator.

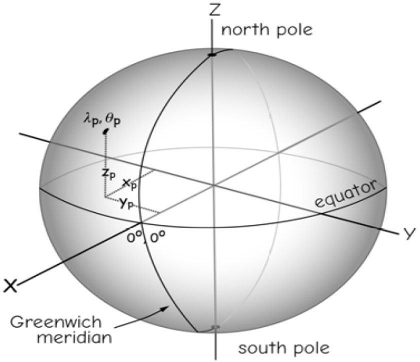


Figure 12: Spherical geographic coordinates (latitude and longitude on a spheroid) and corresponding 3- D Cartesian coordinates (Bolstad (2016)).

Formulas exist to convert between known spherical geographic coordinates (latitude and longitude) and corresponding 3- D Cartesian coordinates:

Lon-lat from known 3-D Cartesian:

$$\lambda_p, \theta_p = F(X_p, Y_p, Z_p) \tag{4}$$

3-D Cartesian from known lon-lat:

$$X_p, Y_p, Z_p = G(\lambda_p, \theta_p) \tag{5}$$

Geographic and Magnetic North

- Magnetic North: location towards which a compass points.
- Geographic North Pole: the average northern location of the Eearth's axis of rotation.
- Magnetic North wanders through time, and has recently increased it's rate of shift.
- How does the Earth wobble: https://www.youtube.com/watch?v=peUrvFFC6Zc

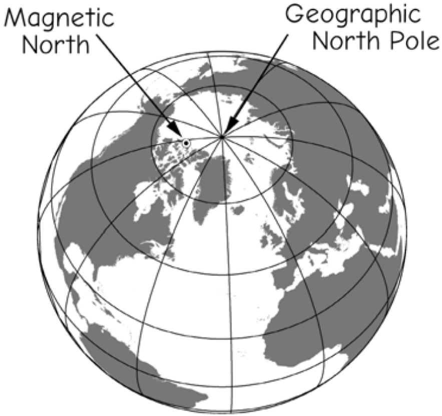
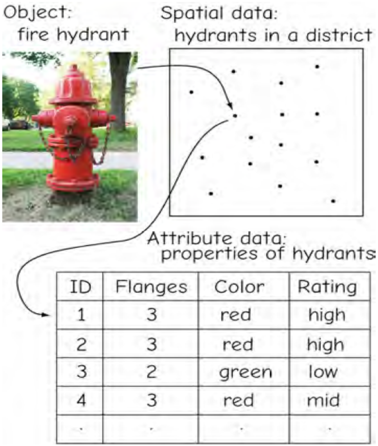


Figure 13: Magnetic north and the geographic North Pole (Bolstad (2016)).

Attribute Data and Types

Attributes are often presented in tables where each row corresponds to a spatial object, and each column corresponds to an attribute (Figure 14).



ID	Flanges	Color	Rating
1	3	red	high
2	3	red	high
3	2	green	low
4	3	red	mid
.	.	.	.

Figure 14: Generally the attributes are envisioned as arranged in columns and rows (Bolstad (2016)).

The attributes can be can be categorized as nominal, ordinal, or interval/ratio attributes.

- *Nominal attributes* variables that provide descriptive information about an object. There is no implied order, size, or quantitative information contained in nominal attributes. Nominal attributes may also be images, film clips, audio recordings, or other descriptive information.
- *Ordinal attributes* imply a ranking or order by their values. An ordinal attribute may be descriptive, such as high, mid, or low, or it may be numeric. The order reflects only rank, and not the scale.
- *Interval/ratio* attributes are measured along a scale. Interval data always appears in the form of numbers or numerical values where the distance between the two points is standardized and equal.

Common Spatial Data Models

Spatial data models begin with a conceptualization, and there are two main conceptualizations used for digital spatial data, the first one:

- Vector data model: Vector data models use discrete elements such as points, lines, and polygons to represent the geometry of realworld entities.

Spatial data models begin with a conceptualization, and there are two main conceptualizations used for digital spatial data, the first one:

- Vector data model: Vector data models use discrete elements such as points, lines, and polygons to represent the geometry of realworld entities.

The second common conceptualization identifies and represents grid cells for a given region of interest:

- Raster data model: Raster cells are arrayed in a row and column pattern to provide wall-to-wall coverage of a study region. Cell values are used to represent the type or quality of mapped variables. The raster model is used most commonly with variables that may change continuously across a region.

The second common conceptualization identifies and represents grid cells for a given region of interest:

- Raster data model: Raster cells are arrayed in a row and column pattern to provide wall-to-wall coverage of a study region. Cell values are used to represent the type or quality of mapped variables. The raster model is used most commonly with variables that may change continuously across a region.

