KIG-C1010 Introduction to geoinformatics

Lecture 3: introduction to spatial data modelling

Aalto-yliopisto
Insinööritieteiden korkeakoulu

Jussi Nikander

18.1.2023

Based on materials developed by Paula Ahonen-Rainio

Topics for today

Basics of spatial modeling

Object and field models

Vector and raster data models

Spatial data in GIS software and databases

Data modeling and data relations: metric and topological relationships

Examples of possible exam questions related to the lecture

Paikkatieto ei voi koskaan esittää täydellisesti todellisen maailma kohteita tai ilmiöitä. Miksi ei, vai voiko?

Geographic data can never fully represent objects or phenomena of real world. Why not, or can it?

Todellisuutta voidaan mallintaa diskreetteinä objekteina (object-based modelling) tai spatiaalisesti jatkuvina ilmiöinä (field-based modelling). Esittele nämä mallinnustavat ja pohdi niiden soveltuvuutta kaupunkiympäristön, maanpinnan korkeuden ja väestötietojen mallinnukseen.

We can model environment, at the conceptual level, as discrete objects (object-based modelling) or spatially continuous phenomena (field-based modelling). Explain these modelling approaches and discuss how they can be used for modelling urban environment, elevation of terrain, and demographic (population) information.

Millä eri tavoilla spatiaalisesti jatkuva ilmiö voidaan esittää vektori-ja rasteritietona?

In which different ways can a spatially continuous phenomenon be represented in vector and raster data?

Spatial data modeling – why, what, and how

Why: what is the purpose of the model?

- **Why is the model created?**
- **What phenomena are included?**
- **What will the model be used for?**
  - Specific analysis
    - E.g. “what would be the best possible location for a new store?”
  - General service
    - E.g. reittiopas
  - Data set
    - E.g. Finnish elevation model

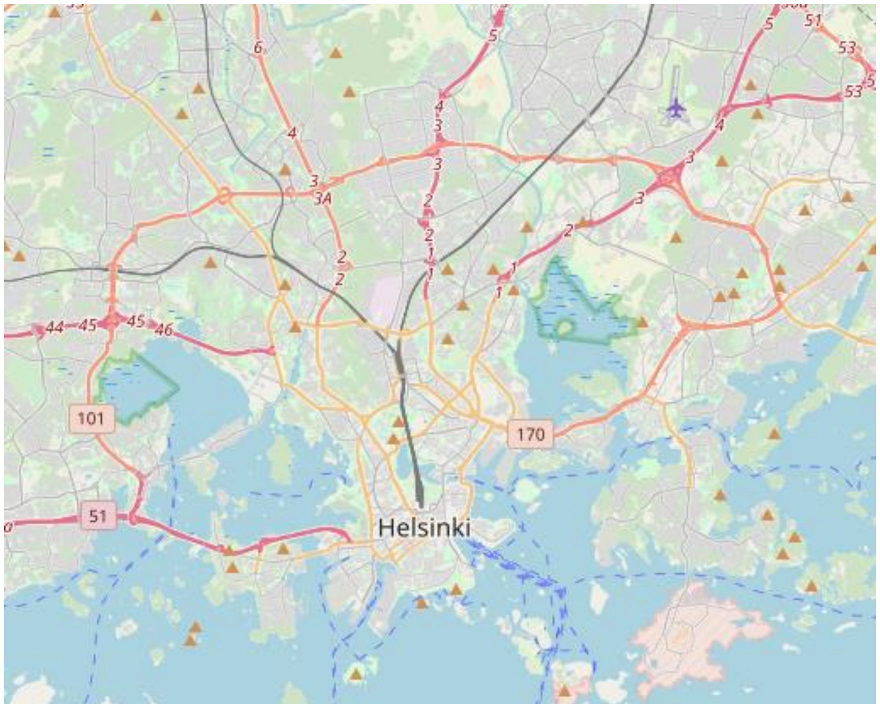
- **Who will be using the model?**
  - General public?
  - Decision makers?
  - Application area experts?
  - GIS professionals?

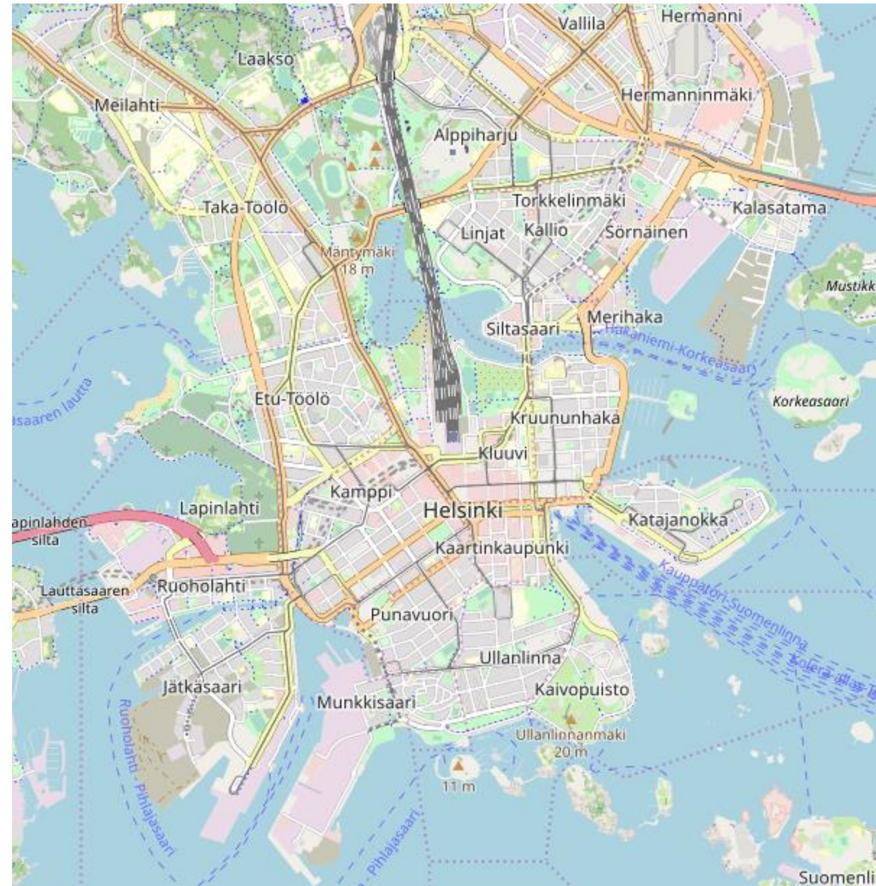
What: elements included in the model

- **The real world is infinitely complex – a model is always limited**
  - A model is a simplified representation of the real world
- **Elements included must be identifiable, relevant, and describable**

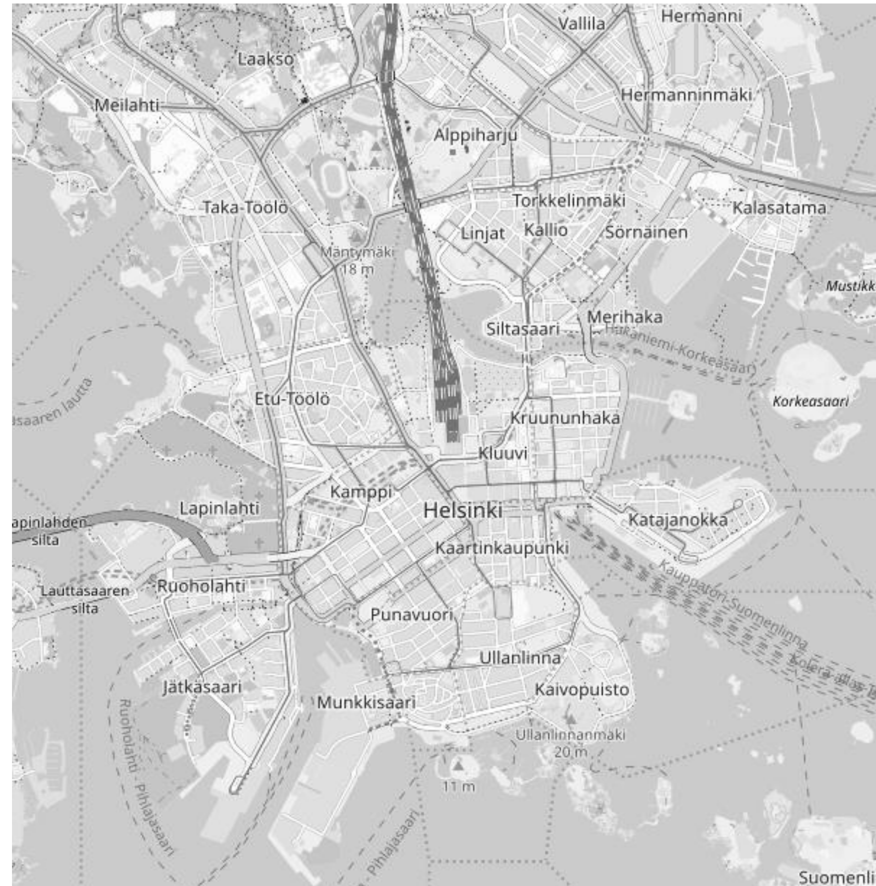
- **Identifiable: there must be means to distinguish between elements of the model**
- **Relevant: the elements must be useful for the intended purpose**
- **Describable: the elements must have characteristics that can be represented formally**

Classroom exercise: where is the Helsinki City centre?



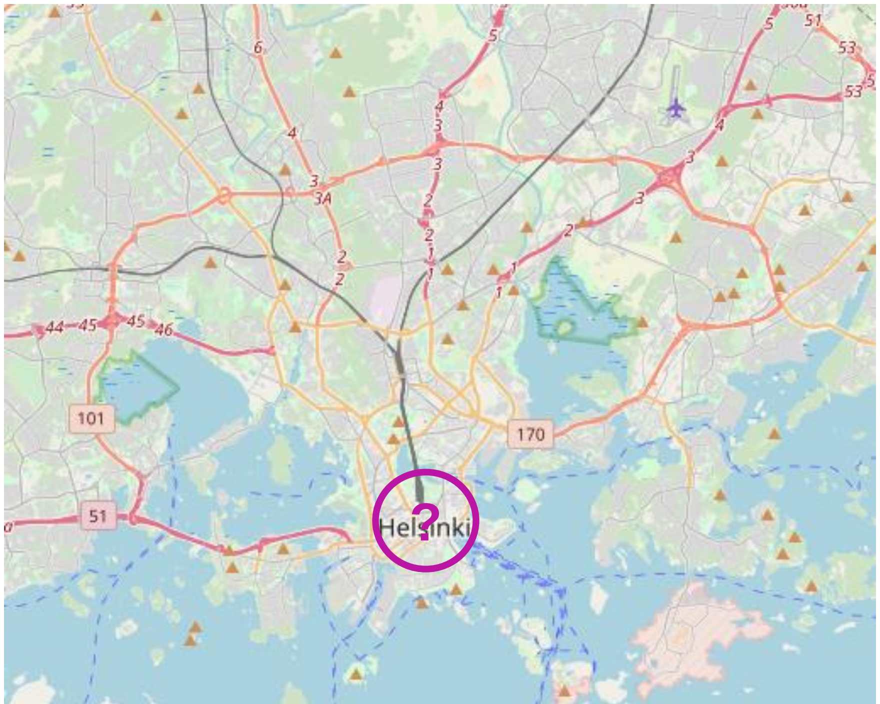


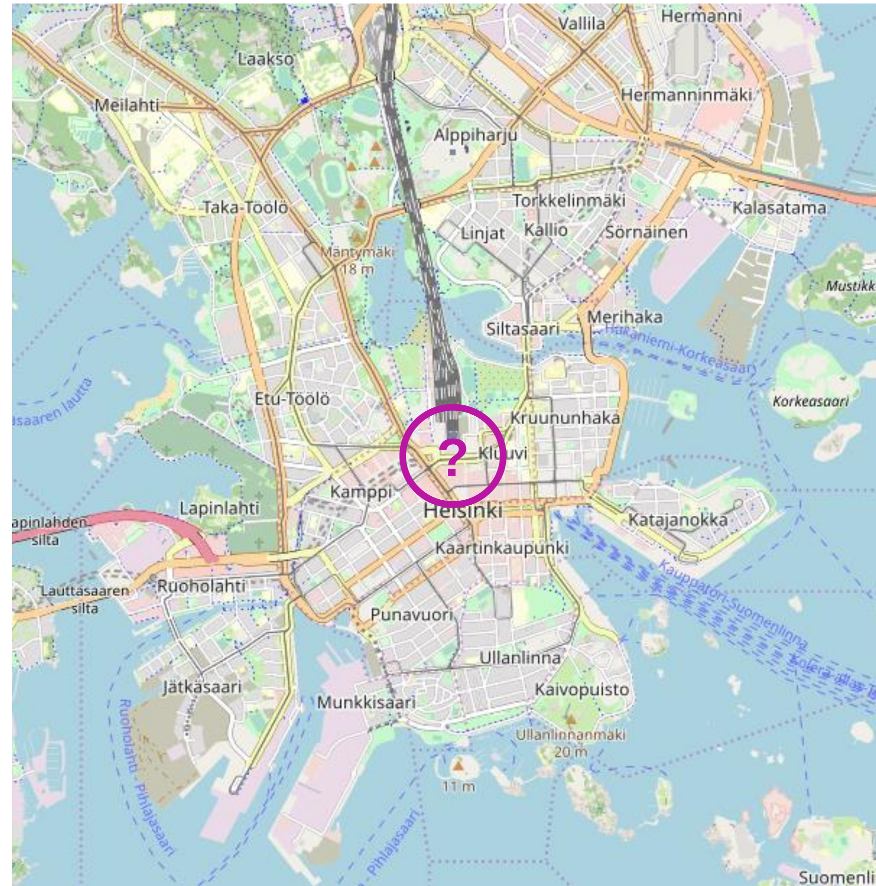
Classroom exercise: where is the Helsinki City centre?



