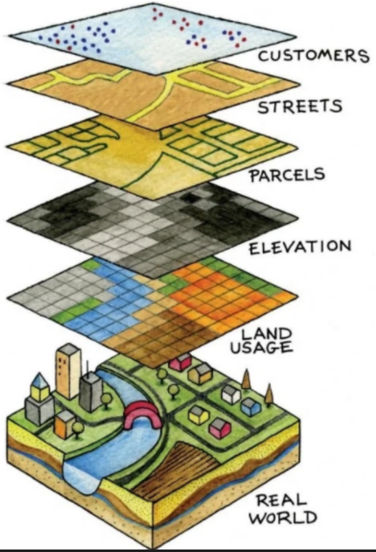
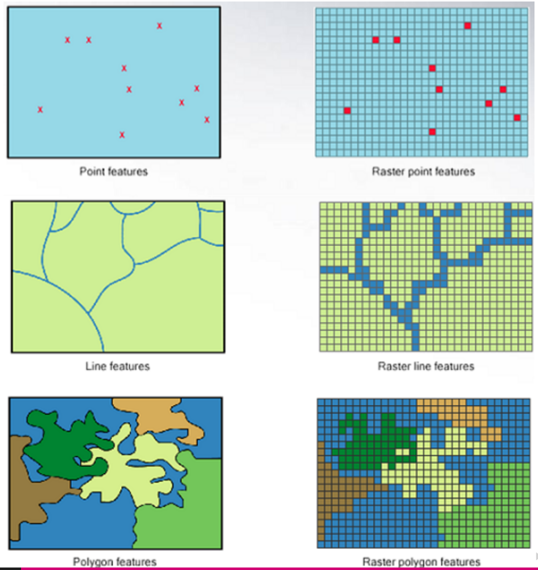
Ekonomická fakulta
Faculty of Economics
Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

Geographic Information Systems 1
Lecture 4: raster data representation

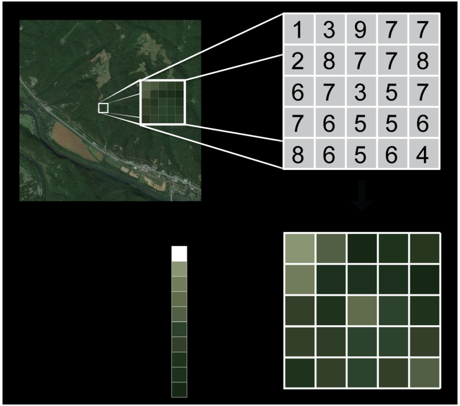
Renata Klufová
University of South Bohemia, Faculty of Economics

March 2021


CUSTOMERS
STREETS
PARCELS
ELEVATION
LAND USAGE
REAL WORLD


Point features
Raster point features
Line features
Raster line features
Polygon features
Raster polygon features

The basis of raster data is the overlap of the studied entity or area with a regular network (grid). The entity is then described by discrete values that are related to the fields of this network. The positional location of an entity is determined by the coordinates of the fields that represent it.

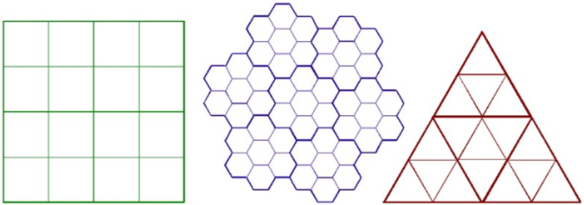


Raster data is especially suitable for the representation of **continuous phenomena**, such as:

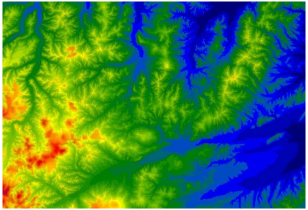
- air and water temperature,
- altitude,
- geologic data,
- precipitation map,
- surface runoff density,
- aerial and satellite imagery and more.

(a) **regular** - the shape of the cells is precisely defined and always the same - square, rectangle, hexagon, triangle
- with the **same resolution level** - individual cells are still the same size,
- with the **unequal resolution level** (hierarchical) - cell size changes in a defined way,

(b) **irregular** - cells of different shape and size.



