



# Geoinformatika

## III. Rastrový datový model

jaro 2016

**Petr Kubíček**

**kubicek@geogr.muni.cz**

**Laboratory on Geoinformatics and Cartography (LGC)  
Institute of Geography  
Masaryk University  
Czech Republic**

# Vektorová data

## Výhody

- lze pracovat s jednotlivými objekty jako se samostatnými celky;
- menší náročnost na paměť;
- dobrá reprezentace jevové struktury dat;
- vysoká geometrická přesnost
- kvalitní grafika, přesné kreslení, znázornění blízké mapám;
- jednoduché vyhledávání, úpravy a generalizace objektů a jejich atributů.

## Nevýhody

- výpočtová náročnost (problémy při náročných analytických operacích);
- komplikovanost datové struktury;
- složitější odpovědi na polohové dotazy;
- obtížná tvorba překryvů vektorových vrstev (overlay)
- problémy při modelování a simulaci jevů.

# Rastrová reprezentace

- **Zaměřuje se na lokalitu jako na celek**
- **Používá se pro reprezentaci jevů, které plošně pokrývají celou oblast, případně se i spojitě mění.**
- **Používá se i pro rasterizované vektorové vrstvy, pokud je následná analýza jednodušší nad rastrem.**
- **RAVE - VERA**

# Rastrová reprezentace

- Základním stavebním prvkem je u rastrové struktury tzv. **buňka** (cell, pixel).
- Buňky jsou organizovány do **mozaiky**.
- Jednotlivé buňky obsahují **hodnoty** (values).
- **Typy tvarů buněk:**
  - čtvercová buňka (lattice, grid)
  - trojúhelníková buňka,
  - hexagonální buňka.

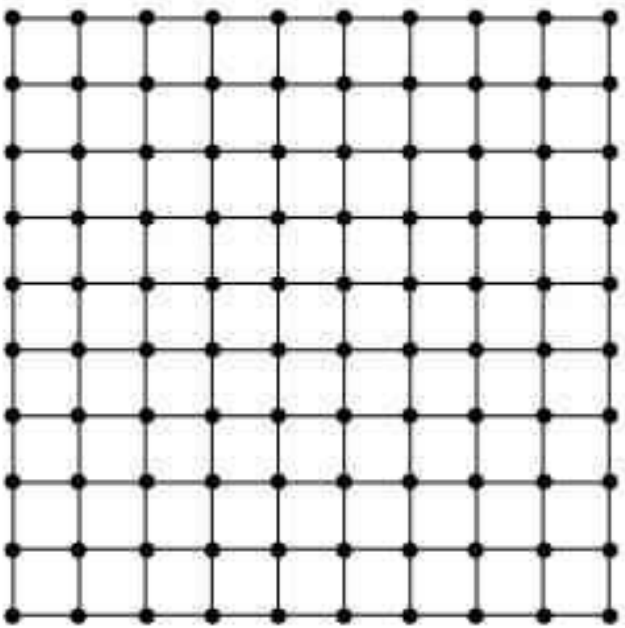


# Typy rastrové reprezentace

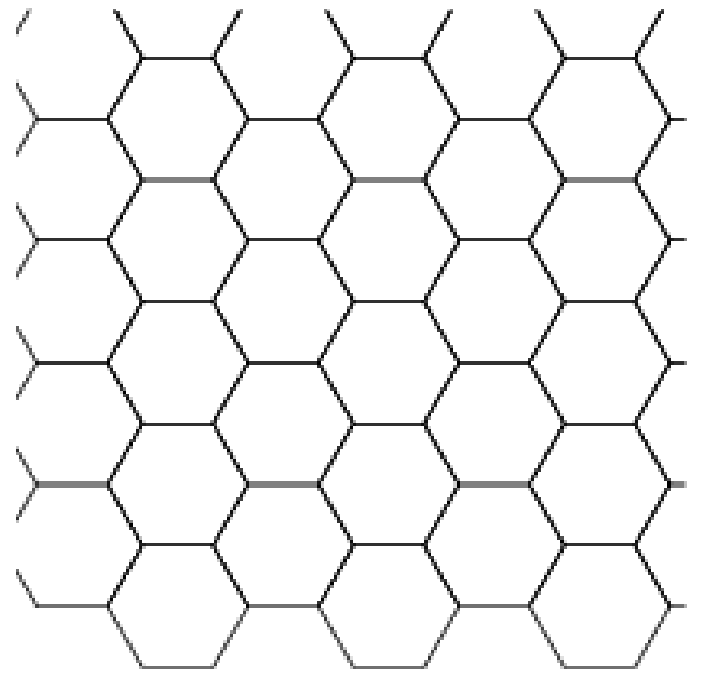
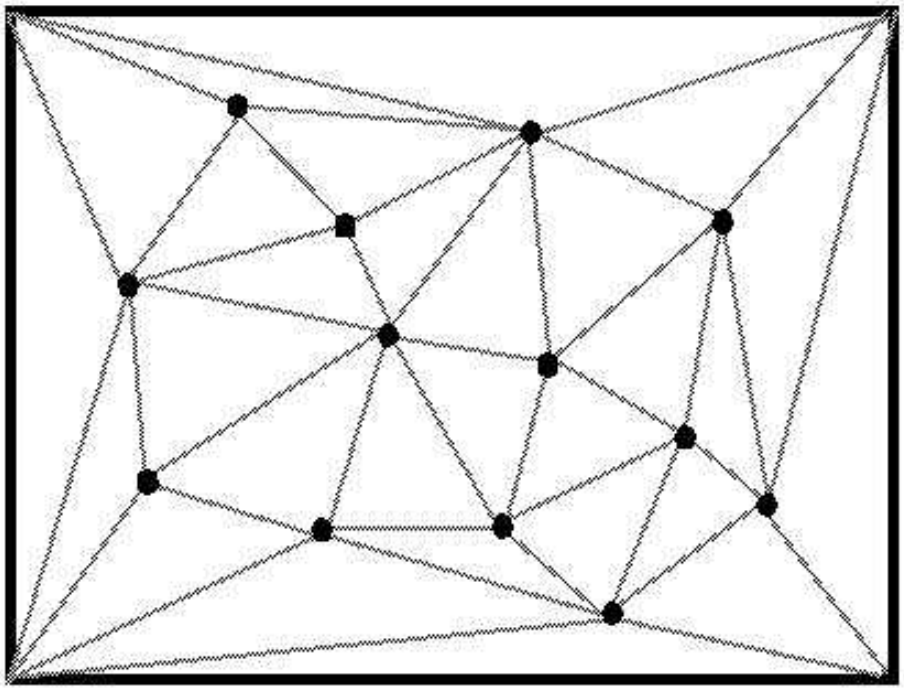
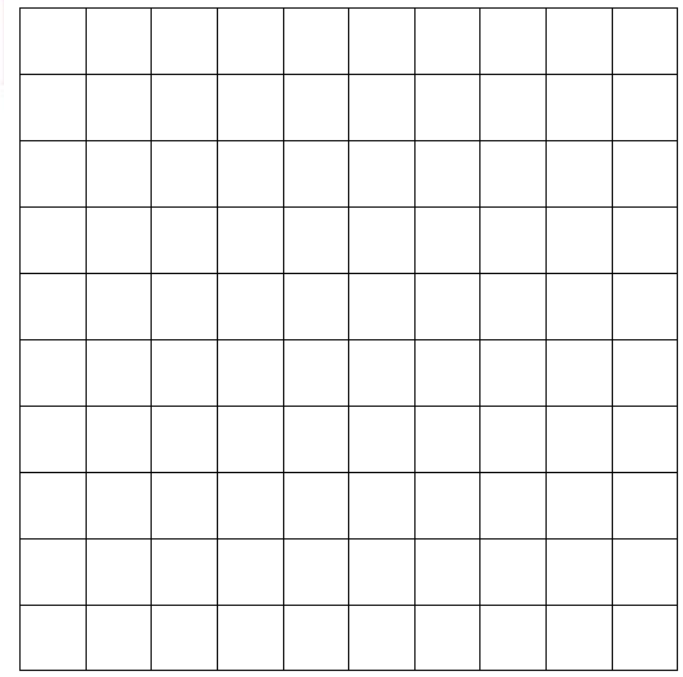
**Rastrovou reprezentaci můžeme rozlišit podle způsobu dělení prostoru na:**