Aalto-yliopisto
Insinööritieteiden
korkeakoulu

Classroom exercise: where is the Helsinki City centre?





Aalto-yliopisto
Insinööritieteiden
korkeakoulu

How: spatial data modeling approaches

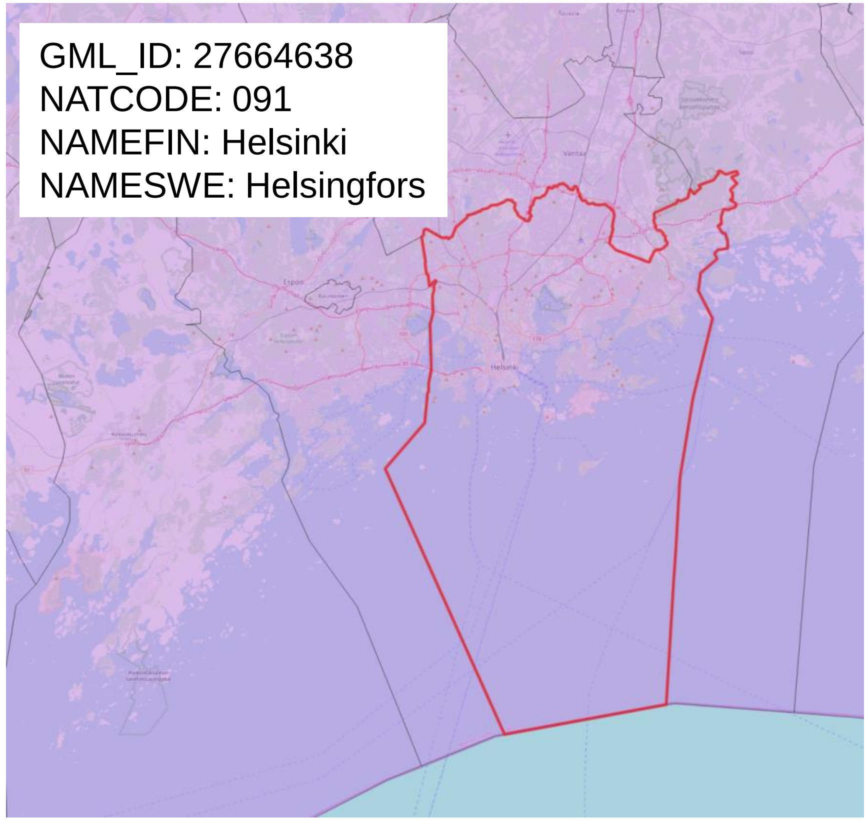
- **There are two main approaches how to model spatial data**
- **Object model represent the reality through distinct entities called objects**
- **Field model represents the reality as continuous phenomena**

Wikimedia

yle.fi (Seppo Savolainen)

The object model

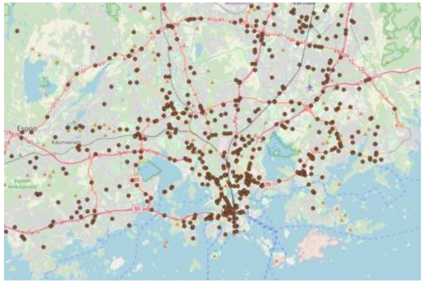
- **Distinct objects in the object model need to be identifiable, relevant, and describable**
- **In a spatial model, all objects have a location, spatial extent, and attribute information**
    - Attributes describe non-spatial aspects of the object



Aalto-yliopisto
Insinööritieteiden korkeakoulu

Discrete objects and geometry types

- **In most spatial models, three spatially distinct types of discrete objects are used**
- **Point has one set of coordinate values and represents single location**
- **Polyline consists of two or more points that are linked together**
- **Polygon consist of an area enclosed by a closed polyline**



Point data (traffic accidents 2016)

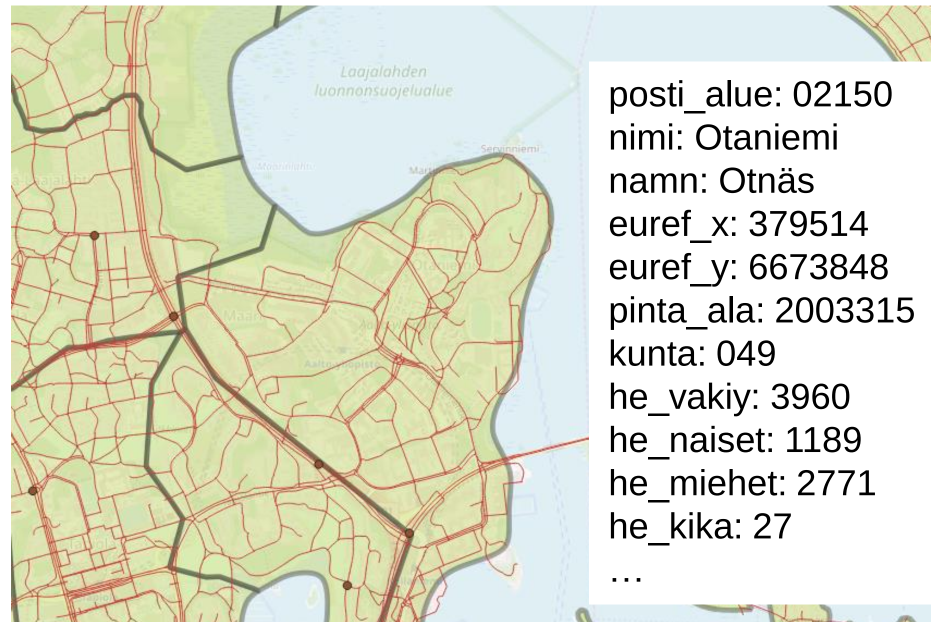
Polyline data (road network)

Polygon data (municipal borders)

Aalto-yliopisto
Insinööritieteiden
korkeakoulu

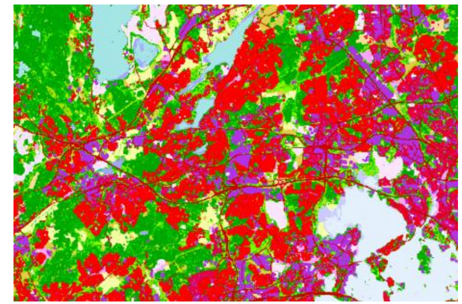
Objects and the vector data model

- **A data model consisting of point, polyline, or polygon data is called a vector model**
- **Each element in a vector data set consists of**
  - Geometry
  - Attributes (incl. unique ID)
  - Typically, only one geometry type per data set
- **Relations between elements may also be included**
  - E.g. topology (covered later)

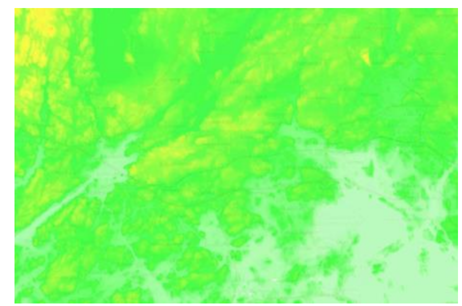


The field model

- **The phenomena in field model need to be continuous and variable with location**
  - Every location has a value (which can be null)
  - A method is required to establish this value in each location
  - Phenomenon cannot be naturally split into homogenous areas



Land cover classes

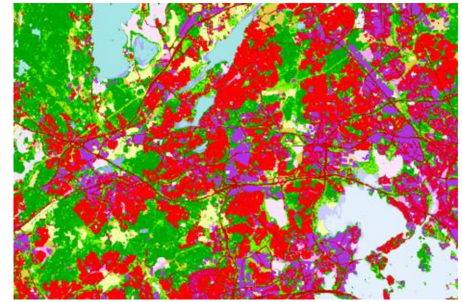


Elevation

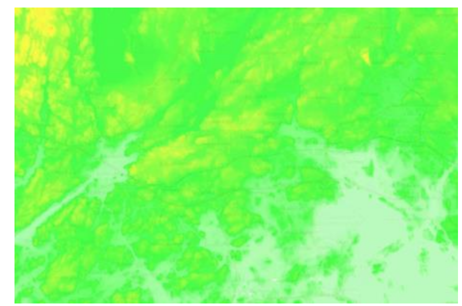
Air temperature in C

Field model values

- **Continuous and categorized values**
  - Continuous (interval or ratio) scale
    - Elevation, temperature, population density
    - Arithmetic can be done on the values
  - Categorized (nominal or ordinal) scale
    - Land cover class, soil type, etc.



Land cover classes (nominal scale)



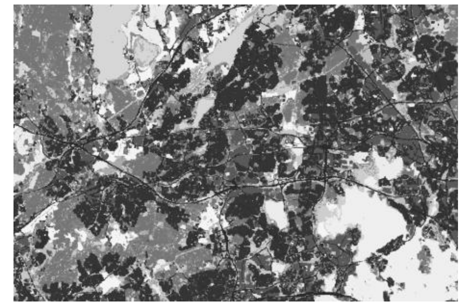
Elevation (ratio scale)

Air temperature in C (interval scale)

Aalto-yliopisto
Insinööritieteiden korkeakoulu

Field model values

- **Continuous and categorized values**
  - Continuous (interval or ratio) scale
    - Elevation, temperature, population density
    - Arithmetic can be done on the values
  - Categorized (nominal or ordinal) scale
    - Land cover class, soil type, etc.



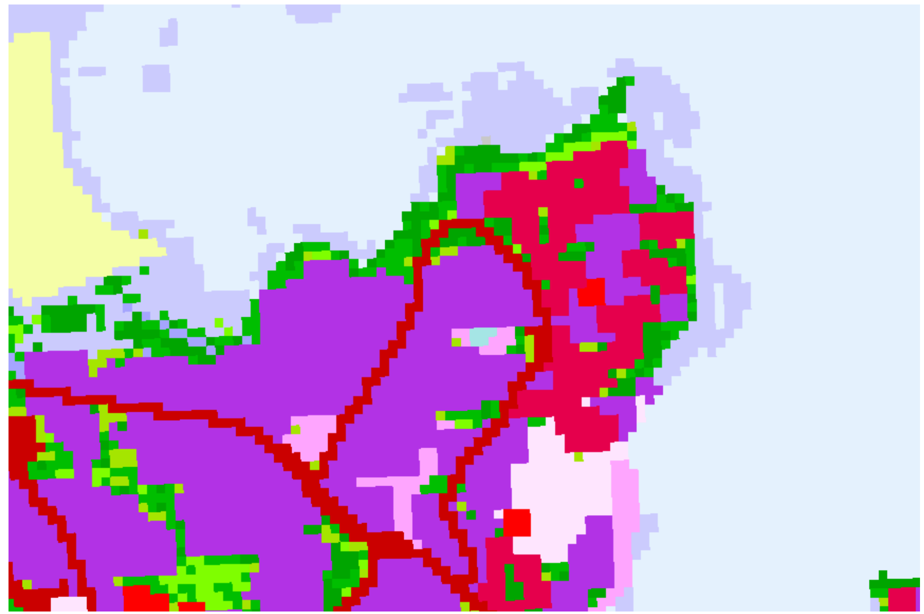
Land cover classes (nominal scale)

Elevation (ratio scale)

Air temperature in C (interval scale)

Fields and the raster data model

- **Field phenomena are often represented using the raster data model**
- **Raster (grid) is a regular tessellation composed of identical cells (pixels)**
  - Model geometry is independent of variations in the phenomena
- **Pixel size is the model resolution**
- **Colors represent different values**