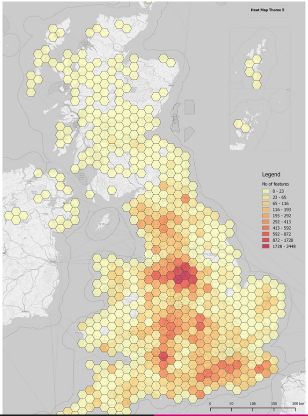
- Practical applications consist mainly in the use of textbf regular divisions, especially with the same level of resolution.
- **Hierarchical structures** are sometimes understood more as data compression methods.



- Historically, the most used are square rasters, mainly for the following reasons:
  - compatible with data structures used in computer technology (matrices),
  - compatible with many HW devices for data recording and output (scanners, printers, plotters),
  - compatible with Cartesian coordinate systems.

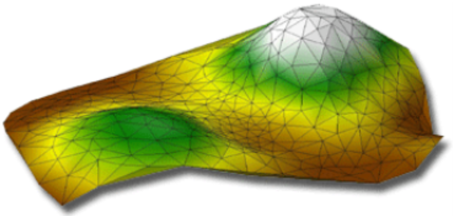
hexagonal grid

- the centers of all adjacent cells equidistant from the center of the cell,
- a model suitable for some analytical functions, eg for radial scanning,
- rarely used in practice - now more and more often.

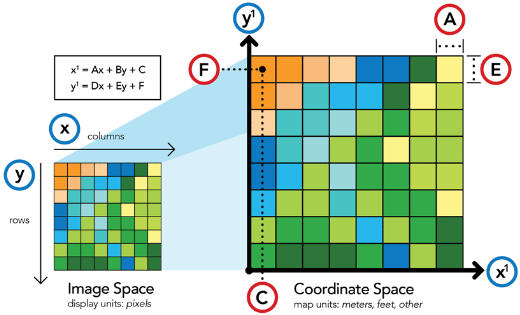


triangular grid:

- the triangles do not have the same orientation, which is advantageous for terrain and other surface representations,
- most often triangles with variable size and shape - so-called **triangular irregular network (TIN)**.

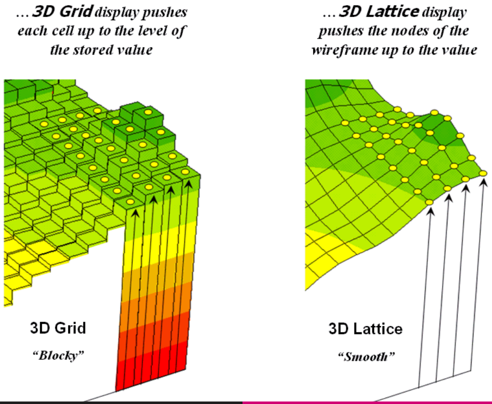


- **raster size (range)** - given by the number of columns and rows in the matrix, and the spatial resolution of the cell,
- **raster georeferencing** - conversion from image coordinates to the coordinates of the relevant cartographic display (real coordinates).

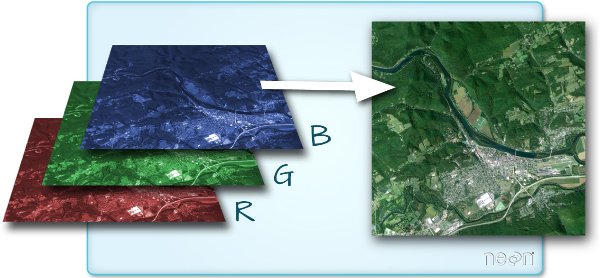


point versus area raster

- **area raster** - cell system (**pixel** = picture element, three-dimensional variant = **voxel**),
- **dot grid (lattice)** - a system at right angles of intersecting lines that delimit individual cells.

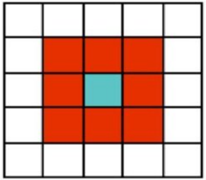


multi-band raster - one raster file can contain multiple sub-images with identical basic parameters of the raster matrix - each of the sub-images then differs in the values that its cells contain = **bands** - satellite or digital aerial images

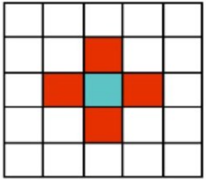


raster cell topology - implicitly defined in raster geometry - each pixel has two types of neighbors:

- full neighbors - adjacent pixels in the same rows or columns - these cells form the so-called **four-member neighborhood (von Neumann)**;
- diagonal neighbors - pixels that touch the person in its corners;
- Both groups of neighbors form the so-called **eight-member neighborhood (Moore's)**.