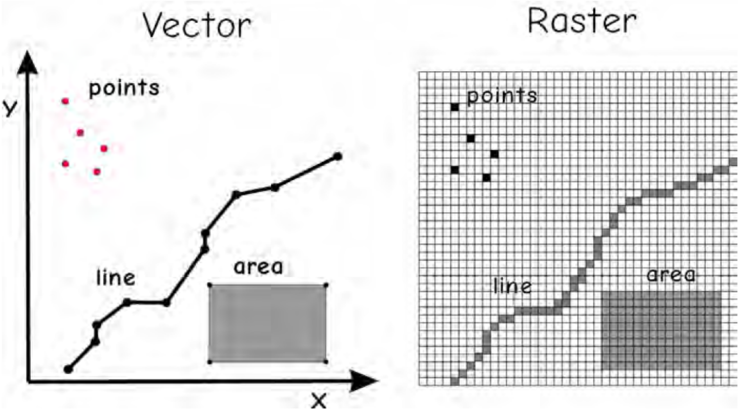


Figure 15: Vector and raster data models (Bolstad (2016)).

Data models are at times interchangeable in that many phenomena may be represented with either the vector or raster approach. The decision to use either a raster or vector conceptualization often depends on the most frequent operations performed.

The best data model for a given application depends on the most common operations, the experiences and views of the GIS users, the form of available data, and the influence of the data model on data quality.

Others data models are also used. A triangulated irregular network (TIN) is an example of such a data model, employed to represent surfaces, such as elevations, through a combination of point, line, and area features.

Vector Data Models

A vector data model uses sets of coordinates and associated attribute data to define discrete objects.

There are three basic types of vector objects:

1. Points

Uses a single coordinate pair to represent the location of an entity that is considered to have no dimension.

2. Lines

Linear features are represented as lines when using vector data models. Lines are most often represented as an ordered set of coordinate pairs.

3. Polygons

Polygons have an interior region and may entirely enclose other polygons in this region.

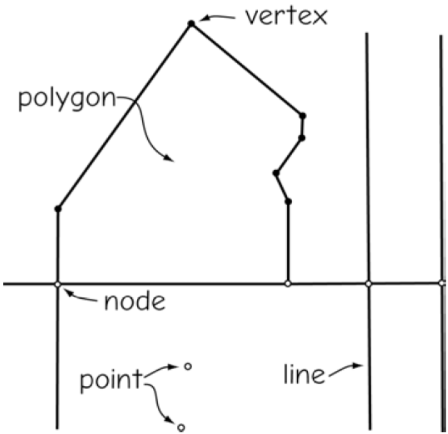


Figure 16: Points, nodes, and vertices define points, line, and polygon features in a vector data model (Bolstad (2016)).

Thus, a single set of features may be represented differently, depending on the interests and purposes of the GIS users (Figure 17).

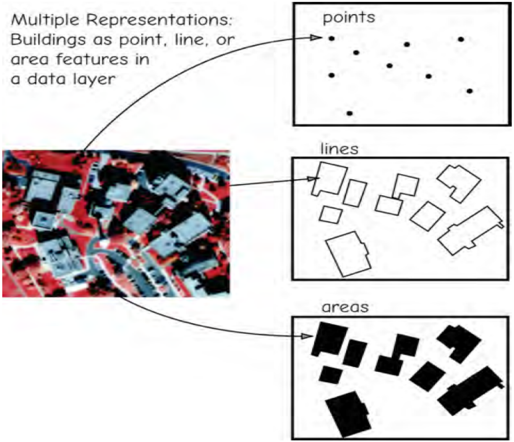
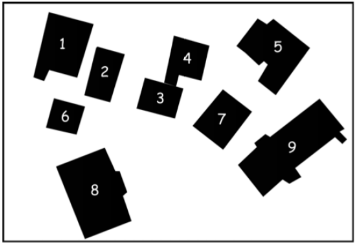


Figure 17: Alternate conceptualizations may be used to represent features (Bolstad (2016)).

Single-part and multi-part features

Single-part features: a single row in the table corresponding to each feature in the data layer (Figure 18).



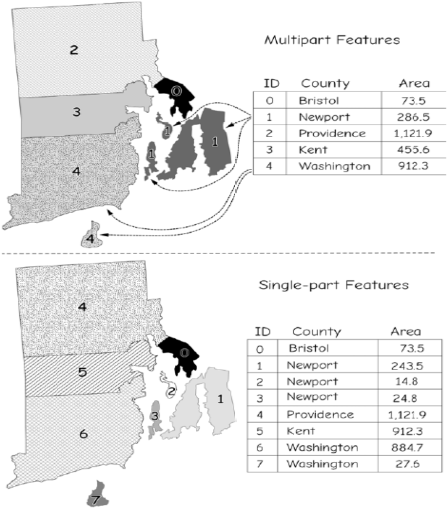
ID	Building Name	Floors	Roof Type
1	Hodson Hall	6.0	flat, sealed tar
2	Borlaug Hall	5.5	pitched 9/12, tile
3	Guilford Technology Bldg.	4.0	flat, gasket
4	Shop Annex	2.5	flat, sealed tar
5	Animal Sciences Bldg.	1.0	pitched 12/12, tile
6	Administration Bldg.	14.0	pitched 6/12, metal
7	Climate Sciences Center	6.0	flat, sealed tar
8	Grantham Tower	1.0	pitched, 9/12, tile
9	Biological Sciences Bldg.	9.0	pitched 12/12, tile

Figure 18: Geographic features correspond to rows in a table with an identifier (ID) and a set of attributes arrayed in columns (Bolstad (2016)).