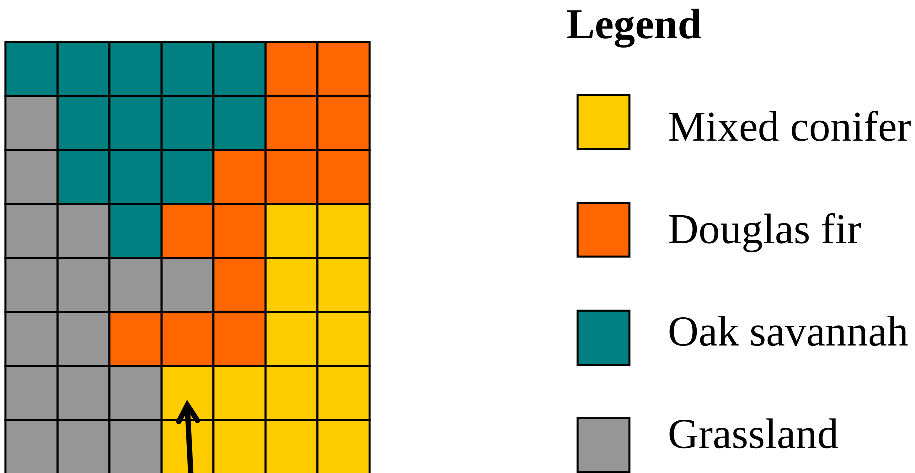


Aalto-yliopisto
Insinööritieteiden korkeakoulu

Fields and the raster data model

- **Field phenomena are often represented using the raster data model**
- **Raster (grid) is a regular tessellation composed of identical cells (pixels)**
  - Model geometry is independent of variations in the phenomena
- **Pixel size is the model resolution**
- **Colors represent different values**

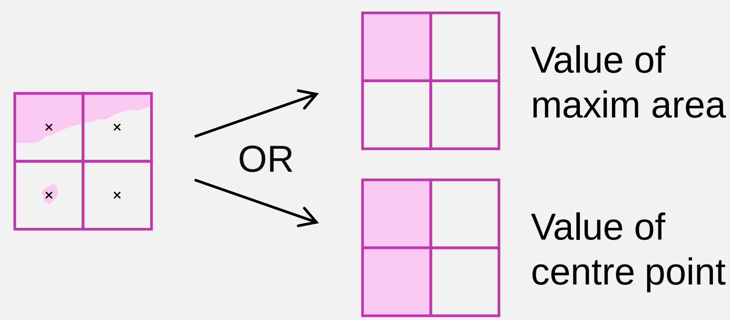
Example: Raster geometry



Longley et al. 2015

Notice:
Colour visualizes the values to users;
the actual data are either codes indicating the classes (as in this case) or metric values.

The value of a cell (pixel) can be the value of the variable in the centre point or the value dominating in the cell area.



Fields and the vector data model

- **Field data is not always represented using the raster model**
  - And object data is not always represented using vector model
- **Triangular Irregular Network (TIN)**
  - A model for elevation
  - Value known at specific points
- Points linked to form triangles
- Values elsewhere interpolated

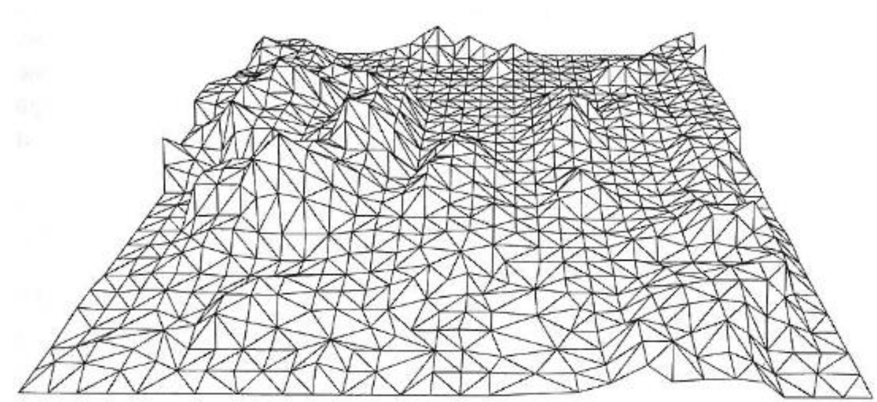


Image source: de Berg et al: Computational Geometry

Aalto-yliopisto
Insinööritieteiden korkeakoulu

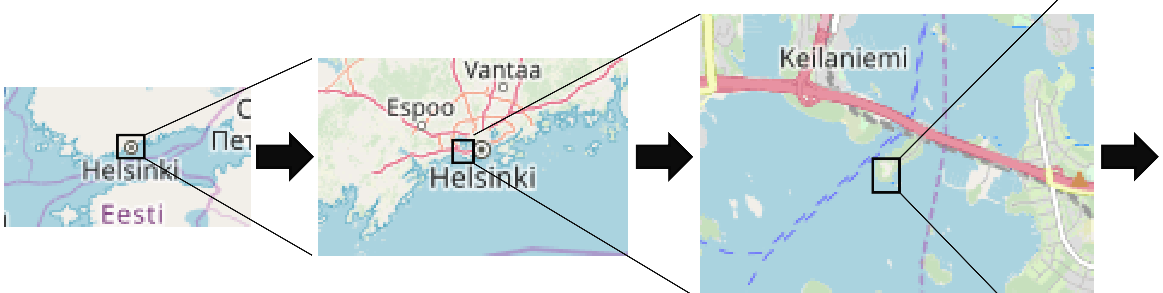
Potential complications in data presentation

Real world entities may…

- **be, and often are, 3 dimensional**
  - => 3D spatial models
- **change and move over time**
  - => Dynamic spatial models
- **Have comples geometries, e.g. a transport network**
  - => E.g. network models
- **be fuzzy and/or have indeterminate boundaries**
  - => Fuzzy modeling
- **have fractal characteristics**
  - => Space-filling models
- **Representation of real world entities in a model may be strongly scale-dependent**

Scale dependency example: the coastline fractal

- **The length of a coastline depends on the level of detail used to model it**
  - Higher resolution leads to longer coastline
  - Real-world coastline does not have a well-defined length (coastline paradox)
  - A model has a maximum resolution, which defines the coastline length in the model



Aalto-yliopisto
Insinööritieteiden
korkeakoulu

Image source:OpenStreetMap

From 2d to 3d: some changes in modeling

- **In 2d each polygon is typically considered separately**
  - Collections are supported, but many applications do not need them
- **In 3d the distinction between identifiable parts of the model that can be distinguished from one other are typically more complex**
  - Most elements of the model are aggregates (polygon meshes, etc.)
- **3d data is unfortunately out of scope for this course**

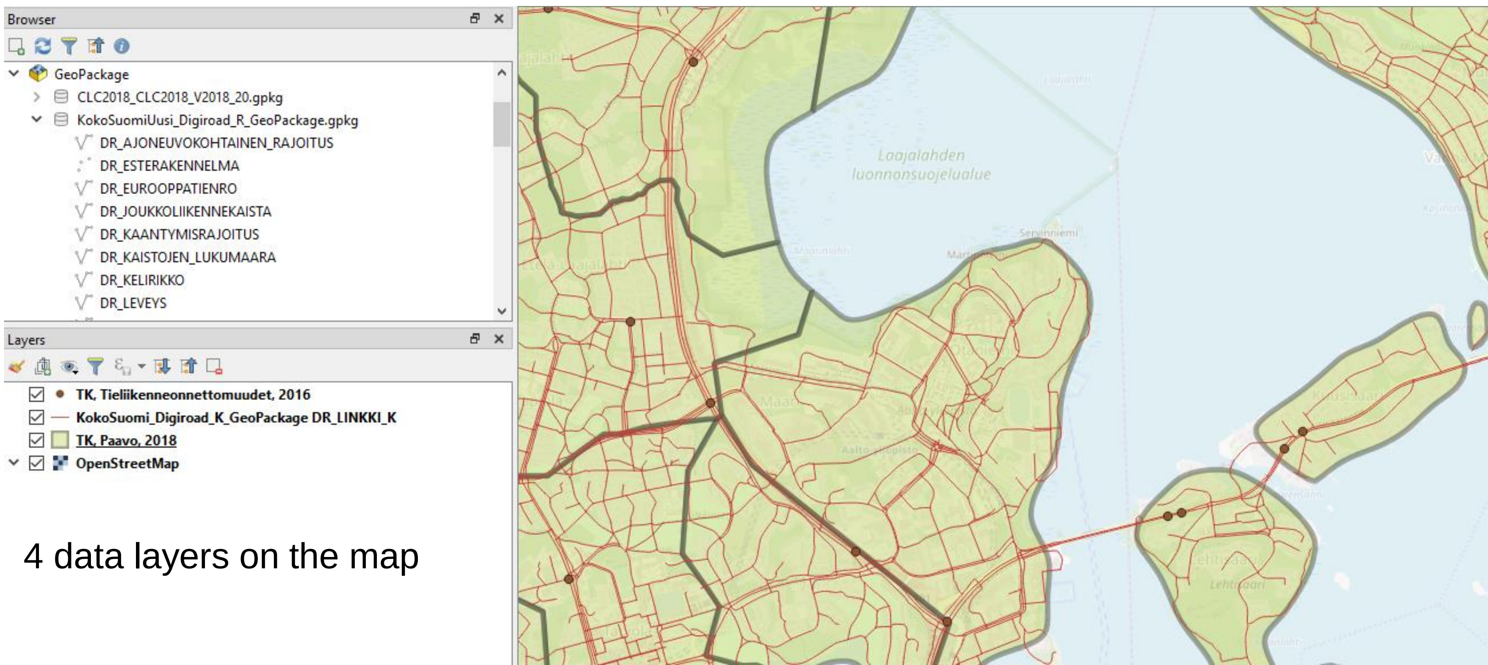
Spatial data modelling: handling and storage

Aalto-yliopisto
Insinööritieteiden korkeakoulu

Spatial data and GIS Software

- **Spatial data analysis is done on GIS software**
  - Traditionally desktop software
  - Increasingly also on the web

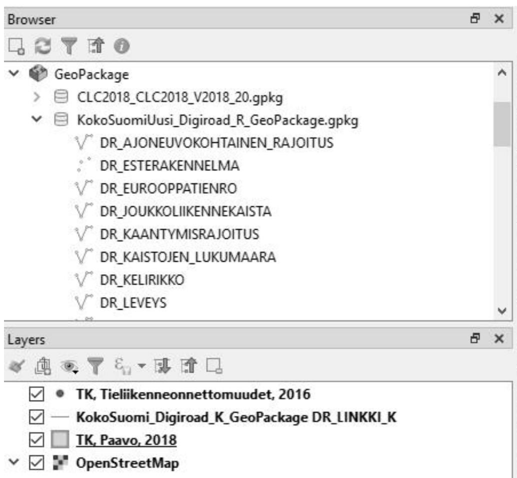
- **GIS data presentation in software is typically based on the concept of layers**
  - Cf. analog maps on paper or transparencies