- **pravidelné (regular) - všechny buňky mají stejnou velikost a tvar.**
  - jednodušší pro ukládání a zpracování údajů, zabírají ovšem na disku mnoho místa.
- **nepravidelné (irregular) - velikost i tvar jednotlivých buněk se liší.**
  - mohou mnohem lépe reprezentovat danou lokalitu (příklad roviny + zvlněná krajina),
  - zpracovávání je algoritmicky i výpočetně náročné. Hlavně pro DMR.

Lattice Network



10 x 10 Grid



# Typy mřížky

## Nejčastěji se používá čtvercová mřížka:

- je kompatibilní s datovými strukturami programovacích jazyků používaných pro tvorbu GIS software,
- je kompatibilní s mnoha zařízeními pro vstup a výstup dat (monitory, scannery, plottery),
- je kompatibilní s **kartézským (pravoúhlým) souřadnicovým systémem**.

## Trojúhelníková mozaika

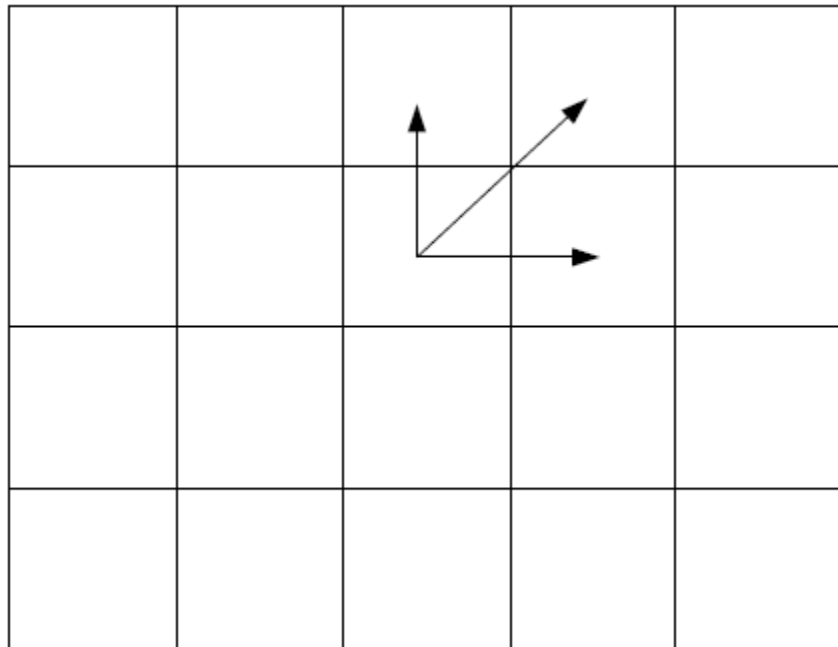
- jednotlivé buňky nemají stejnou orientaci – výhoda při reprezentování digitálního modelu reliéfu (terénu), kde je každému vrcholu o souřadnicích  $x, y$  přiřazena funkční hodnota  $z$  (výška  $z = f(x, y)$ ).
- Jednotlivé trojúhelníky pak implicitně obsahují údaje o svém sklonu a směru tohoto sklonu.

## Hexagonální mozaika

- středy všech sousedních buněk jsou ekvidistantní (stejně od sebe vzdálené), což je výhodné pro některé analytické funkce (např.: paprskové vyhledávání).