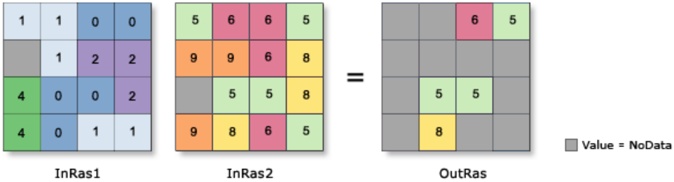
a



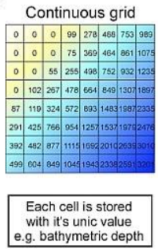
b

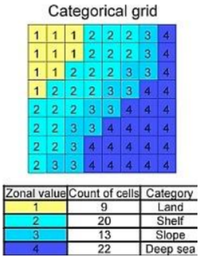
In the raster model, there are cases when the cell does not contain **no value** - possible solutions:

- NoValue/NoData,
- -9999

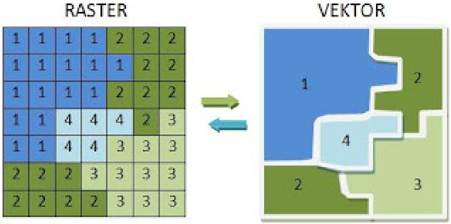


- **quantitative** raster - describes the size of the phenomenon recorded in a given cell. Cells can take real number values. The values used in the quantitative raster are usually from the ratio or interval domain;
- **qualitative** raster - expresses the quality (type, kind, class, category) of the phenomenon. Cells carry an integer value. The cell values used in these rasters are from the enumerated (nominal) domain.





- **rasterization**: vector → raster (see below - assigning values to a cell),
- **vectorization**: raster → vector.



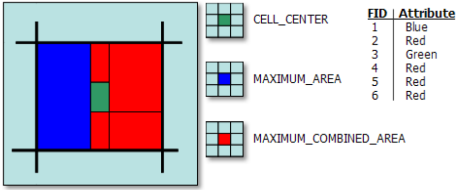
The quality of the real-world display using a raster data model is affected by several factors:

- the method of assigning the values of the displayed attribute to individual cells,
- **spatial resolution** - size of the basic raster cell,
- **color resolution** - the number of values that a pixel can take (record attribute values), is correctly called in the case of a digital image "**radiometric resolution**" in ESRI help, this number is then called **bit depth**.

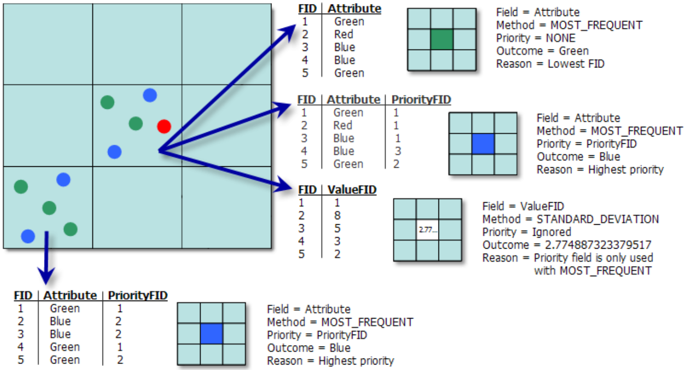
Method of assigning the values of the displayed attribute to individual cells

Attribute values in individual cells can be determined in different ways:

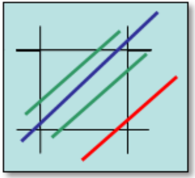
- a point value measured anywhere in the cell area,
- arithmetic mean of several point measurements,
- weighted arithmetic mean, where the weight is the area of individual values,
- maximum or minimum value of the attribute,
- the value of the attribute with the highest weight,
- the most frequently recurring value (mode) and other options.

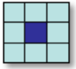


Method of assigning the values of the displayed attribute to individual cells



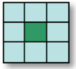
Method of assigning the values of the displayed attribute to individual cells





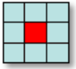
FID	Attribute
1	Green
2	Blue
3	Green
4	Red

Field = Value
Method = MAXIMUM_LENGTH
Priority = NONE
Outcome = Blue
Reason = Longest length