4 data layers on the map

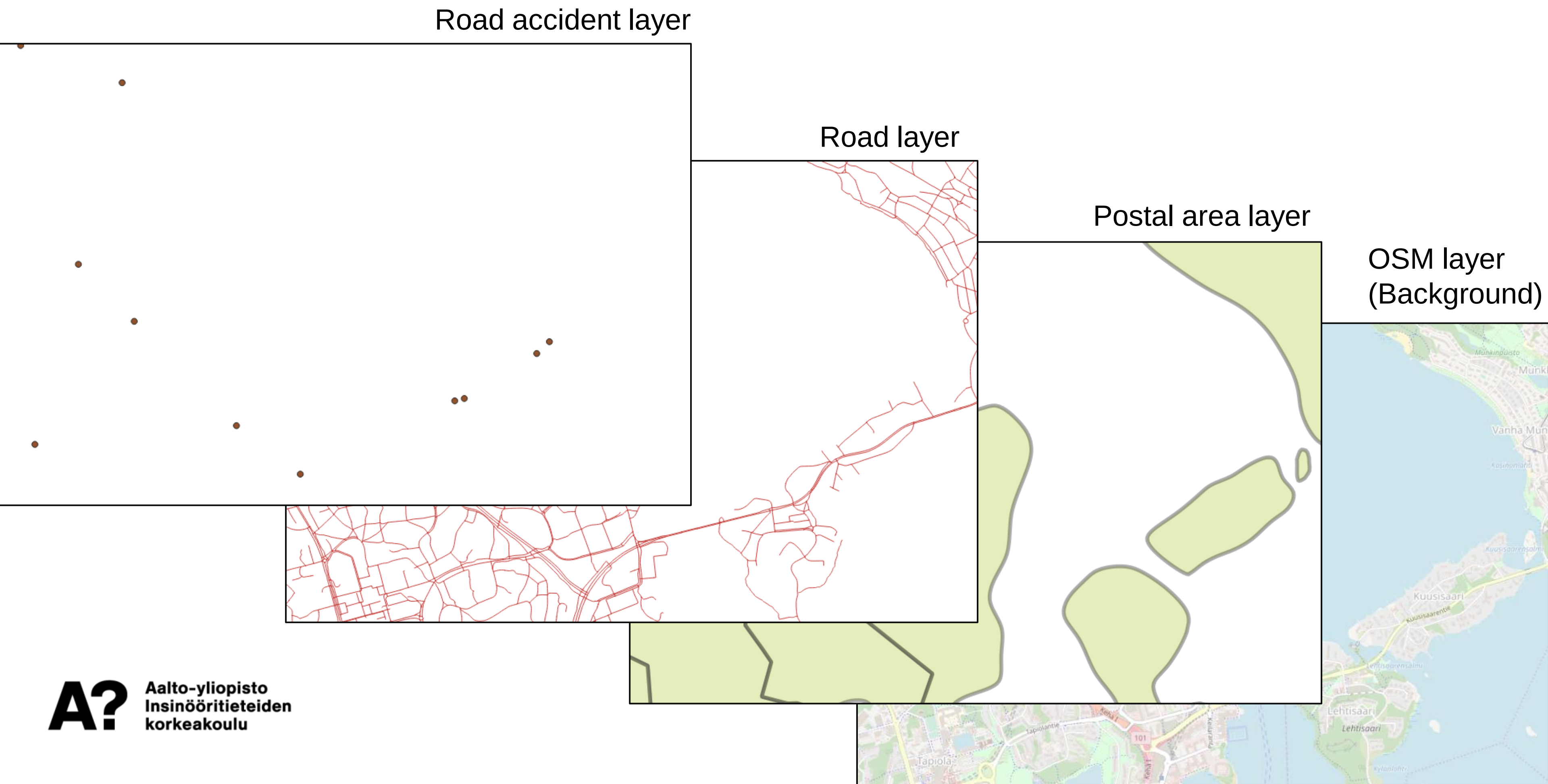
Spatial data and GIS Software

- **Spatial data analysis is done on GIS software**
  - Traditionally desktop software
  - Increasingly also on the web
- **GIS data presentation in software is typically based on the concept of layers**
  - Cf. analog maps on paper or transparencies

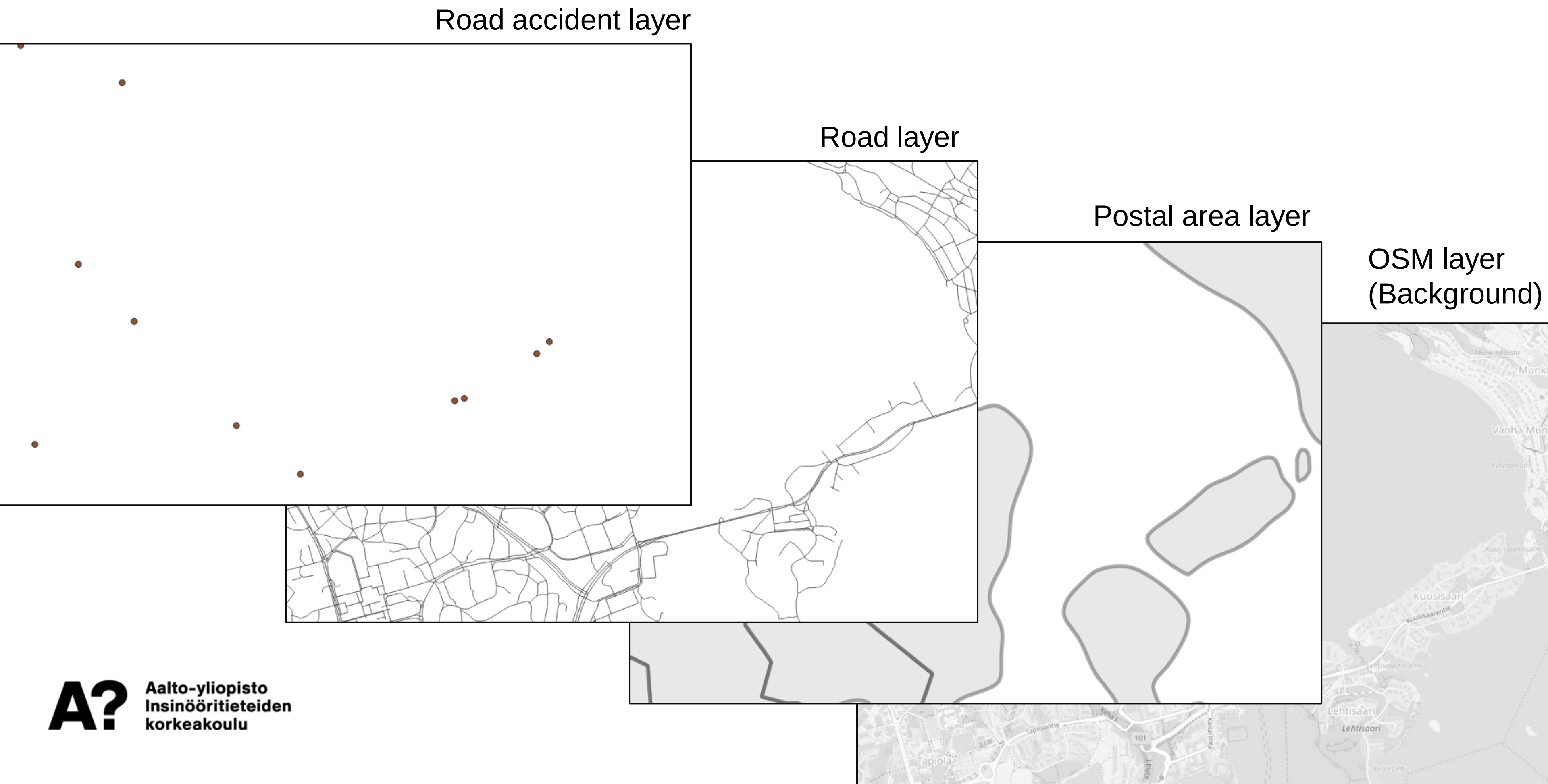


4 data layers on the map

Data layers example

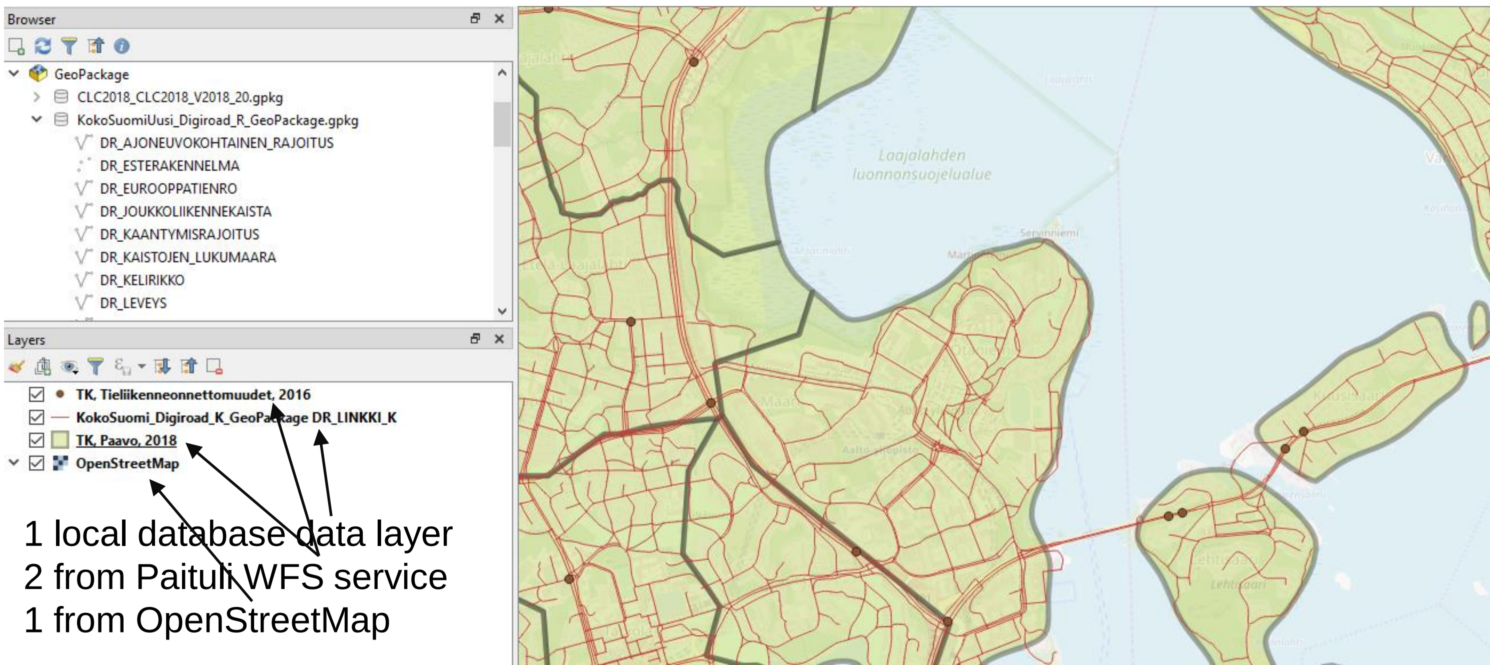


Data layers example



Spatial data and GIS Software

- **Spatial data is stored in databases**
  - Local databases
  - On-line databases
- **Database typically contain data set(s) related to specific purpose**

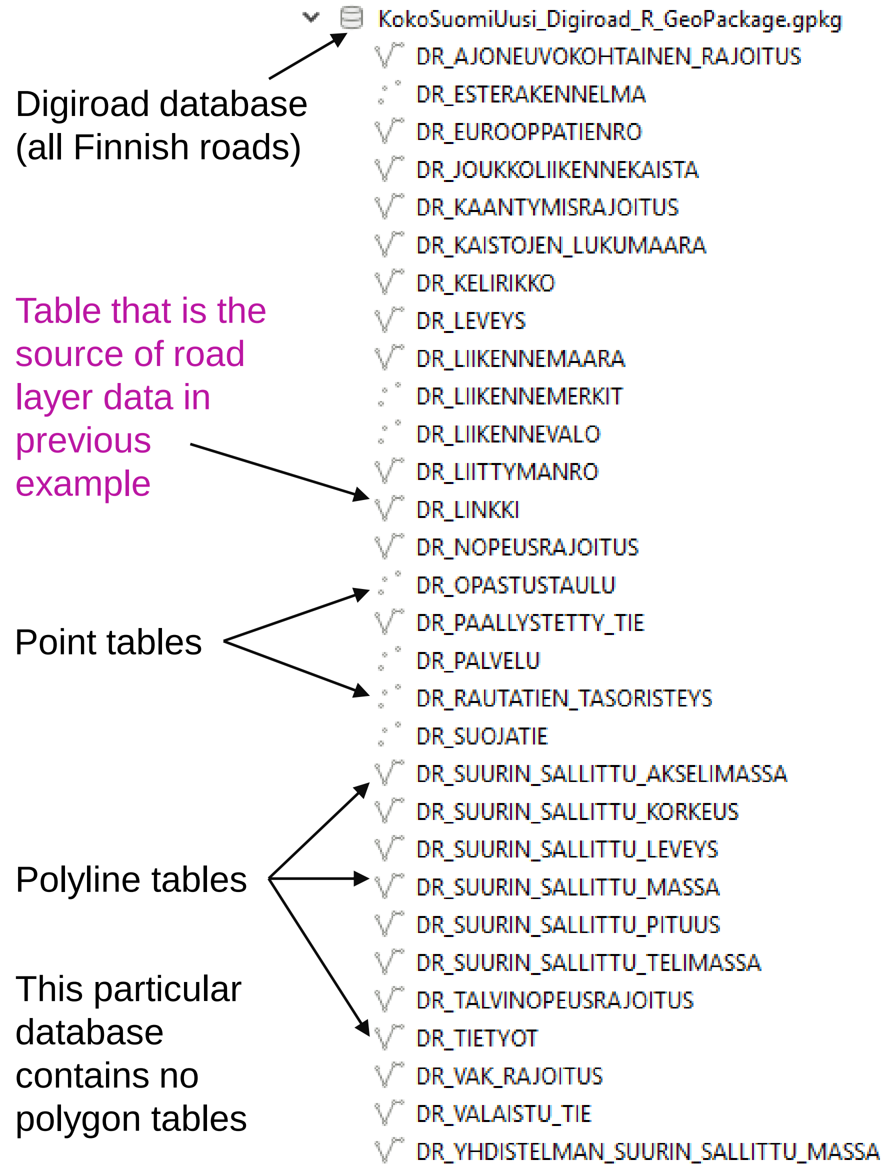


1 local database data layer
2 from Paituli WFS service
1 from OpenStreetMap

A?