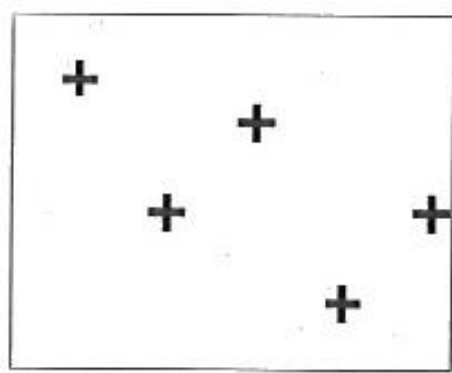
# Topologie v rastru

- **Topologie je v rastrovém modelu definována implicitně (je jasné, kdo je čí soused), tudíž není nutné ji explicitně ukládat jako pro vektorový model!**

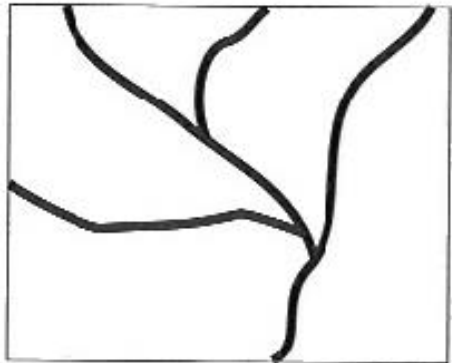
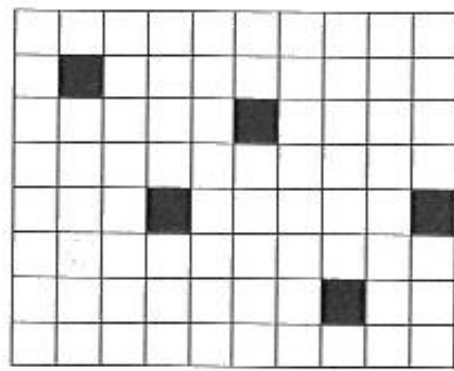




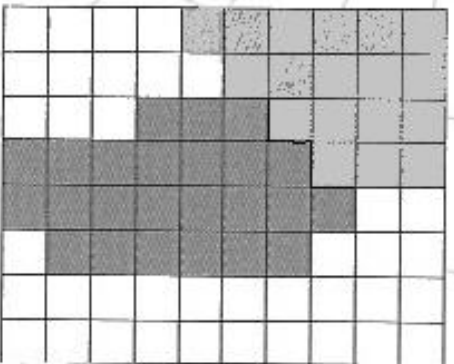
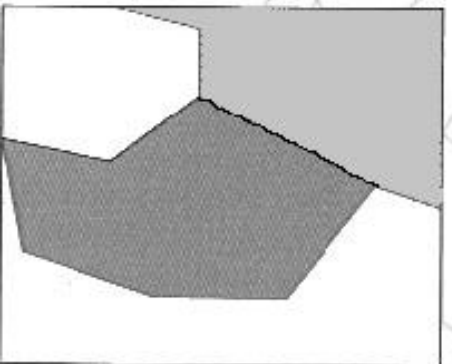
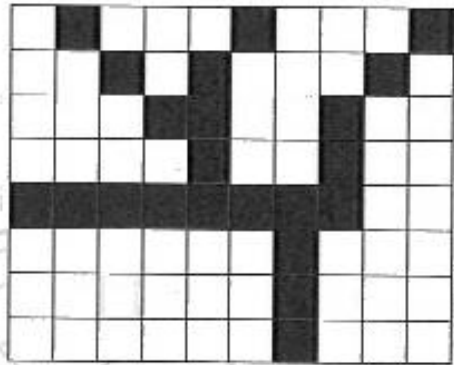
# Reprezentace geometrie v rastru



*Point features represented in a grid.*



*Linear features represented in a grid.*



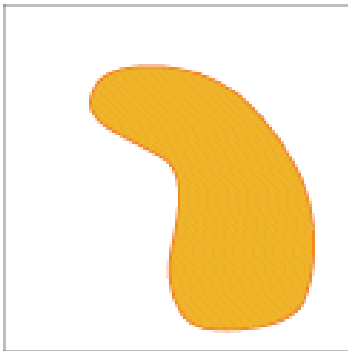
- **rastrová datová struktura může nést informace o bodech, liniích a plochách.**
- **Odlišné možnosti převodu mezi vektorem a rastrem.**



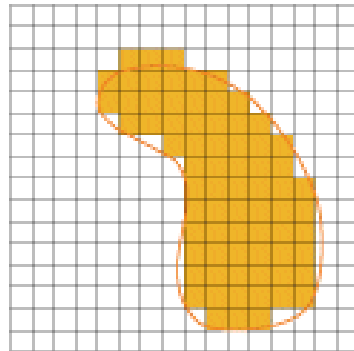
# Faktory ovlivňující vyjádření v rastru - rozlišení

- Vliv velikosti buňky ( $\sim$  rozlišení) na tvar objektů (+ a -)

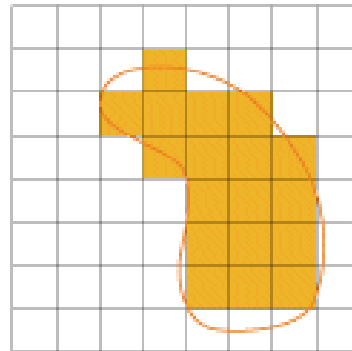
71 m<sup>2</sup>  
polygon



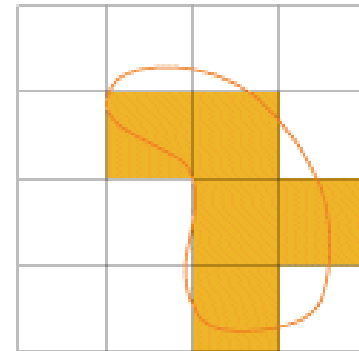
73 m<sup>2</sup>  
1 m cell  
16 x 16 cells



72 m<sup>2</sup>  
2 m cell  
8 x 8 cells



80 m<sup>2</sup>  
4 m cell  
4 x 4 cells



**PRO**

- Smaller cell size
- Higher resolution
- Higher feature spatial accuracy

**PROTI**

- Larger cell size
- Lower resolution
- Lower feature spatial accuracy

**PROTI**

- Slower display
- Slower processing
- Larger file size

**PRO**

- Faster display
- Faster processing
- Smaller file size



# Faktory ovlivňující vyjádření v rastru

## způsob přiřazení hodnot zobrazovaného atributu (kvantitativní data)

- jako bodová hodnota změřená **kdekoli v ploše buňky**
- jako **aritmetický průměr** u několika bodových měření
- jako **vážený aritmetický průměr**, kde váhou je plošný rozsah jednotlivých hodnot
- jako **maximální nebo minimální hodnota atributu** v ploše buňky
- jako hodnota atributu **s největší váhou** (i pro kvalitativní).