Multi-part features: Vector layers sometimes have a **many to one** relationship between geographic features and table rows (Figure 19). In these instances, many spatially distinct features are matched with a row, and the row attributes apply to all the distinct features. This is common when representing:

- Islands
- Groups of buildings
- Other clusters of features

These are referred to as *multipart features*, because multiple geographic objects may correspond to one row.



ID	County	Area
0	Bristol	73.5
1	Newport	286.5
2	Providence	1,121.9
3	Kent	455.6
4	Washington	912.3

ID	County	Area
0	Bristol	73.5
1	Newport	243.5
2	Newport	14.8
3	Newport	24.8
4	Providence	1,121.9
5	Kent	912.3
6	Washington	884.7
7	Washington	27.6

Figure 19: Example of multipart and single-part features (Bolstad (2016)).

Care is warranted when converting multipart features to single-part features. The most common problems arise for aggregate variables in polygon layers, such as total counts.

Figure 20 shows how the population is associated with the aggregated set of polygons corresponding to a state.

Care is warranted when converting multipart features to single-part features. The most common problems arise for aggregate variables in polygon layers, such as total counts.

Figure 20 shows how the population is associated with the aggregated set of polygons corresponding to a state.

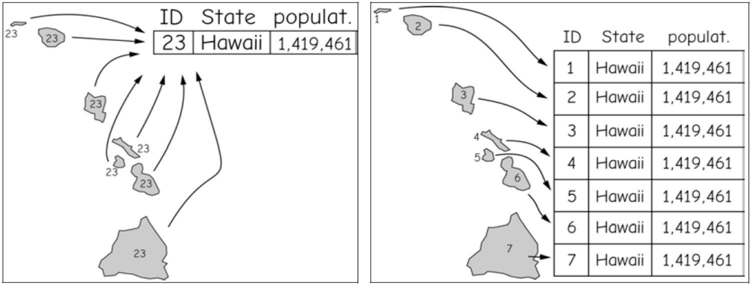


Figure 20: Multipart to single-part conversion may lead to errors in subsequent analysis because attributes may be copied from the original, multi-part cluster (left, above), to each single-part component (above right) (Bolstad (2016)).

Polygon Inclusions and Boundary Generalization

Vector data frequently exhibit two characteristics, polygon inclusions and boundary generalization.

- Polygon inclusions are areas in a polygon that are different from the rest of the polygon, but still part of the polygon (Figure 21, **a)** and **b)**).
- Boundary generalization is the incomplete representation of boundary locations. This problem stems from the typical way we represent linear and area features in vector data sets (Figure 21, **c)**).

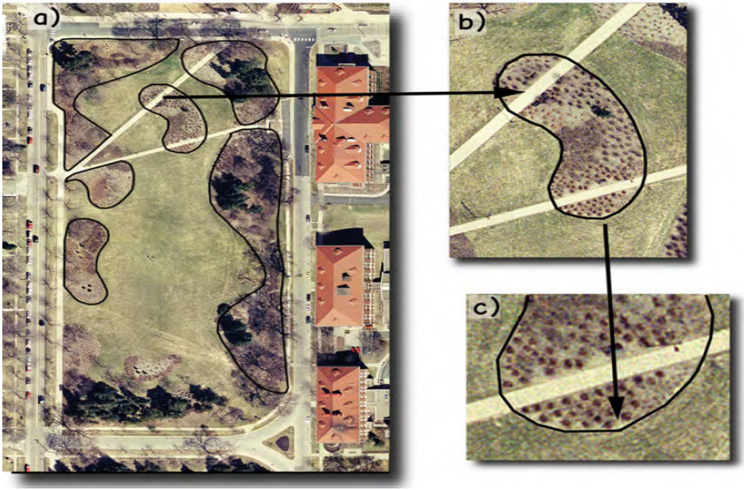


Figure 21: Examples of polygon inclusions (sidewalk inclusion in flower bed shown in a and b), and boundary generalization (c) in a vector data model (Bolstad (2016)).

Vector Topology

Vector data often contain *vector topology*, enforcing strict connectivity and recording adjacency and planarity. Early systems employed a spaghetti data model (Figure 22, **a)**), in which lines may not intersect when they should, and may overlap without connecting.

Topological models create an intersection and place a node at each line crossing, record connectivity and adjacency, and maintain information on the relationships between and among points, lines, and polygons in spatial data.