FID	Attribute
1	Green
2	Blue
3	Green
4	Red

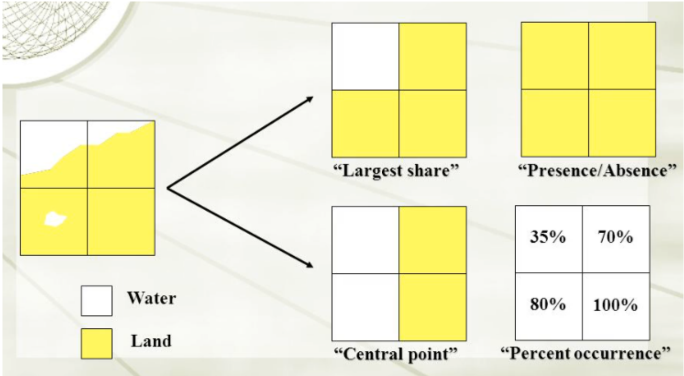
Field = Value
Method = MAXIMUM_COMBINED_LENGTH
Priority = NONE
Outcome = Green
Reason = Length of two green



FID	Attribute	PriorityFID
1	Green	1
2	Blue	1
3	Green	2
4	Red	3

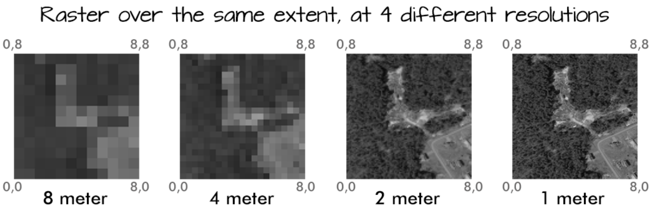
Field = Value
Method = MAXIMUM_LENGTH
Priority = PriorityFID
Outcome = Red

Method of assigning the values of the displayed attribute to individual cells



The size of the raster base cell

- In general, the smaller the basic cell of a raster, the better (more accurately) the course of the boundaries of individual geoelements can be captured in this raster.
- However, the same is true: halving the side length of the raster base cell quadruples the memory space required to store the raster.



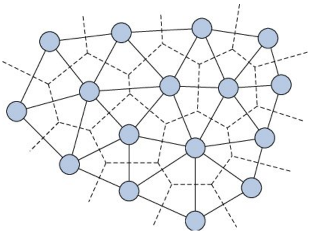
- "**minimum cartographic length rule**": the length of the cell side or step size that defines the resolution of the raster to choose as half of the smallest length needed to represent objects existing in real conditions.

"Color depth" (resolution) of the raster

When working with rasters, we use different resolutions in individual cells to record the values of the monitored attribute. According to the resolution used, we distinguish the following types of rasters:

- **binary** - only the presence, resp. absence of attribute (most often values 0 and 1) - to record the value of one cell we always need one bit - e.g. scanned cadastral maps;
- **8-bit** - we distinguish 256 different integer values of the monitored attribute in the cell - to record the value of one raster cell we need 1 Byte - eg scanned color masters, panchromatic aerial and satellite images;
- **24-bit** - there are about 1.6 million different integer values of the monitored attribute in the cell - to record one cell we need 3 bytes - e.g. multispectral satellite images;
- **continuous** - in a cell we distinguish an almost unlimited number of real values of the monitored attribute - to record one cell we usually need 4, resp. 6 Byte.

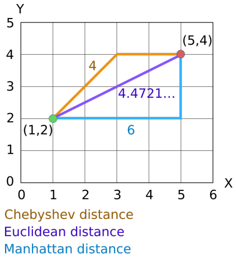
- A large number of triangular networks can be created from a certain set of entry points = **triangulation**.
- The so-called **Delaunay triangulation** is of great importance - no other point falls into the circle circumscribed by each of the triangles. The vertex of each triangle of the network is bounded by a convex triangle.



- The duality of this triangulation is called the **Voronoi diagram** or **Thiessen polygons** or **Dirichlet tesselace**. The vertices of the Voronoi polygons are also the centers of the circles circumscribed by the Delaunay triangles. The perpendiculars leading to the centers of the sides of the Delaunay triangulation form the edges of the Voronoi polygons.

The following metrics are most commonly used in rasters:

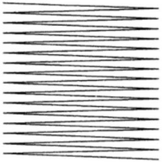
- **edge metric (block metric)** - the distance between two cells is defined as the minimum number of overcome cell edges, sometimes also referred to as **Manhattan metric**;
- **metric of edges and centers (checkerboard metric)** - the distance between two cells is defined as the minimum number of overcome edges or centers (so the diagonal direction of movement is permissible),



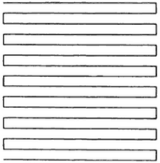
- **Euclidean metric** - the position of each cell is represented by the position of its center, the distance between cells is then defined as the distance of their centers.