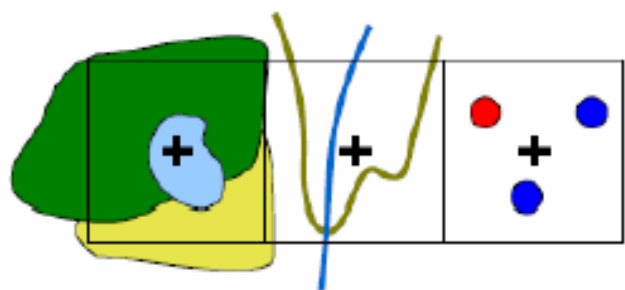
## datové rozlišení („barevná hloubka“ rastru)

- **binární rastr** (0x1, výskyt x nevýskyt) – záznam 1 bitem.
- **8bitový rastr** – 256 různých celočíselných hodnot, záznam 1 bajtem.
- **24bitový rastr** – 1,6 milionu různých celočíselných hodnot, 3 bajty.
- **kontinuální rastr** – hodnoty v reálných číslech, záznam 4 nebo 6 bajty.

# Řešení konfliktů

Problém - **jedna výsledná buňka obsahuje více různých objektů**. Pro řešení této se používají 3 základní metody, z čehož první dvě se používají pro převod bodů, linií i polygonů a zbývající jen pro převod polygonů:

- **Metoda dominantního typu** vychází z principu, že u buňky, do které zasahuje více objektů, se vyjádří podíl její plochy, zabíraný každým z objektů a hodnota objektu s největším podílem je pak buňce přiřazena (u bodů a linií se podíl plochy nahrazuje počtem a příp. délkou objektů, které buňka obsahuje).
- **Metoda nejdůležitějšího typu** buňce přiřadí hodnotu, která je považovaná za nejdůležitější z hlediska aplikace.
- **Metoda centroidu**, buňka má přiřazenou hodnotu definovanou polohou jejího středu při průmětu do vektorové reprezentace.



Dominantní typ



Nejdůležitější typ

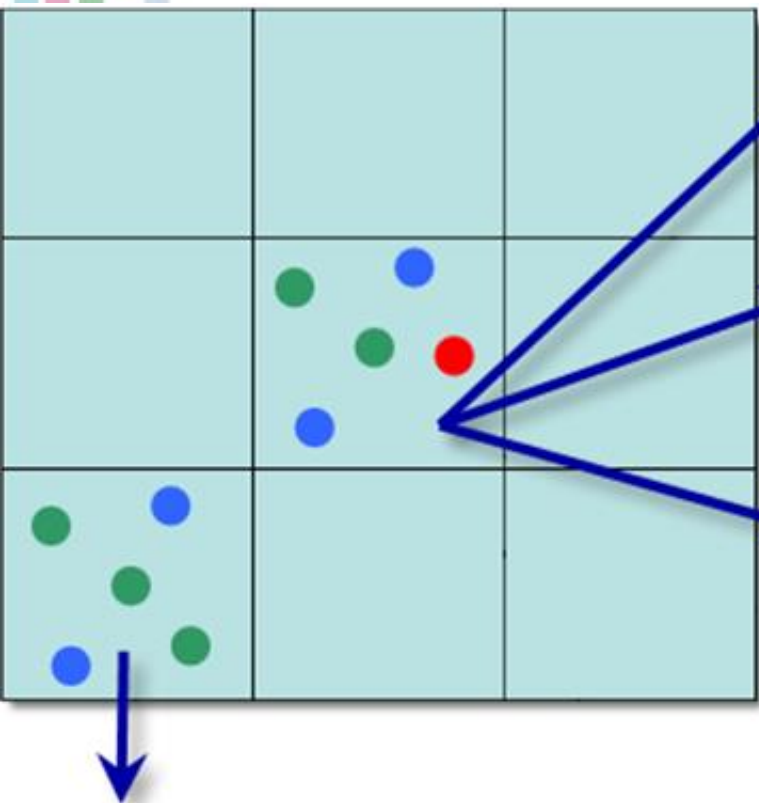


Centroidy

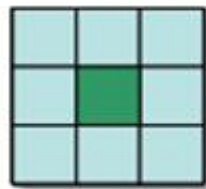




# Příklady – pravidla pro rasterizaci bodů v ArcGIS

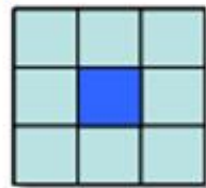


FID	Attribute
1	Green
2	Red
3	Blue
4	Blue
5	Green



Field = Attribute  
 Method = MOST\_FREQUENT  
 Priority = NONE  
 Outcome = Green  
 Reason = Lowest FID

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



Field = Attribute  
 Method = MOST\_FREQUENT  
 Priority = PriorityFID  
 Outcome = Blue  
 Reason = Highest priority