Spatial data and databases

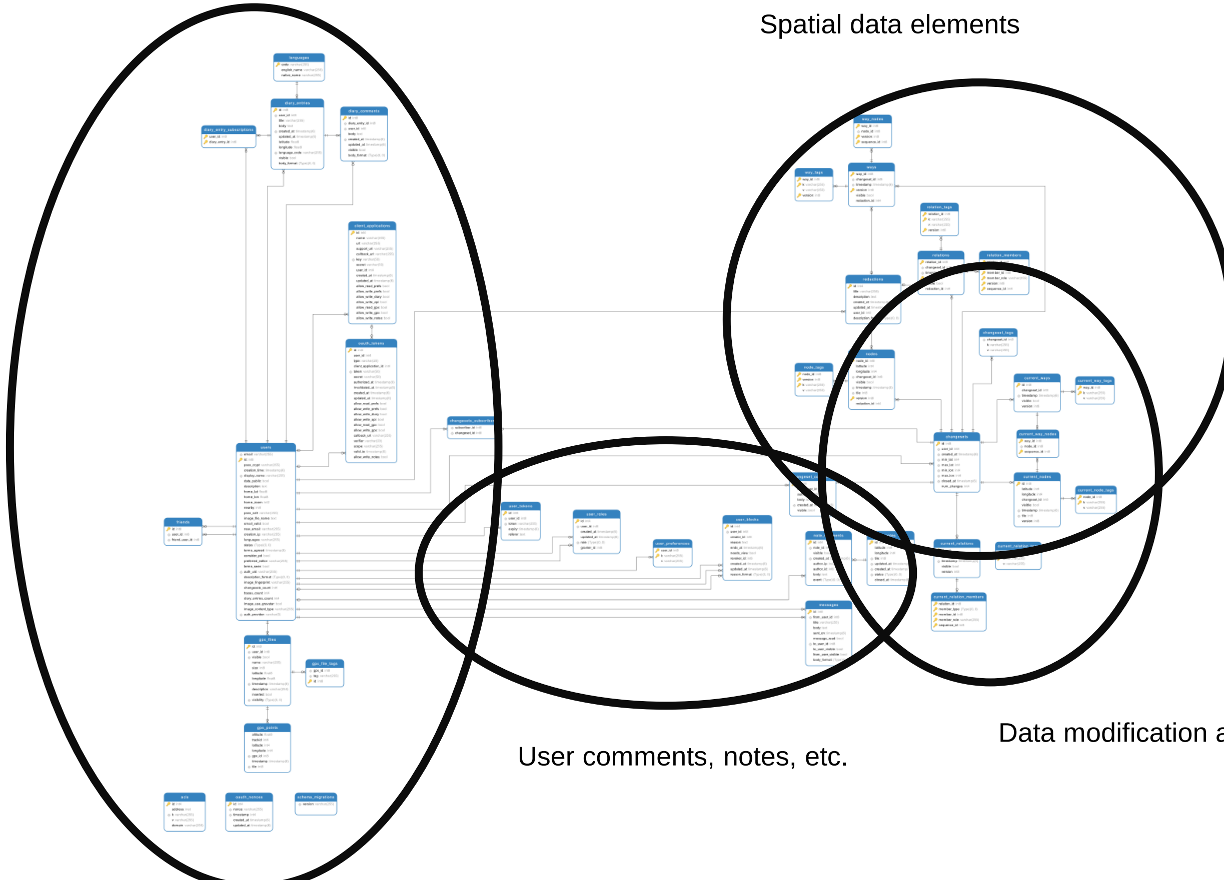
- **Spatial data is stored in databases**
  - Many providers and formats
  - File formats sometimes used for data transfer
    - Shapefile, geotiff
- **Spatial databases often use relational (SQL) format**
  - Data divided into tables
  - Each table has a geometry type
  - All elements in a table have the same attributes
  - Tables joined by relations



Example: OSM Main Database

Spatial data elements

OSM user who contributes data to OSM DB



User comments, notes, etc.

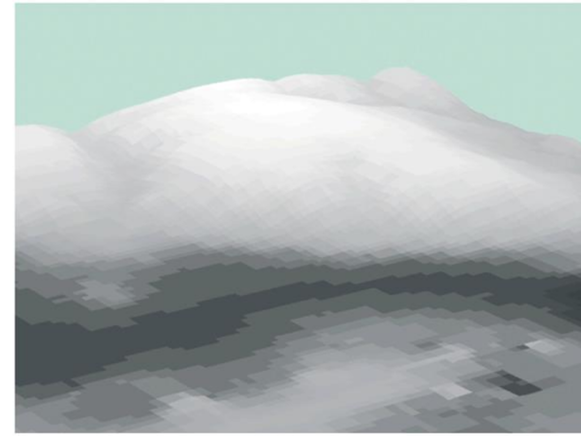
Data modification and history

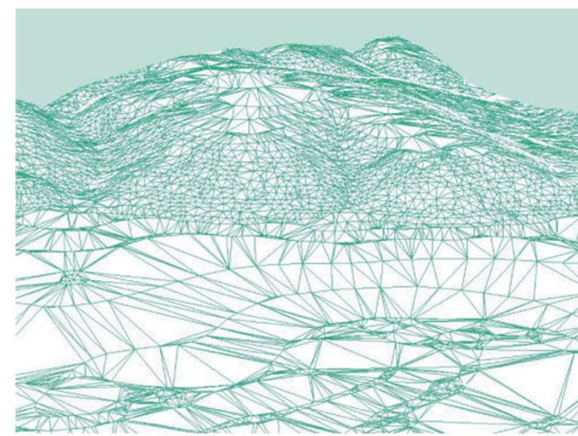
Aalto-yliopisto
Insinööritieteiden
korkeakoulu

Source: OpenStreetMap documentation

Different data models for the same phenomenon

- **Elevation can be modeled using DEM (Digital Elevation Model) or a TIN (Triangular Irregular Network)**
- **DEM is a raster model; data is in regular tessellation**
- **TIN is a vector model; data point density varies**





Heywood et al. 2006

Classroom exercise: DEM vs TIN

How does the model influence the potential accuracy of representation?

- **DEM with 25m * 25m raster cells compared to 1m * 1m cells?**

- **TIN with average distance between measured points being 0,5m, 5m, or 50m?**

Classroom exercise: DEM vs TIN

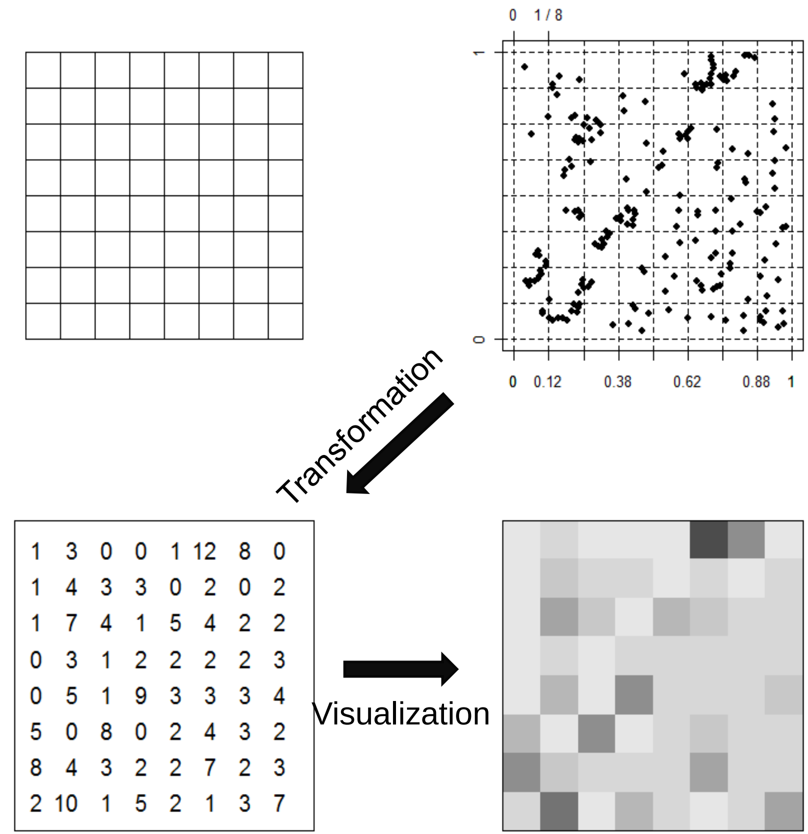
How does the model influence the potential accuracy of representation?

- **DEM with 25m * 25m raster cells compared to 1m * 1m cells?**
  - The smaller the cell, the better the model is able to take into account small landforms
  - Is the value the same everywhere in the cell?
  - If not, how can we estimate values in different locations?

- **TIN with average distance between measured points being 0,5m, 5m, or 50m?**
  - Closer the points, the smaller the individual triangles in the TIN and the better the model is able to take into account small landforms
  - If the point density is high in locations where elevation changes rapidly, the model can be good regardless of average point density

Spatial data transformations

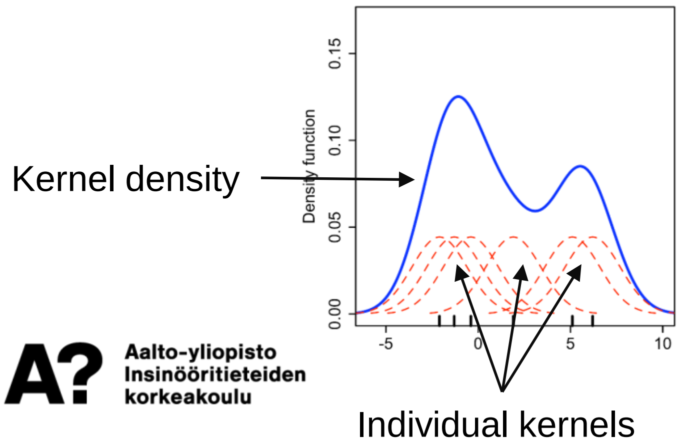
- **Often spatial data is transformed in various ways**
  - Data model transformations: vector-to-raster, raster-to-vector
  - Transformations that change semantics (meaning) of data
- **For example, point objects can be transformed into a field model by counting the sum of points in each cell of the given grid**
  - Point data becomes a data density raster (or, more generally, a surface)



1	3	0	0	1	12	8	0
1	4	3	3	0	2	0	2
1	7	4	1	5	4	2	2
0	3	1	2	2	2	2	3
0	5	1	9	3	3	3	4
5	0	8	0	2	4	3	2
8	4	3	2	2	7	2	3
2	10	1	5	2	1	3	7

Spatial data transformations

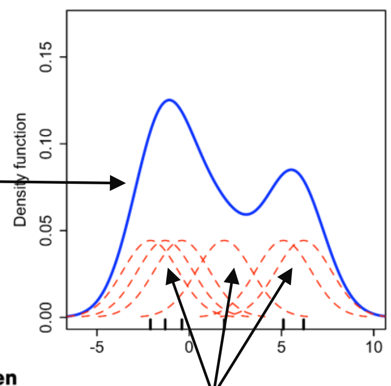
- Another way to turn point data into a density surface is to use kernel function
- A kernel is calculated around each point
- Kernels are summed at each location (in practice: pixel) to get the Kernel density