Different data structures are used to store raster data:

- saving data **by cells** (most often in a text file) - the simplest: three data on a line: row and column index (resp. coordinates i, j, or x, y) and the value represented by the cell - in terms of disk space requirements the least preferred method;
- in the form of **matrix** - disadvantages: memory intensive, it is necessary to attach accompanying data to the stored data, which inform the user, for example, about the dimensions of the raster - rewriting of matrix values:
  - **in rows or columns** (row order),
  - **backwards in rows or columns** (row-prime order).



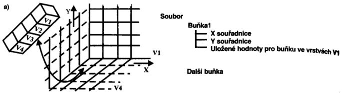
(a) Row Order



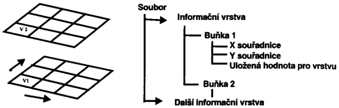
(b) Row-prime Order

Different data structures are used to store raster data:

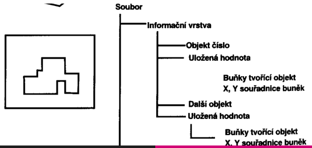
(a) direct cell dating,



(b) direct layer dating,



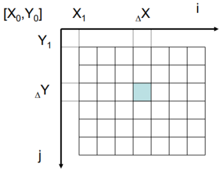
(c) direct object dating.



- to define the position of the cell with respect to the beginning of the raster, a pair of indices i, j is sufficient: index j - row index, index i - column index

  very important in terms of the required memory space
- calculation of indices:

  $[X_1, Y_1] = [i, j] = [0, 0] \qquad X = X_1 + p \cdot \Delta X \rightarrow i = p \qquad Y = Y_1 + q \cdot \Delta Y \rightarrow j = q$



where $\Delta X, \Delta Y$ are the step distances (basic raster dimensions) that determine the resolution of the raster. Reducing them increases the resolution. The coordinates of the indices can be calculated by reversing the rules.

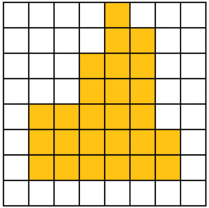
It is common for several identical values stored in consecutive cells to be repeated in the order thus formed.

- **data compression** = reduction of the volume of stored data by excluding their coordinates or indexes in the system of rows and columns of the raster (eg large uniform area - a large number of cells with the same value)
- for direct dating of each cell or direct dating of the information layer,
- principle: in the data file only the content of individual cells and their the position can be identified on the basis of the data of a special file in which the spatial parameters of the data file are defined (min. and max. values of coordinates of the examined area, number of rows and columns of the raster, cell size, etc.) - e.g. in its introductory part, so-called **header**. e.g. IDRISI

run-length encoding method - the original values of data from cells are replaced by the so-called **tupples**, which are formed by pairs of numbers: the first indicates the stored value and the second the number of its repetitions in order (lossless):

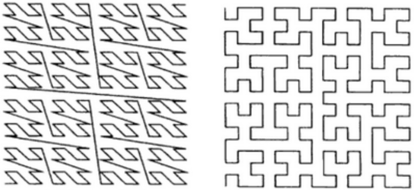
- suitable for simple block patterns, but not for continuous-tone photographic images;
- e.g. row of values: 1 1 1 1 1 3 3 3 4 2 2 2 2 2 3 3 3 3 3 3 is replaced as follows: (1 5)(3 3)(4 1)(2 5)(3 6).

The original 20 values were replaced by 10, ie we reached a compression factor of 50%.



Other options for determining the order of raster values according to which they will be overwritten from the data matrix to a file (space-filling curves):

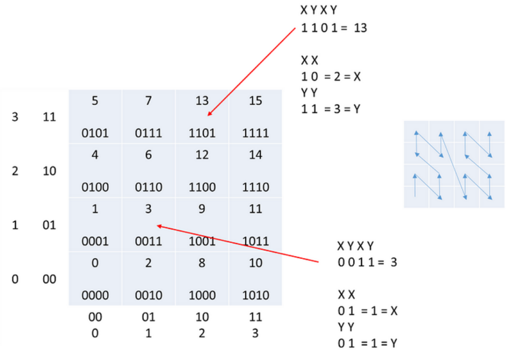
- **Morton's order** - used for rasters whose dimensions (number of rows and columns) are divisible by 4 and are the same - in contrast to row order, elements close to each other in space (area) become close in order, which is of great importance for some analyzes.
- **Peano order (Peano curve)**.