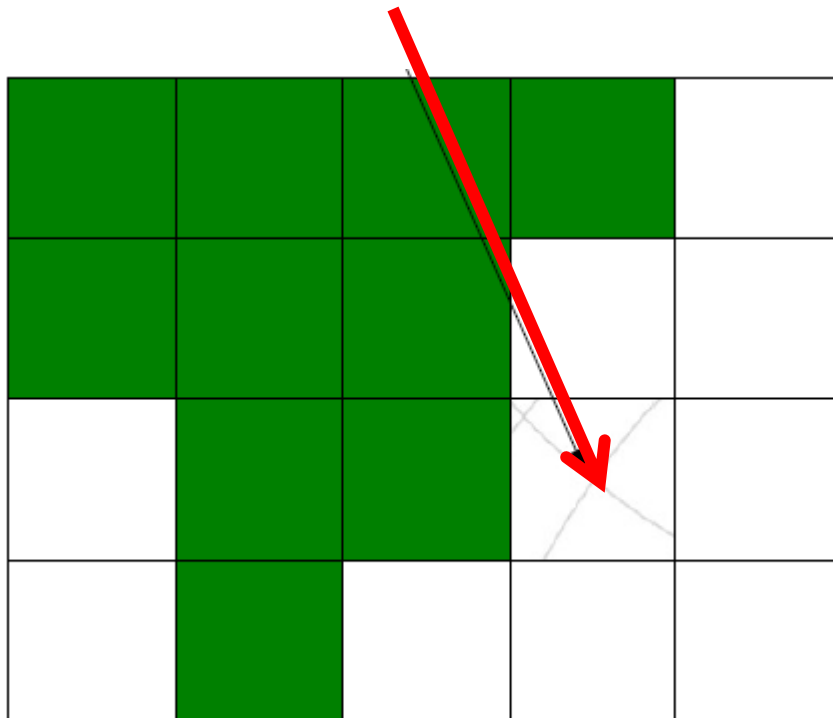
FID	ValueFID
1	1
2	8
3	5
4	3
5	2



Field = ValueFID  
 Method = STANDARD\_DEVIATION  
 Priority = Ignored  
 Outcome = 2.774887323379517  
 Reason = Priority field is only used with MOST\_FREQUENT

# Prázdné buňky

- Pokud je hodnota buňky definována jako prázdná (NoData), znamená to, že tato **buňka nese žádnou informaci o prostoru**, který reprezentuje.
- **0 je validní hodnota!**
- 999 obvykle použito pro No data



# Rastrová data

## Obrazová data

- Snímek dálkového průzkumu Země (DPZ)
- Ortofoto
- Skenované plány
- Dokumentace

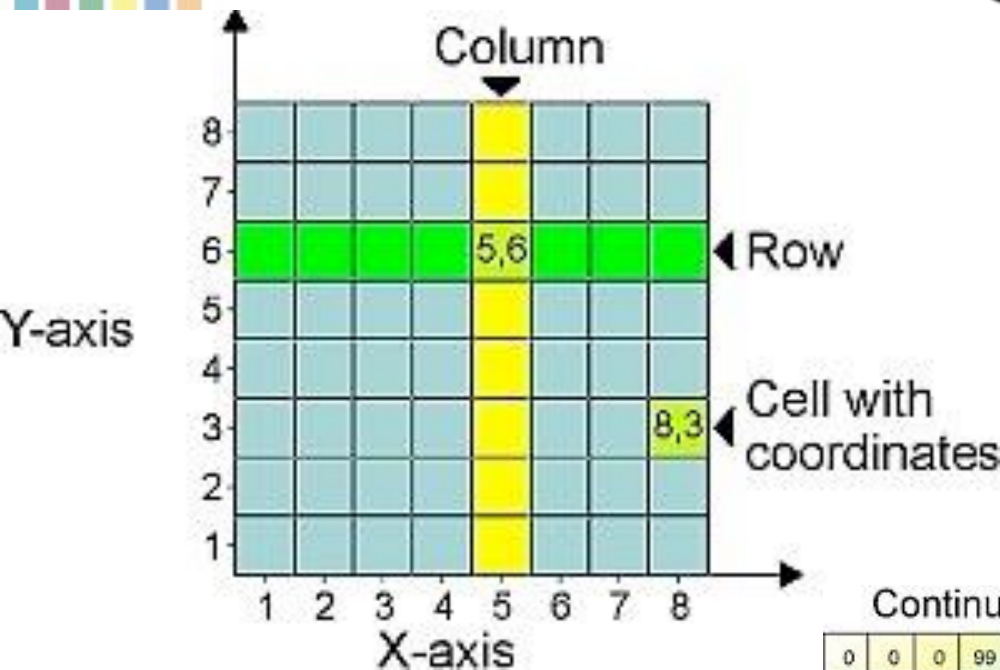
## Klasické rastry

- jednopásmová data
- znázorňují rozložení vždy jen **jednoho geografického jevu** (nadmořská výška-DMT , vodstvo , lesy, ...).
- mohou být získána např. převodem z vektorů nebo vyhodnocením obrazových dat.





# Čtvercová mřížka



Continuous grid

0	0	0	99	278	468	753	989
0	0	0	75	369	464	861	1075
0	0	55	255	498	752	932	1235
0	102	267	478	664	849	1307	1897
57	119	324	572	893	1483	1987	2335
291	425	766	954	1257	1537	1979	2476
392	482	877	1115	1592	2010	2639	3010
499	604	849	1045	1943	2338	2591	3203