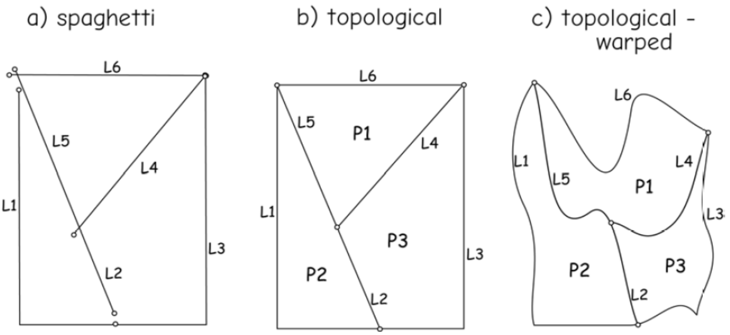


Figure 22: Spaghetti (a), topological (b), and topological warped (c) vector data. Figure b and c are topologically identical because they have the same connectivity and adjacency (Bolstad (2016)).

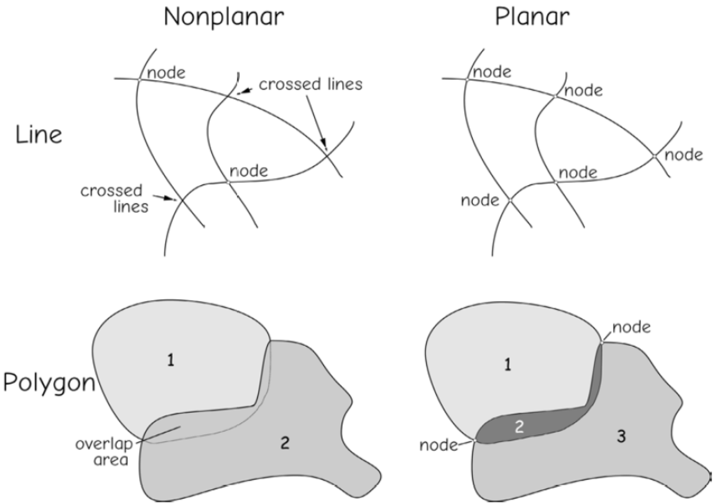
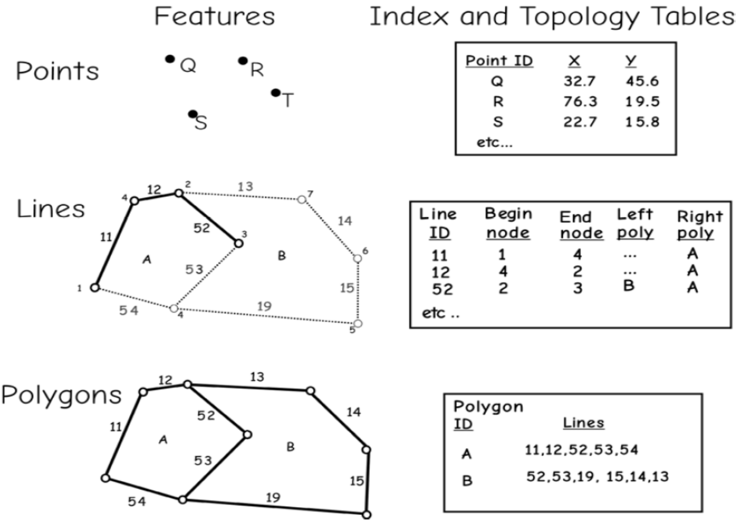


Figure 23: Nonplanar and planar topology in lines and polygons (Bolstad (2016)).

Other topological vector models: *planar topology*. It requires that all features occur on a two-dimensional surface. There can be no overlaps among lines or polygons in the same layer. When planar topology is enforced, lines may not cross over or under other lines. At each line crossing there must be an intersection.

Topological data models often have an advantage of smaller file sizes, largely because coordinate data are recorded once. For example, a nontopological approach often stores polygon boundaries twice. Lines 52 and 53 at the bottom of Figure 24 will be recorded for both polygon A and polygon B.



Point ID	X	Y
Q	32.7	45.6
R	76.3	19.5
S	22.7	15.8
etc...		

Line ID	Begin node	End node	Left poly	Right poly
11	1	4	...	A
12	4	2	...	A
52	2	3	B	A
etc ..				

Polygon ID	Lines
A	11,12,52,53,54
B	52,53,19, 15,14,13

Figure 24: An example of vector features and corresponding topology tables (Bolstad (2016)).

There are limitations and disadvantages to topological vector models:

- Computational costs in defining the topological structure of a vector data layer.
- The data must be very clean, in that all lines must begin and end with a node, all lines must connect correctly, and all polygons must be closed.

However, the limitations and the extra editing are far outweighed by the gains in efficiency and analytical capabilities provided by topological vector models.