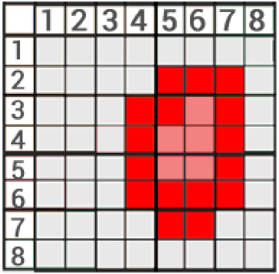
(c) Morton Order (d) Pi-Order

Data compression methods
Morton order - bit interleaving

If we keep the numbering of rows and columns from 0 and their number will be divisible by 2, it is possible to use textbf binary addressing (bit interleaving) to determine the position of a cell in the raster row and column system based on its position in a series of values created using Morton's order:



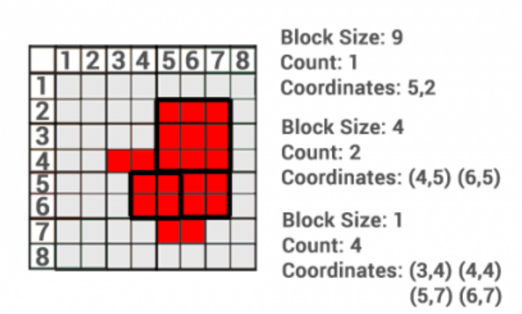
Chain codes - define the boundaries of each polygon by encoding the direction (orientation) of the boundary from the specified starting point. The disadvantage of this procedure is the multiple placement of boundary sections common to multiple polygons.



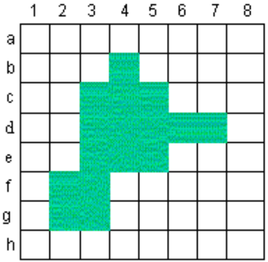
(5,2)

(E3, S4, W1, S1, W1, N1, W1, N3, E1, N1)

Block codes - indicate the position of reference points and the size of square blocks from which the whole object can be created.

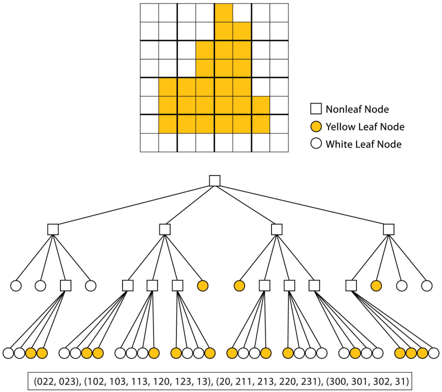


Run length codes - defines the affiliation of raster cells to an object row by row or column, specifying only the beginning and end of a section of cells in a system of rows and columns that have the same value stored. Some authors also refer to this procedure as the ”standard run-length codes”.



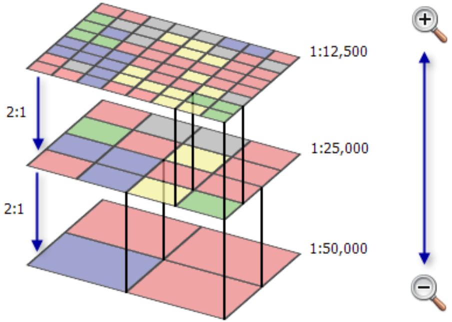
řádek b 4,4
řádek c 3,5
řádek d 3,7
řádek e 3,5
řádek f 2,3
řádek g 2,3

Quadtrees method - is based on the gradual regular division of the raster into quadrants and sub-quadrants until either the same value of the monitored attribute is the same in the whole quadrant or the smallest possible quadrant size (equal to the basic cell dimension) is reached.



three-dimensional space - cube → so called octtree

pyramid - stores all cells in the hierarchy, while the value represented in higher cells is usually determined as the average value of cells lying one level lower



From the point of view of the implementation of individual components of the description of geoelements, the raster model is the worst off (Rapant, 2002). Most of the problems arise because in the raster model it is not possible to work directly with individual geoelements, but only with rasters showing the distribution of properties of geoelements in the area of interest:

- **the geometric component** is included in this model only implicitly, explicit expression is practically impossible,
- **thematic component** is realized in the form of individual rasters, showing the distribution of properties in the area of interest,
- **time component** can only be captured as a sequence of rasters showing the distribution of the same attribute at different points in time,
- **relational component** can be realized only to a limited extent, to the extent corresponding to the raster possibilities,
- **functional component** can be implemented in the form of programs processing rasters.

Thank you for your attention.