Each cell is stored with its unique value e.g. bathymetric depth

Categorical grid

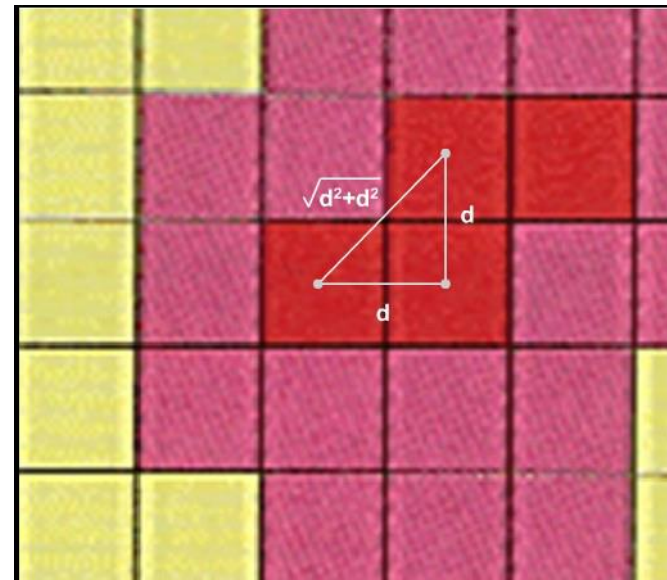
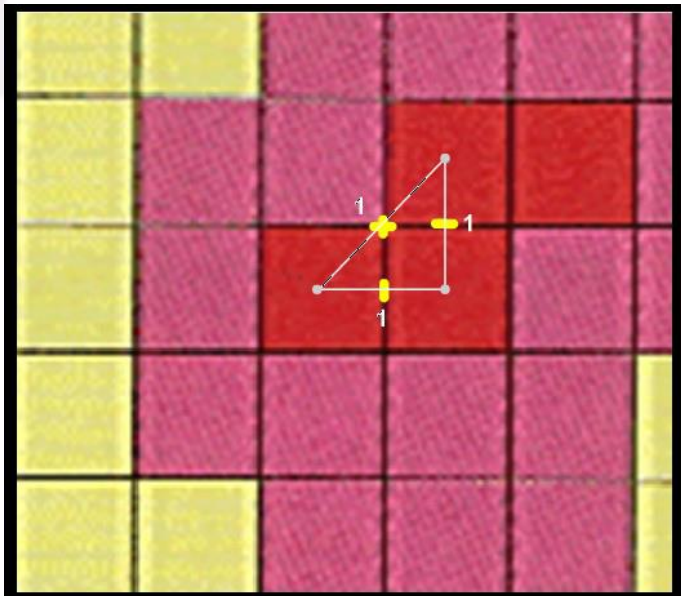
1	1	1	2	2	2	3	4
1	1	1	2	2	2	3	4
1	1	2	2	2	3	3	4
1	2	2	2	3	3	4	4
2	2	2	3	3	4	4	4
2	2	3	3	4	4	4	4
2	2	3	4	4	4	4	4
2	3	3	4	4	4	4	4

Zonal value	Count of cells	Category
1	9	Land
2	20	Shelf
3	13	Slope
4	22	Deep sea

# Metrika čtvercové mřížky

- V geometrii nastává problém metriky (způsob definice vzdálenosti dvou buněk) – odlišná vzdálenost středu čtverců.
- Euklidovská metrika

$$d = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$



# Rastrová data výhody a nevýhody

## výhody

- jednoduchost datové struktury
- snadné překrývání a kombinace obrazů s různým obsahem
- rychlé dotazování
- snadná tvorba uživatelských nastaveb
- jednoduchá kombinace s jinými daty rastrové povahy (DPZ)
- snadné provádění analytických operací

## nevýhody

- **značná paměťová náročnost (velký objem dat)**
- omezená přesnost, daná rozlišením rastru a orientací rastru (výpočty délek, vzdáleností, ploch ...)
- kvalita výstupů závislá na rozlišení rastru (nižší vizuální kvalita rastrových výstupů)
- nevhodnost pro síťové analýzy

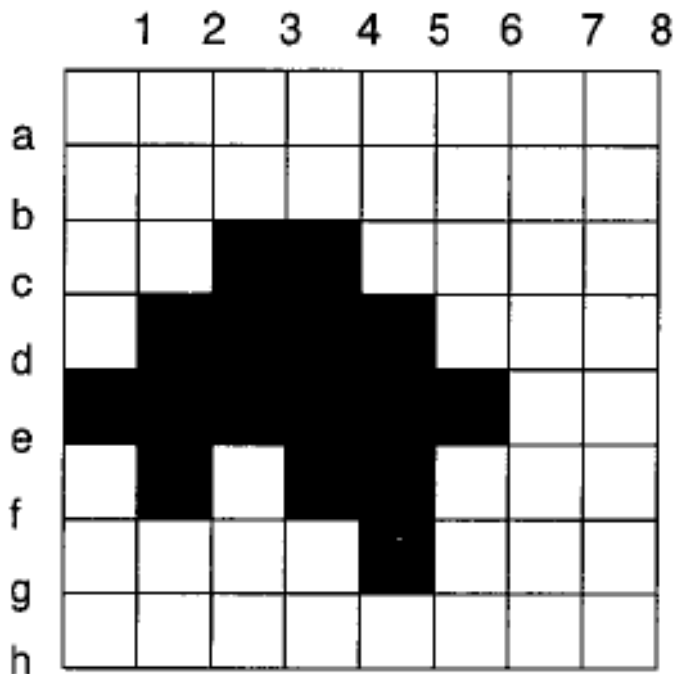


# Kompresní techniky pro rastry

- **Ztrátové**
  - komprimují lépe než neztrátové
  - dochází ke ztrátě informace => někdy nevhodné!
- **Neztrátové**
  - Run Length Codes – RLC
  - Run Length Encoding – RLE
  - Čtyřstrom – QuadTree
  - Adaptivní komprese

# Run Length Codes

- Definuje příslušnost buněk rastru k objektu po řádcích nebo sloupcích, přičemž udává jen **začátek a konec úseku buněk** v řádku či sloupci.