There are limitations and disadvantages to topological vector models:

- Computational costs in defining the topological structure of a vector data layer.
- The data must be very clean, in that all lines must begin and end with a node, all lines must connect correctly, and all polygons must be closed.

However, the limitations and the extra editing are far outweighed by the gains in efficiency and analytical capabilities provided by topological vector models.

There are limitations and disadvantages to topological vector models:

- Computational costs in defining the topological structure of a vector data layer.
- The data must be very clean, in that all lines must begin and end with a node, all lines must connect correctly, and all polygons must be closed.

However, the limitations and the extra editing are far outweighed by the gains in efficiency and analytical capabilities provided by topological vector models.

There are limitations and disadvantages to topological vector models:

- Computational costs in defining the topological structure of a vector data layer.
- The data must be very clean, in that all lines must begin and end with a node, all lines must connect correctly, and all polygons must be closed.

However, the limitations and the extra editing are far outweighed by the gains in efficiency and analytical capabilities provided by topological vector models.

Vector Features, Tables, and Structures

Topological vector models are commonly used to define spatial features in a data layer. As we described earlier in this chapter, geographic features are associated with nonspatial attributes (Figure 25).

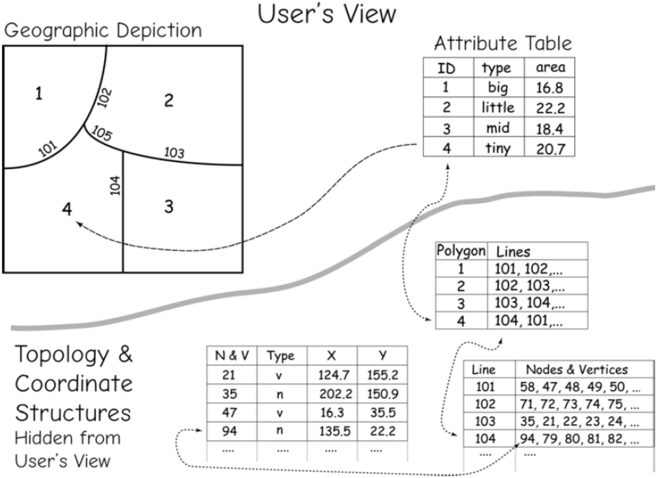


Figure 25: Features in a topological data layer typically have a one-to-one relationship with entries in an associated attribute table. The attribute table typically contains a column with a unique identifier, or ID, for each feature (Bolstad (2016)).

Raster Data Models

Raster data models define the world as a regular set of cells in a grid pattern (Figure 26). Typically these cells are square and evenly spaced in the x and y directions. The phenomena or entities of interest are represented by attribute values associated with each cell location.

Raster data models are the natural means to represent continuous spatial features or phenomena. Elevation, precipitation, slope, and pollutant concentration are examples of continuous spatial variables.

Raster data models define the world as a regular set of cells in a grid pattern (Figure 26). Typically these cells are square and evenly spaced in the x and y directions. The phenomena or entities of interest are represented by attribute values associated with each cell location.

Raster data models are the natural means to represent continuous spatial features or phenomena. Elevation, precipitation, slope, and pollutant concentration are examples of continuous spatial variables.

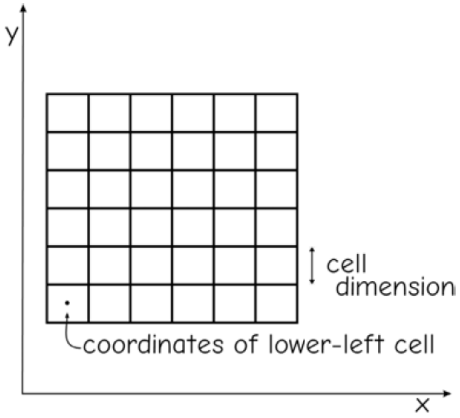


Figure 26: Important defining characteristics of a raster data model (Bolstad (2016)).

A cell location may be calculated from the cell size, known corner coordinates, and cell row and column number as:

$$\begin{aligned} \mathrm{N}_{\text{cell}} &= \mathrm{N}_{\text{lower-left}} + \text{row} * \text{cell size} && (6) \\ \mathrm{E}_{\text{cell}} &= \mathrm{E}_{\text{lower-left}} + \text{column} * \text{cell size} && (7) \end{aligned}$$

where N is the coordinate in the north direction (y), E is the coordinate in the east direction (x), and the row and column are counted starting with zero from the lower left cell.

Note that there is often a trade-off between spatial detail and data volume in raster data sets. The number of cells needed to cover a given area increases four times when the cell size is cut in half. Smaller cells provide greater spatial detail, but at the cost of larger data sets (Figure 27).