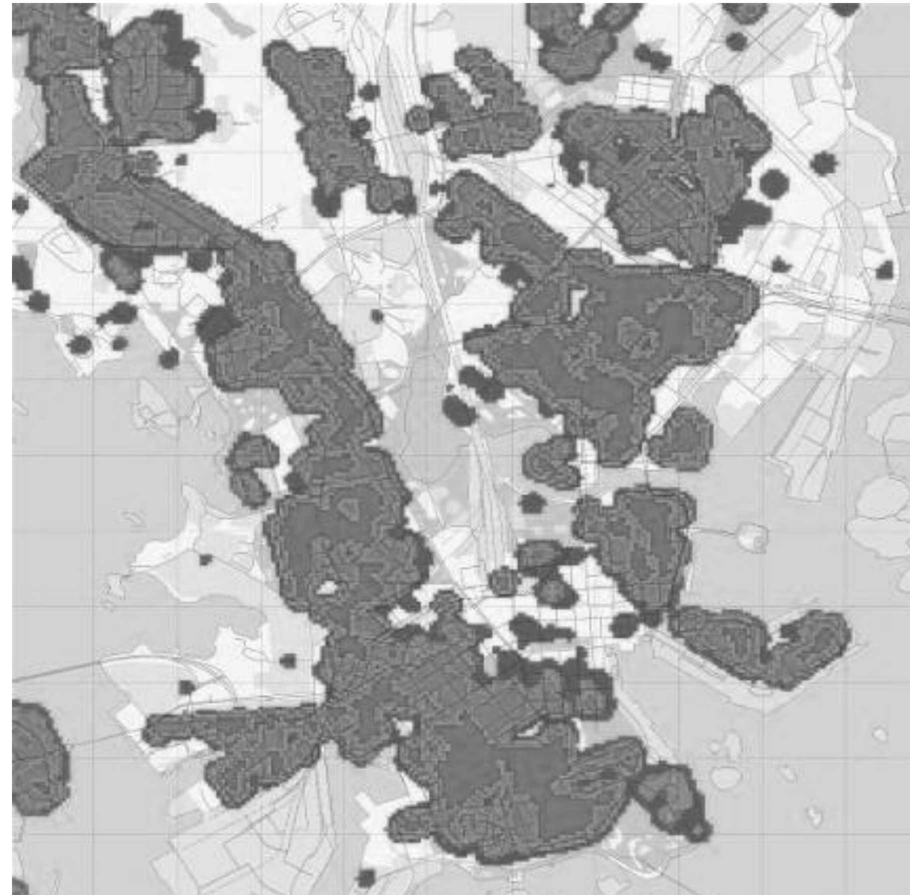


(Spatenkova, 2009)

Spatial data transformations

- **Another way to turn point data into a density surface is to use kernel function**
- **A kernel is calculated around each point**
- **Kernels are summed at each location (in practice: pixel) to get the Kernel density**





(Spatenkova, 2009)

Spatial data modelling: distributions and relations

Spatial distribution of elements

- **Distribution of objects in space is often important in spatial analysis**
  - described in terms of spatial relationships between objects
- **These relationships can be for example:**
  - Distance
    - Euclidian distance, distance in network, travel time, etc.
    - metric relationship
- Adjacency
  - polygons that share an edge (common boundary)
  - **topological** relationship
  - other definitions are possible, such as the nearest neighbour or distance-based adjacency
- Interaction
  - Interaction between two objects is inverse-weighted by their distance; only pairs of adjacent polygons are considered

Spatial distribution examples

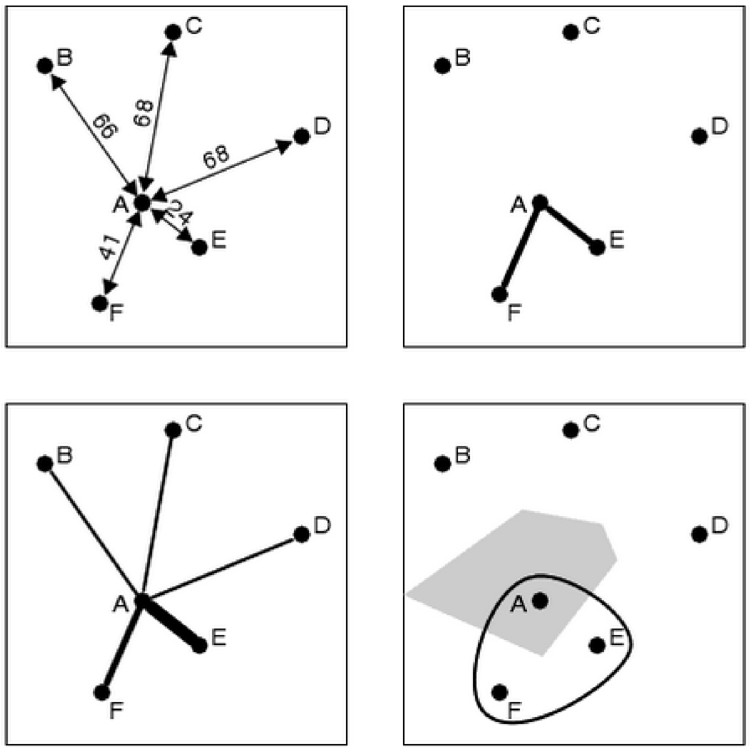


Figure 2.2 A schematic representation of the distance, adjacency, interaction, and neighborhood concepts.

Distance matrix D for distances between elements A, B, C, D, E, and F

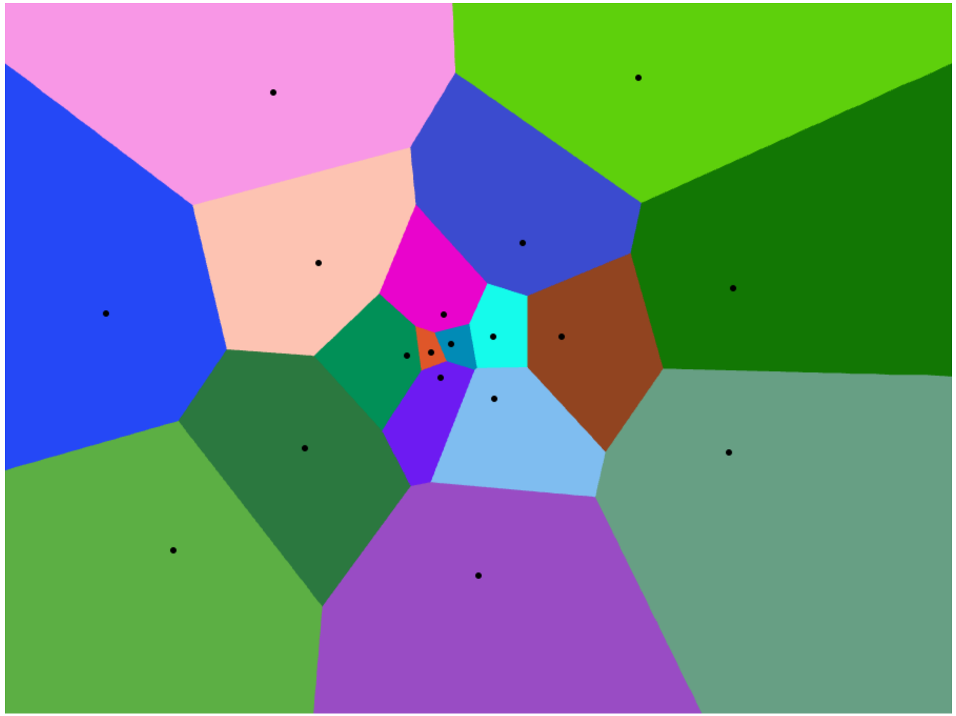
$$\mathbf{D} = \begin{bmatrix} 0 & 66 & 68 & 68 & 24 & 41 \\ 66 & 0 & 51 & 110 & 99 & 101 \\ 68 & 51 & 0 & 67 & 91 & 116 \\ 69 & 110 & 67 & 0 & 60 & 108 \\ 24 & 99 & 91 & 60 & 0 & 45 \\ 41 & 101 & 116 & 108 & 45 & 0 \end{bmatrix}$$

(Remember: point distances are something you need in WA-1)

Source: O’Sullivan & Unwin (2010) p. 41-50, figure from p. 46

Proximity polygons: Voronoi diagram

- **A proximity polygon represents the region of space closest to a given point in a point set**
- **A Voronoi diagram (Thiessen polygon) is a set of proximity polygons for a point set**
- **You will be using these in WA-3**
- **Basic Voronoi diagram uses straight-line distance, not e.g. travel time**



Generated using http://alexbeutel.com/webgl/voronoi.html

Spatial relations between objects: metric

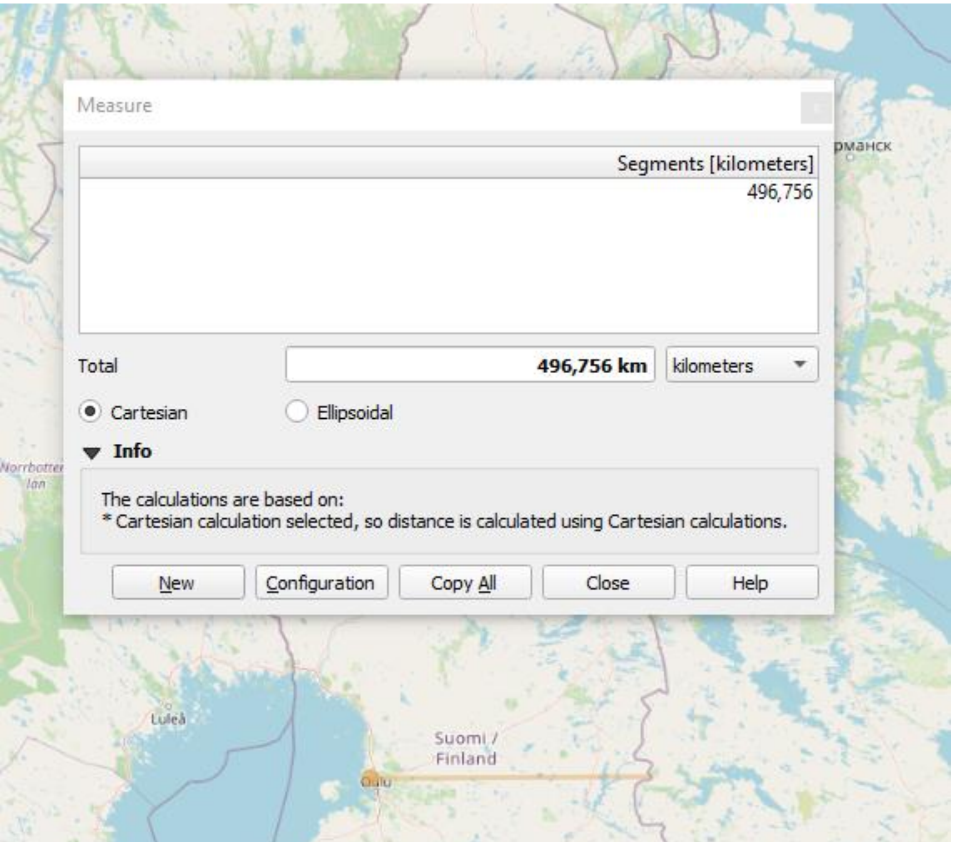
- **Metric relationships**
  - Distance, direction
  - Measured for a specific geometry (e.g. geometric coordinate system)
  - When geometry is distorted (e.g. map projection changed), (apparent) values change

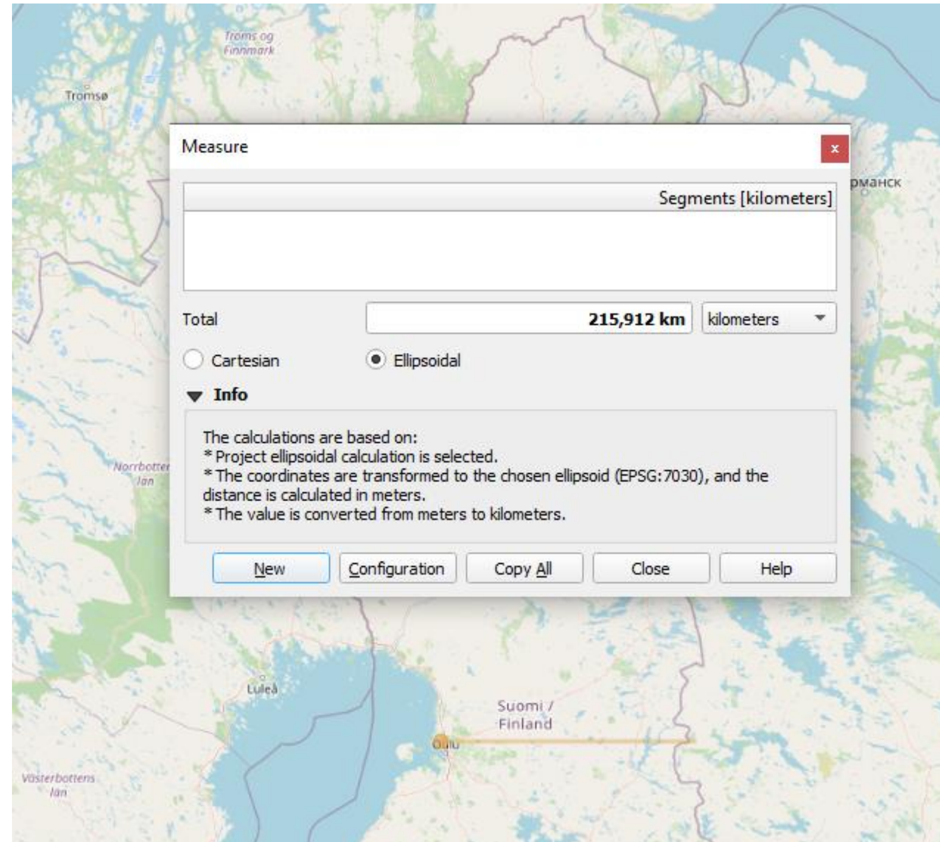
Map in the image uses Mercator projection

The great circle route is shorter than the Rhumb line route

Image source: Zhang, Q., & Zhang, K. (2015, June). 2D and 3D webpage visualization of the great circle and rhumb line. In *2015 23rd International Conference on Geoinformatics* (pp. 1-3). IEEE.