Řádek c 3,4  
Řádek d 2,5  
Řádek e 1,6

Řádek f 2,2 4,5  
Řádek g 5,5

# Run Length Encoding

- Využití maticového zápisu dat.
- Efektivní při rozsáhlých homogenních oblastech dat

**1 1 1 1 5 5 9 9 9 9 9 9 9 2 9 9 9**  
**(4 1)(2 5)(7 9)(1 2)(3 9)**

- Heterogenní ☹

**0 1 0 1 2 3 5 2 1 4**

**(1 0)(1 1)(1 0)(1 1)(1 2)(1 3)(1 5)(1 2)(1 1)(1 4)**

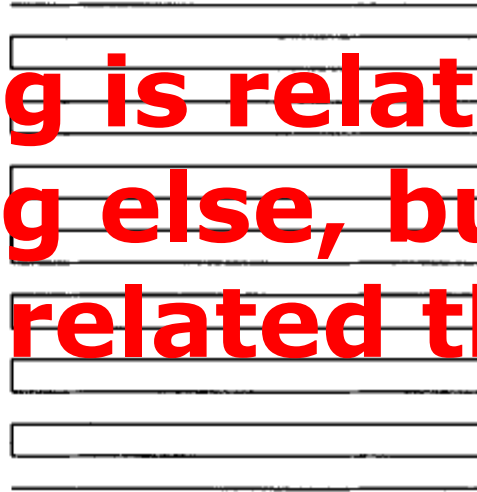
**Jak zefektivnit kompresi?**

# Způsob procházení rastru

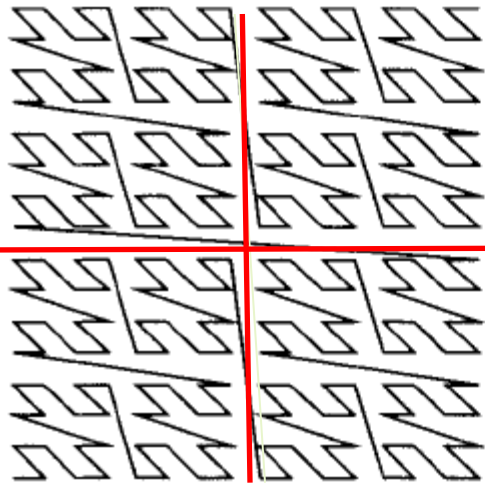
**Everything is related with everything else, but near things are more related than distant things.**



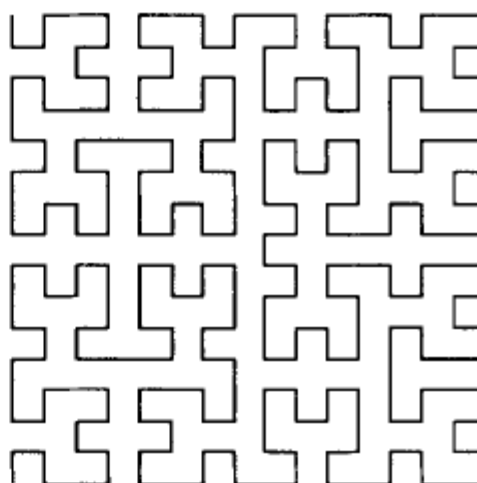
(a) Row Order



(b) Row-prime Order



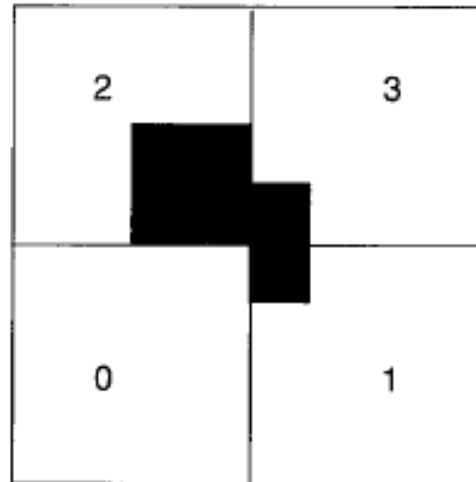
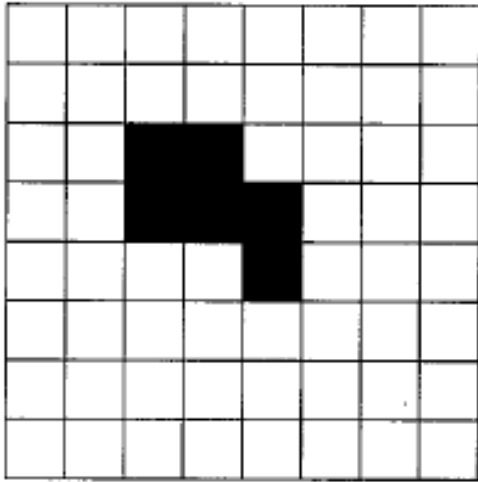
(c) Morton Order



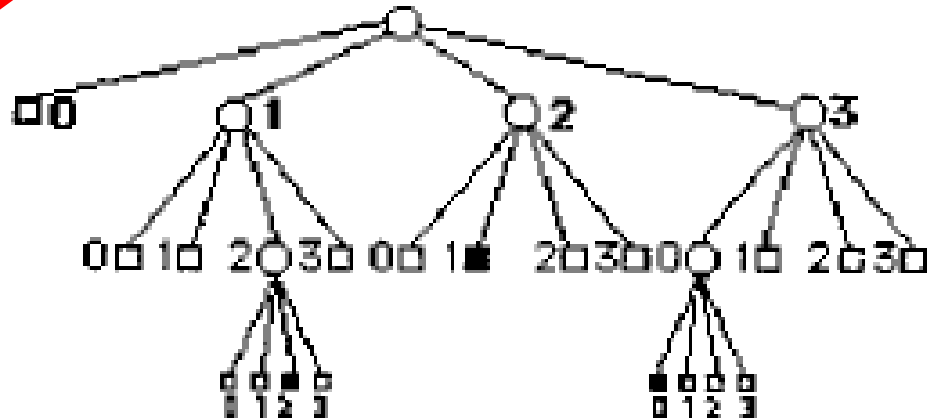
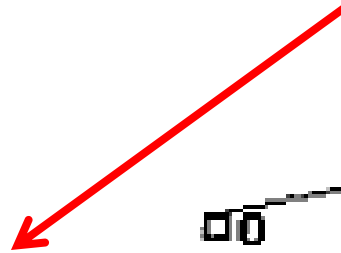
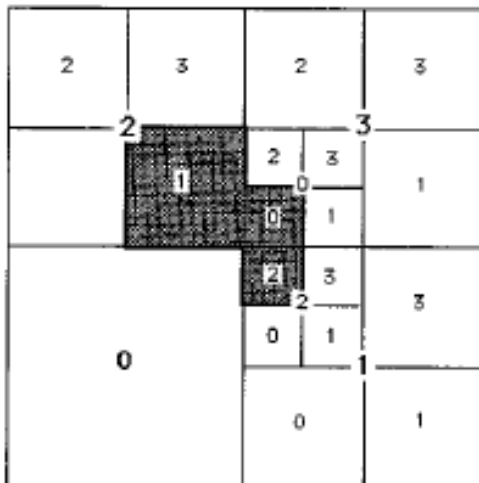
(d) Pi-Order