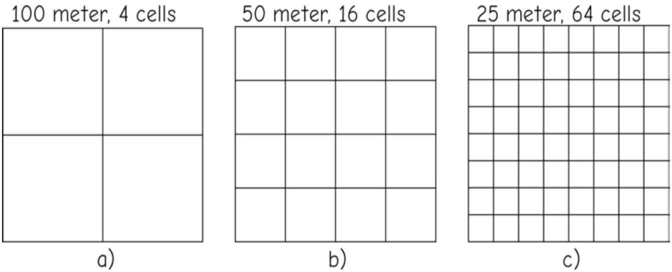


Figure 27: The number of cells in a raster data set depends on the cell size. For a given area, a linear decrease in cell size causes an exponential increase in cell number, e.g., halving the cell size causes a four fold increase in cell number (Bolstad (2016)).

A raster data model may also be used to represent discrete data, for example, to represent landcover in an area (Figure 28). Raster cells typically hold numeric or singleletter alphabetic characters. A coding scheme defines what land cover type the discrete values signify. Each code may be found at many raster cells and the raster cell values may be assigned and interpreted in at least seven different ways (Figure 29).

a	a	a	a	r	f	f	a	a	a	a	a
a	a	a	a	r	f	f	a	a	a	a	a
a	a	a	f	r	f	f	a	a	a	a	a
a	a	a	r	r	f	f	a	a	a	a	a
a	a	a	r	f	f	f	a	a	a	a	a
a	f	f	r	f	f	f	a	a	a	a	a
a	f	f	r	f	u	f	a	a	a	a	a
h	h	h	h	h	h	h	h	h	h	h	h
f	f	r	u	u	u	u	a	a	a	a	a
f	f	r	f	u	u	a	a	a	a	a	a
f	f	f	r	f	f	a	a	a	a	a	a
f	f	f	f	r	f	a	a	a	a	a	a

a = agriculture u = developed
f = forest r = river
h = highways

Figure 28: Discrete or categorical data may be represented by codes in a raster data layer (Bolstad (2016)).

Data Type	Description	Example
point ID	alpha-numeric ID of closest point	hospital
line ID	alpha-numeric ID of closest line	nearest road
contiguous region ID	alpha-numeric ID for dominant region	state
class code	alpha-numeric code for general class	vegetation type
table ID	numeric position in a table	row
physical analog	numeric value representing surface value	elevation
statistical value	numeric value from a statistical function	population density

Figure 29: Types of data represented by raster cell values (from L. Usery, pers. comm.) (Bolstad (2016)).

Point and line assignment to raster cells may be complicated when there are multiple features within a single cell.

- When light poles are represented in a raster data layer, cell value assignment is straightforward when there is only one light in a cell (Figure 30, near A).
- When there are multiple poles in a single cell there is some ambiguity, or generalization in the assignment (Figure 30, near B).
- When two or more roads meet, they will do so within a raster cell, and some set of attributes must be assigned (Figure 30, C).

For the two first points a common solution represents one feature from the group, and retains information on the attributes and characteristics of that feature. This entails some data loss.

Point and line assignment to raster cells may be complicated when there are multiple features within a single cell.

- When light poles are represented in a raster data layer, cell value assignment is straightforward when there is only one light in a cell (Figure 30, near A).
- When there are multiple poles in a single cell there is some ambiguity, or generalization in the assignment (Figure 30, near B).
- When two or more roads meet, they will do so within a raster cell, and some set of attributes must be assigned (Figure 30, C).

For the two first points a common solution represents one feature from the group, and retains information on the attributes and characteristics of that feature. This entails some data loss.

Point and line assignment to raster cells may be complicated when there are multiple features within a single cell.

- When light poles are represented in a raster data layer, cell value assignment is straightforward when there is only one light in a cell (Figure 30, near A).
- When there are multiple poles in a single cell there is some ambiguity, or generalization in the assignment (Figure 30, near B).
- When two or more roads meet, they will do so within a raster cell, and some set of attributes must be assigned (Figure 30, C).

For the two first points a common solution represents one feature from the group, and retains information on the attributes and characteristics of that feature. This entails some data loss.

Point and line assignment to raster cells may be complicated when there are multiple features within a single cell.

- When light poles are represented in a raster data layer, cell value assignment is straightforward when there is only one light in a cell (Figure 30, near A).
- When there are multiple poles in a single cell there is some ambiguity, or generalization in the assignment (Figure 30, near B).
- When two or more roads meet, they will do so within a raster cell, and some set of attributes must be assigned (Figure 30, C).

For the two first points a common solution represents one feature from the group, and retains information on the attributes and characteristics of that feature. This entails some data loss.

Point and line assignment to raster cells may be complicated when there are multiple features within a single cell.

- When light poles are represented in a raster data layer, cell value assignment is straightforward when there is only one light in a cell (Figure 30, near A).
- When there are multiple poles in a single cell there is some ambiguity, or generalization in the assignment (Figure 30, near B).
- When two or more roads meet, they will do so within a raster cell, and some set of attributes must be assigned (Figure 30, C).