Metric relation example



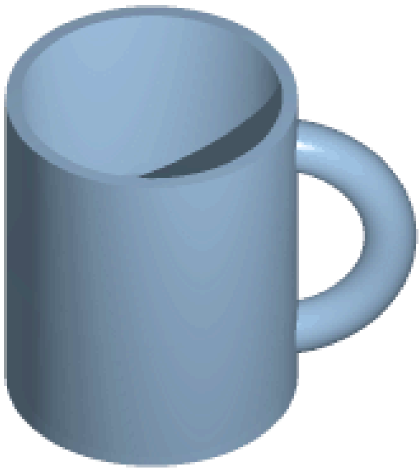


Which value is the correct one? Why is one of the calculations incorrect?

Spatial relations between objects: topological

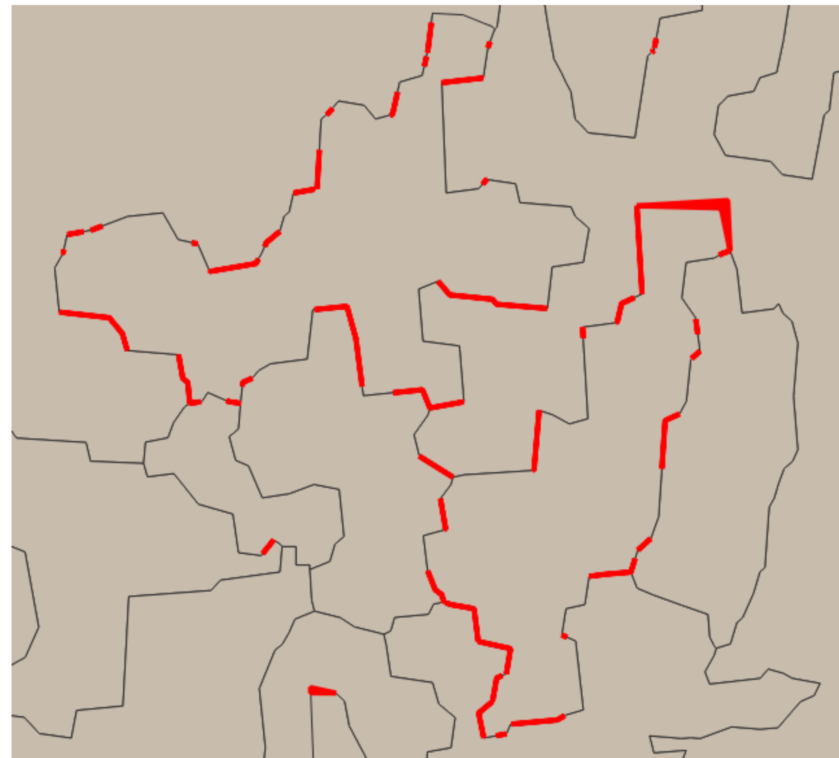
- Topological relations are preserved through deformations of objects.
- The common description in two dimensions is to imagine a sheet of rubber. Topological relations are preserved no matter how the sheet is twisted, or stretched, or otherwise modified (in ways that do not involve breaking the sheet)
- In GIS, topology is related to terms such as *adjacency, connectivity,* or *containment*

- **Topological relations can be used to e.g. validate aspects of vector geometry**



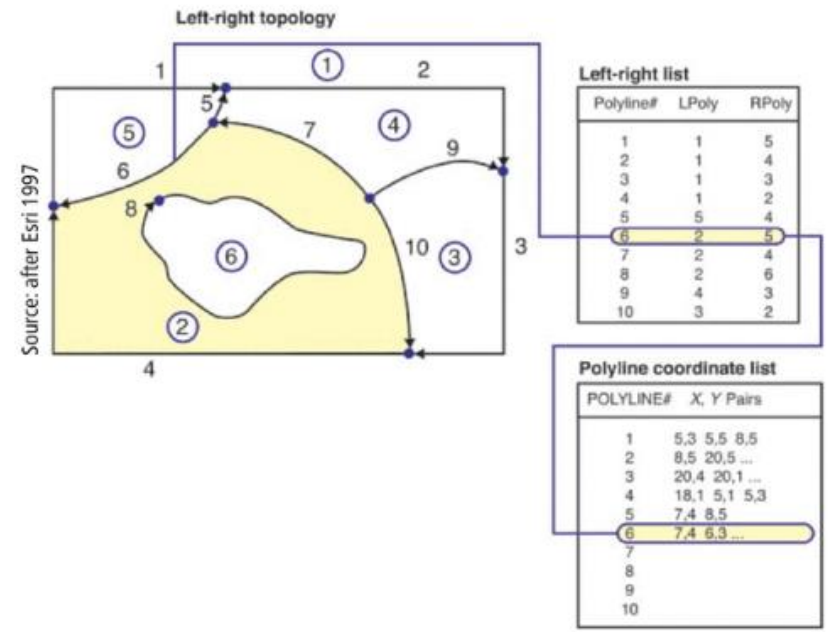
Topological relations and geometry validation: adjacency

- **The adjacency relation is often used in geometry validation**
  - E.g. are polygons in a polygon network really adjacent?
- **Lines are adjacent if they share an endpoint**
- **Polygons are adjacent if they share a boundary line**
  - The boundary line must be identical for the entire length of shared boundary (between shared line endpoints)



Topological data

- **For many applications topology is explicitly stored in the data model**
- **In this example**
  - Each line knows its left and right area
  - In order to have semantics for “left” or “right”, the line has a direction
  - Two neighboring areas share an edge



Left-right list

Polyline#	LPoly	RPoly
1	1	5
2	1	4
3	1	3
4	1	2
5	5	4
6	2	5
7	2	4
8	2	6
9	4	3
10	3	2

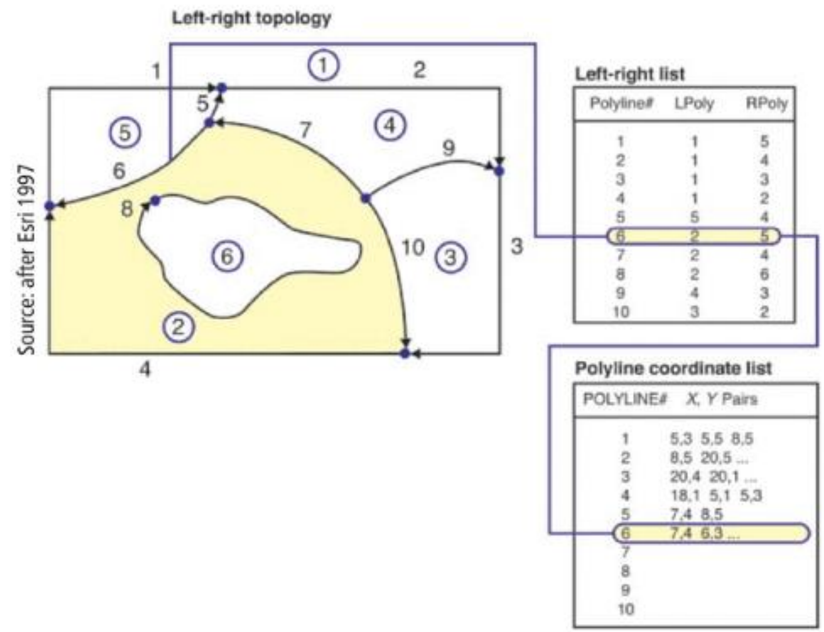
Polyline coordinate list

POLYLINE#	X, Y Pairs
1	5,3 5,5 8,5
2	8,5 20,5 ...
3	20,4 20,1 ...
4	18,1 5,1 5,3
5	7,4 8,5
6	7,4 6,3 ...
7	
8	
9	
10	

Longley et al. 2015

Classroom exercise: Adjacency

- **Which polygon has the largest number of adjacent polygons?**
- **How many lines bound polygon 1?**
- **Which line is adjacent to smallest number of other lines?**



Polyline#	LPoly	RPoly
1	1	5
2	1	4
3	1	3
4	1	2
5	5	4
6	2	5
7	2	4
8	2	6
9	4	3
10	3	2

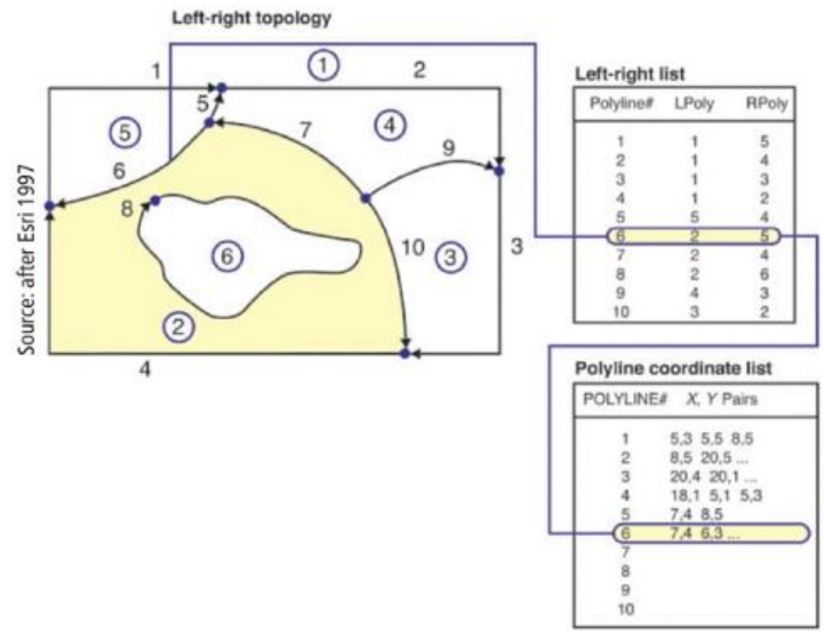
POLYLINE#	X, Y Pairs
1	5,3 5,5 8,5
2	8,5 20,5 ...
3	20,4 20,1 ...
4	18,1 5,1 5,3
5	7,4 8,5
6	7,4 6,3 ...
7	
8	
9	
10	

Aalto-yliopisto
Insinööritieteiden korkeakoulu

Longley et al. 2015

Classroom exercise: Adjacency

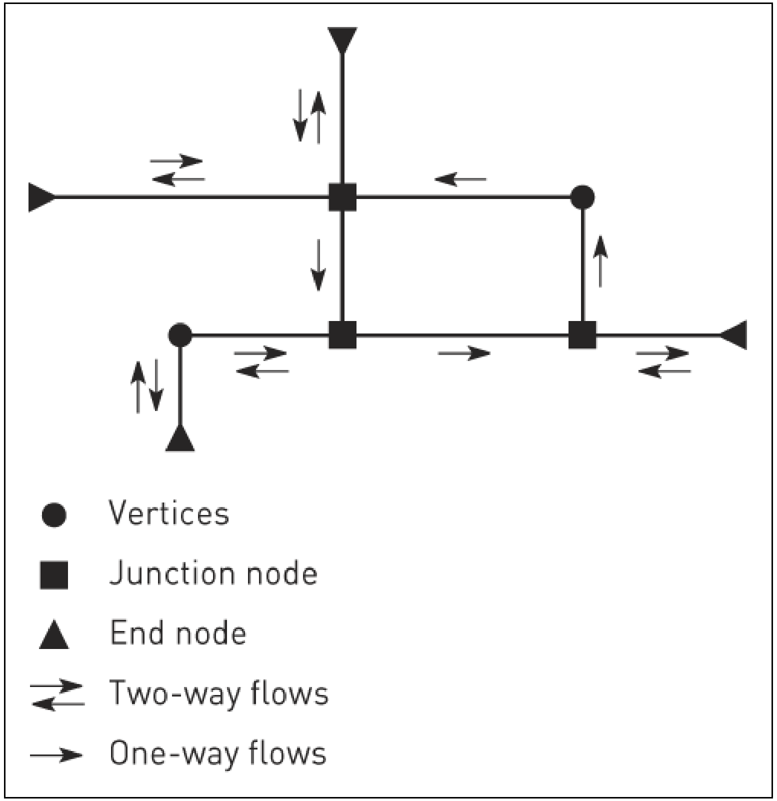
- **Which polygon has the largest number of adjacent polygons? (number 2)**
- **How many lines bound polygon 1? (4 lines)**
- **Which line is adjacent to smallest number of other lines? (line 8)**