For the two first points a common solution represents one feature from the group, and retains information on the attributes and characteristics of that feature. This entails some data loss.

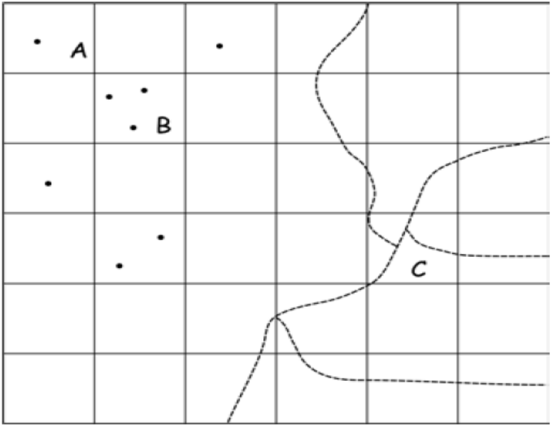


Figure 30: Raster cell assignment requires decisions when multiple objects occur in the same cell (Bolstad (2016)).

Raster Features and Attribute Tables

Raster layers can be associated attribute tables. This is most common when nominal data are represented, but may also be used with ordinal or interval/ratio data.

Features in the raster layer may be linked to rows in an attribute table, and these rows may describe the essential nonspatial characteristics of the features.

Raster layers can be associated attribute tables. This is most common when nominal data are represented, but may also be used with ordinal or interval/ratio data.

Features in the raster layer may be linked to rows in an attribute table, and these rows may describe the essential nonspatial characteristics of the features.

a) Raster, one-to-one

A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	B	B	B	B	B	B	B
A	A	A	C	C	B	B	B	B	B
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	E	E
C	C	C	C	C	D	D	E	E	E

attribute table
(cell 1 is upper-left corner)

cell-ID	IDorg	class	area
1	A	10	0.8
2	A	10	0.8
3	A	10	0.8
4	A	10	0.8
5	B	11	0.8
6	B	11	0.8
7	B	11	0.8
⋮	⋮	⋮	⋮
100	E	10	0.8

b) Raster, many-to-one

10	10	10	10	11	11	11	11	11	11
10	10	10	10	11	11	11	11	11	11
10	10	10	10	11	11	11	11	11	11
10	10	10	11	11	11	11	11	11	11
10	10	10	15	15	11	11	11	11	11
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	10	10
15	15	15	15	15	21	21	10	10	10

attribute table

class	area
10	18.4
11	24.0
15	21.6
21	13.6

Figure 31: Raster data models one-to-one relationship between cells and attributes (a). Raster data models many-to-one relationship between cells and table rows (b) (Bolstad (2016)).

Differences between raster data models and vector data models?

Characteristic
data structure
storage requirements
coordinate conversion
analysis
spatial precision
accessibility
display and output

Figure 32: Comparison between raster and vector data models (Bolstad (2016)).

Differences between raster data models and vector data models?

Characteristic	Raster	Vector
data structure	usually simple	usually complex
storage requirements	larger for most data sets without compression	smaller for most data sets
coordinate conversion	may be slow due to data volumes, and require resampling	simple
analysis	easy for continuous data, simple for many layer combinations	preferred for network analyses, many other spatial operations more complex
spatial precision	floor set by cell size	limited only by positional measurements
accessibility	easy to modify or program, due to simple data structure	often complex
display and output	good for images, but discrete features may show “stairstep” edges	maplike, with continuous curves, poor for images

Figure 33: Comparison between raster and vector data models (Bolstad (2016)).

Conversion Between Raster and Vector Models

UNIVERSITÄT BAYREUTH

Vector to raster:

- Vector to raster conversion involves assigning a cell value for each position occupied by vector features.
- The vector point features commonly have not dimension.
- Points in a raster data set must be represented by a value in a raster cell.

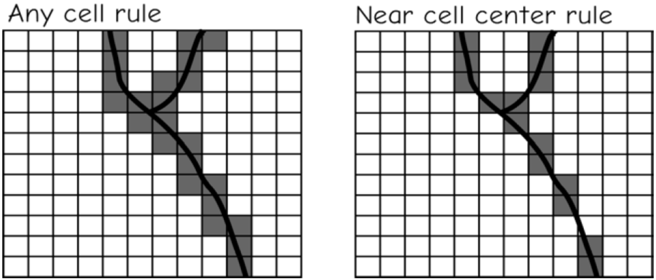


Figure 34: Vector line to raster (1 is left and 2 is right) (Bolstad (2016)).

The output from vector-to-raster conversion depends on the algorithm used, even though you use the same input.

The output often depends in subtle ways on the spatial operation.

The output from vector-to-raster conversion depends on the algorithm used, even though you use the same input.

The output often depends in subtle ways on the spatial operation.

Raster to vector data:

- Point features: each vector point feature is usually assigned the coordinate of the corresponding cell center (Figure 35, a)).
- Linear features: conversion to vector lines typically involves identifying the continuous connected set of grid cells that form the line (Figure 35, b)).

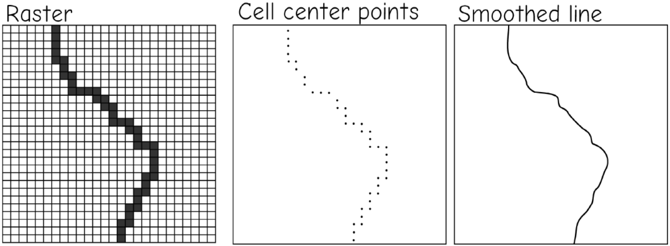


Figure 35: Raster to vector conversion (points and line respectively) (Bolstad (2016)).

Other data models

Triangulated Irregular Networks

A triangulated irregular network (TIN) is a data model commonly used to represent terrain heights. Typically the x, y, and z locations for measured points are entered into the TIN data model. The TIN forms a connected network of triangles. *Delaunay triangles* are created such that the lines from one triangle do not cross the lines of another. (Figure 36).

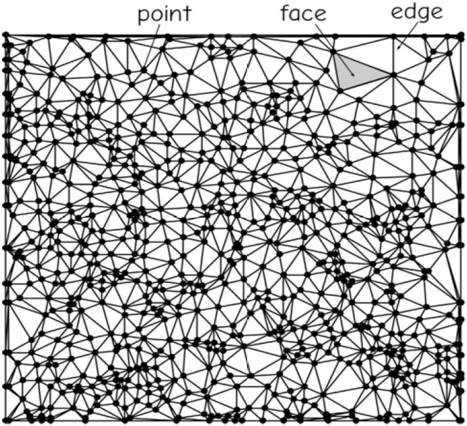


Figure 36: A TIN data model defines a set of adjacent triangles over a sample space. Sample points, facets, and edges are components of TIN data models. (Bolstad (2016)).

Three-Dimensional Data Models

GIS in built environments are increasingly integrating three-dimensional (3D), this is in part to support analysis, and in part to generate visualizations from at or near ground-level, for example, building appearance from a nearby road.

While vector 3D models are more common, no one model form or standard has been widely adopted, with representations generally software-specific.

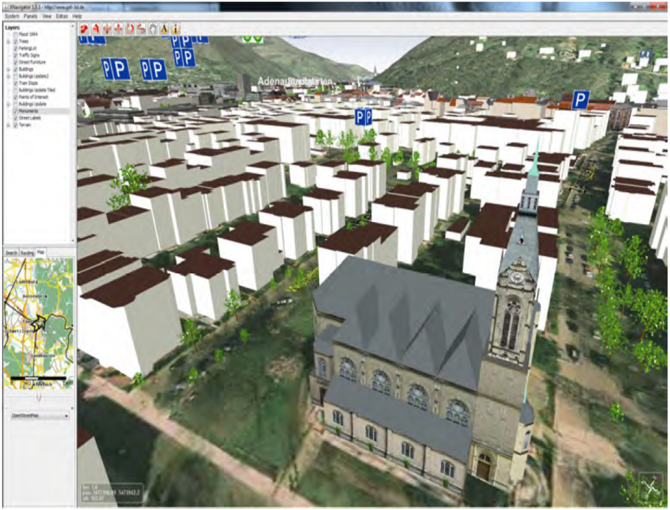


Figure 37: An example of 3D spatial data displayed for a region in Germany (Bolstad (2016)).

Multiple Models

Because of this widespread importance, digital elevation data are commonly represented in a number of data models.

Raster grids, TINs, and vector contours are the most common data structures used to organize and store digital elevation data. Raster and TIN data are often called digital elevation models (DEMs) or digital terrain models (DTMs) and are commonly used in terrain analysis.

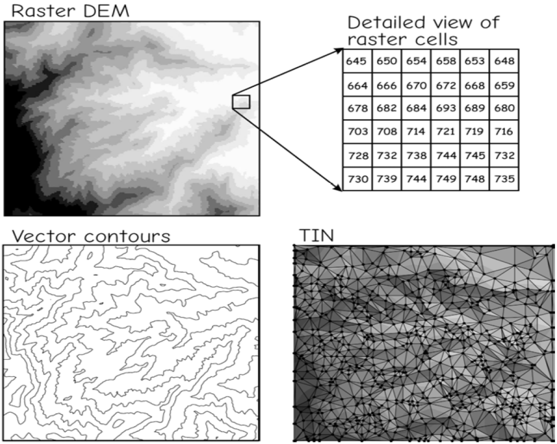


Figure 38: Data may often be represented in several data models. Digital elevation data are commonly represented in raster (DEM), vector (contours), and TIN data models (Bolstad (2016)).

Thank You

References I

Bolstad, P. (2016). *GIS fundamentals: A first text on geographic information systems*. Eider (PressMinnesota).