Longley et al. 2015

Topological data and networks: connectivity

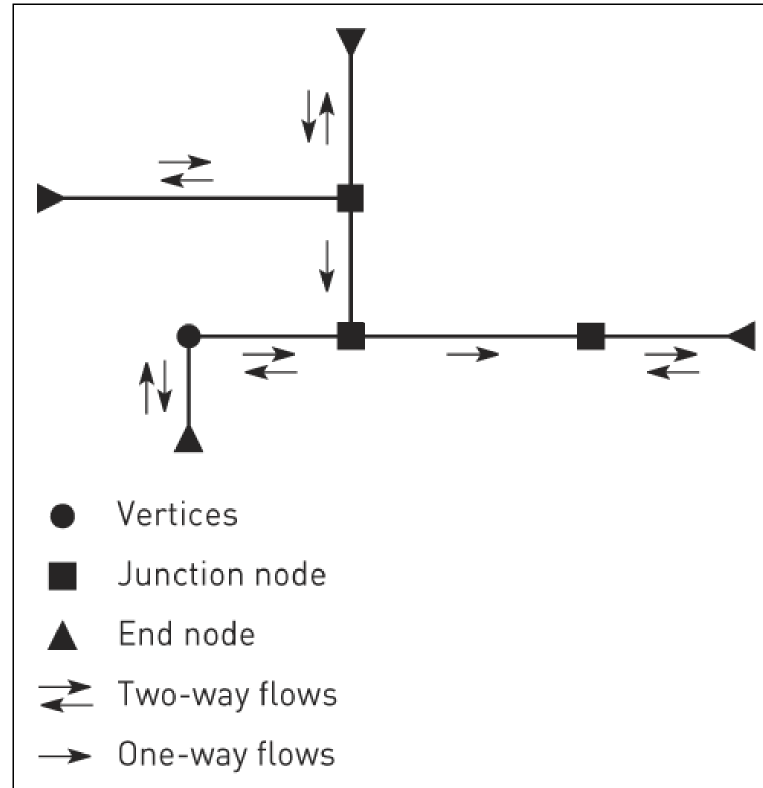
- **Graphs or networks are data sets consisting of vertices (nodes) and edges that connect them**
  - Edges can have a direction
- **Two vertices are connected if there is a path between them**
  - A path consists of one or more edges
- **In the example, all vertices are connected**



Heywood et al. 2006

Topological data and networks: connectivity

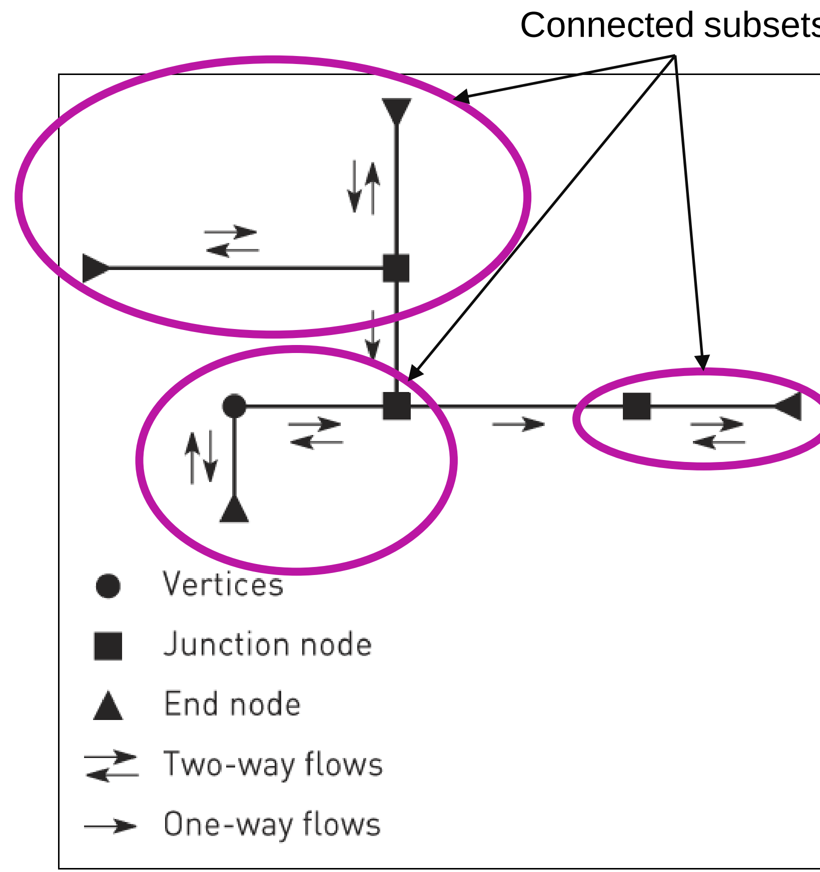
- **Are all vertices still connected?**



Heywood et al. 2006

Topological data and networks: connectivity

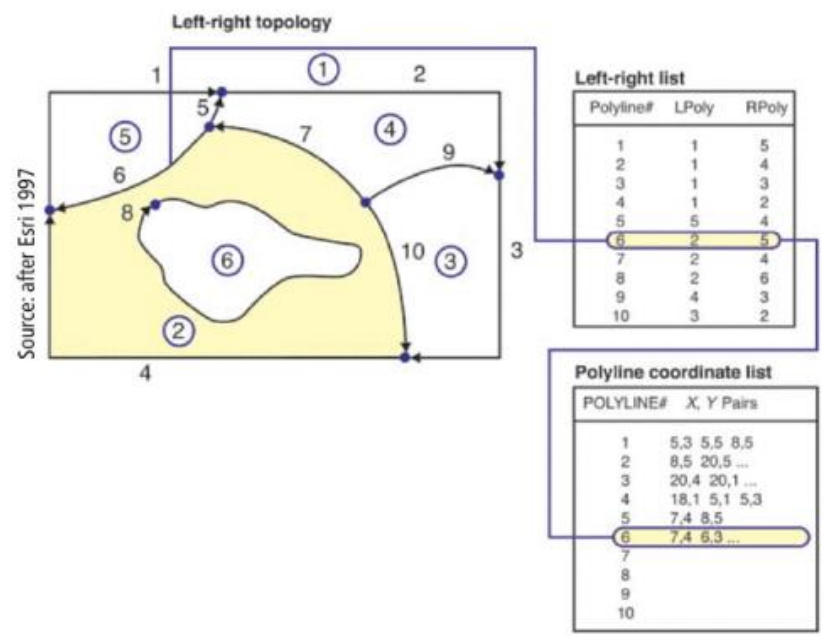
- **Are all vertices still connected?**
- **No, the one-way edges prevent the whole network from being connected**



Heywood et al. 2006

Topological data: containment

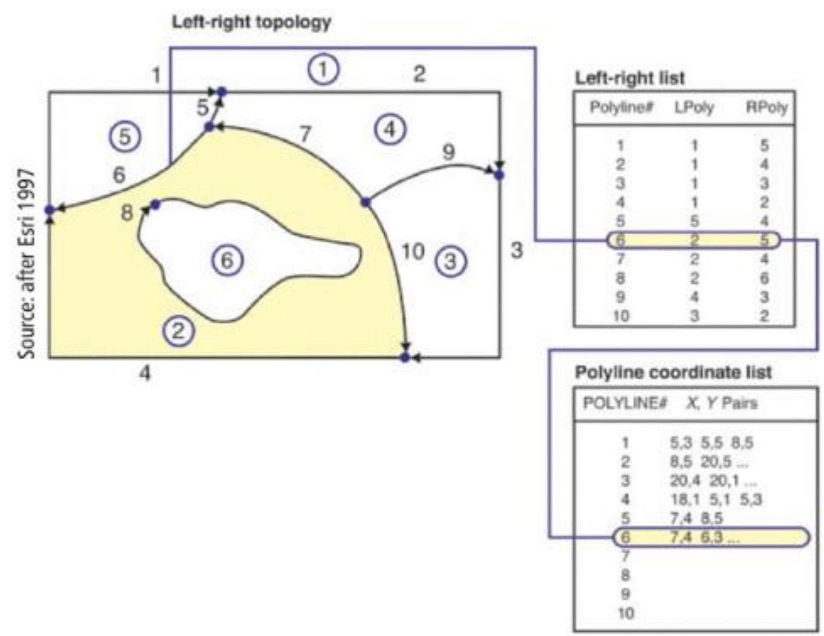
- **A polygon contains a data element, if the data element is completely within the polygon**
  - The element is completely inside the polygon
- **In the figure, are there polygons that contain other polygons?**



Longley et al. 2015

Topological data: containment

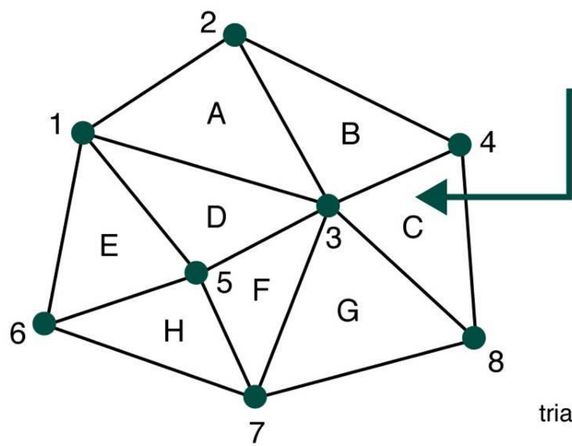
- **A polygon contains a data element, if the data element is completely within the polygon**
    - The element is completely inside the polygon
- **In the figure, are there polygons that contain other polygons?**
    - Polygons 1 and 2



Longley et al. 2015

Topological data structure example: TIN

A TIN is a topologic data structure that manages information about the nodes that comprise each triangle and the neighbours to each triangle



Triangle	Node list	Neighbours
A	1, 2, 3	-, B, D
B	2, 4, 3	-, C, A
C	4, 8, 3	-, G, B
D	1, 3, 5	A, F, E
E	1, 5, 6	D, H, -
F	3, 7, 5	G, H, D
G	3, 8, 7	C, -, F
H	5, 7, 6	F, -, E

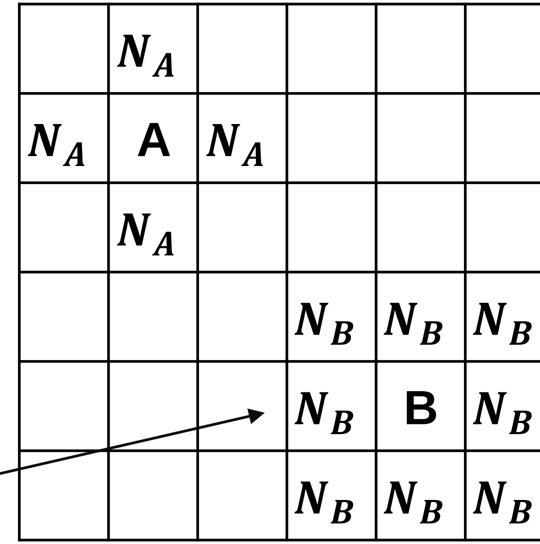
Triangles always have three nodes and usually have three neighbouring triangles. Triangles on the periphery of the TIN can have one or two neighbours.

Periphery as a neighbour of a triangle is marked with ”-” in the table.

Aalto-yliopisto
Insinööritieteiden
korkeakoulu

Raster data topology

- **In a raster grid topology is based on cell (pixel) adjacency**
- **Typically adjacency is defined for a cell and its immediate**
  - 4 neighbors
  - 8 neighbors
- **More complex neighborhoods are also possible**



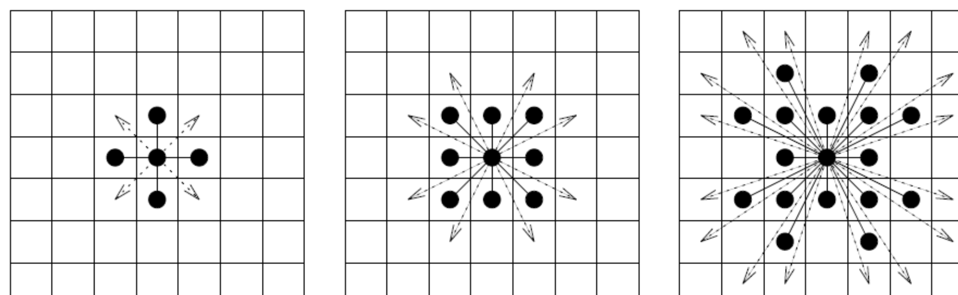


Fig. 4: 8-connected. Fig. 5: 16-connected. Fig. 6: 32-connected.

Image source: Van Bemmelen et al.

Reading for the lecture

- **Longley et al. (2015): Geographic Information Science and Systems Ch. 3.4-3.6, 7.23**

- **O'Sullivan & Unwin (2010): Geographic information analysis pp. 5-18, 41-50**