- A) a B) - alternativy postupného procházení.
- C) a D) - prostor vyplňující křivky (space-fill).
- B) a D) jsou více efektivní - souvislost s Toblerovým zákonem (First law of Geography).

# Quad tree - čtyřstrom



- Hierarchické uložení
- Dělení kvadrantů až do doby, kdy jsou homogenní.





# Adaptivní

- **Rozdělení dat do bloků využívajících metodu s nejvyšší účinností.**
- **Kombinace více metod v jedné datové sadě.**
- **Příklad – LZW(Lempel-Ziv-Welch) obecná komprese i pro neobrazové formáty.**
  - Princip spočívá v nahrazení vzorků vstupních dat binárními kódy proměnné (postupně rostoucí) délky.
  - Vstupní vzorky se překládají pomocí slovníku, který je průběžně doplňován o nové vzorky.
  - Délka slovníku je dána aktuálním počtem bitů použitých pro kódování.
  - Slovník přitom není zapisován do výstupních dat.

# Hlavička rastru

- NCOLS xxx
- NROWS xxx
- XLLCORNER xxx
- YLLCORNER xxx
- CELLSIZE xxx
- NODATA\_VALUE xxx
- row 1
- row 2
- .
- .
- row n



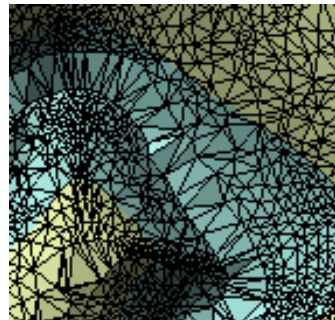
# Příklady rastrových formátů

Formát	přípona	lokalizační soubor	nekomprimovaný	komprese	
				neztrátová	ztrátová
	tif	tfw	X	X	
	<b>jpg</b>	jgw / jpw / hgr			X
	<b>png</b>	pgw		X	
	gif	gfw		X	
	bmp	bpw	X		
	cit	přímo uvnitř		X	
	img	?	?		
	MrSID	?		X	
	DjVu	?		X	



# Nepravidelná trojúhelníková síť

- Nepravidelné rastrové reprezentace - problémy s tvorbou, analýzou i uložením – prakticky se nepoužívají.
- Výjimkou je Nepravidelná trojúhelníková síť (Triangulated Irregular Network)
- Reprezentuje povrch jako soubor trojúhelníků (**trojúhelníková**), které jsou definovány třemi body umístěnými kdekoliv v prostoru (**nepravidelná**) a pro tyto trojúhelníky uchovává topologické vztahy (**síť**).
- Často se používá pro reprezentaci povrchů, například digitálního modelu reliéfu – DMR.



# Reprezentace TIN

## Triangulated Irregular Network (TIN)

seznam trojúhelníků

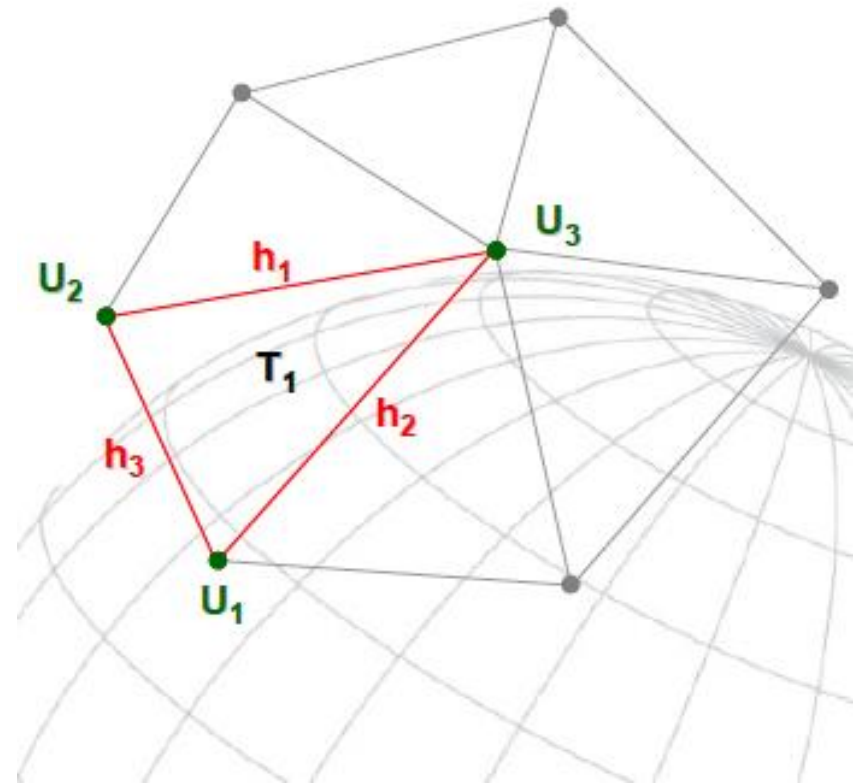
[  $T_1, (h_1, h_2, h_3),$   
 $T_2, ( \dots ),$   
 $T_3, ( \dots ), \dots ]$

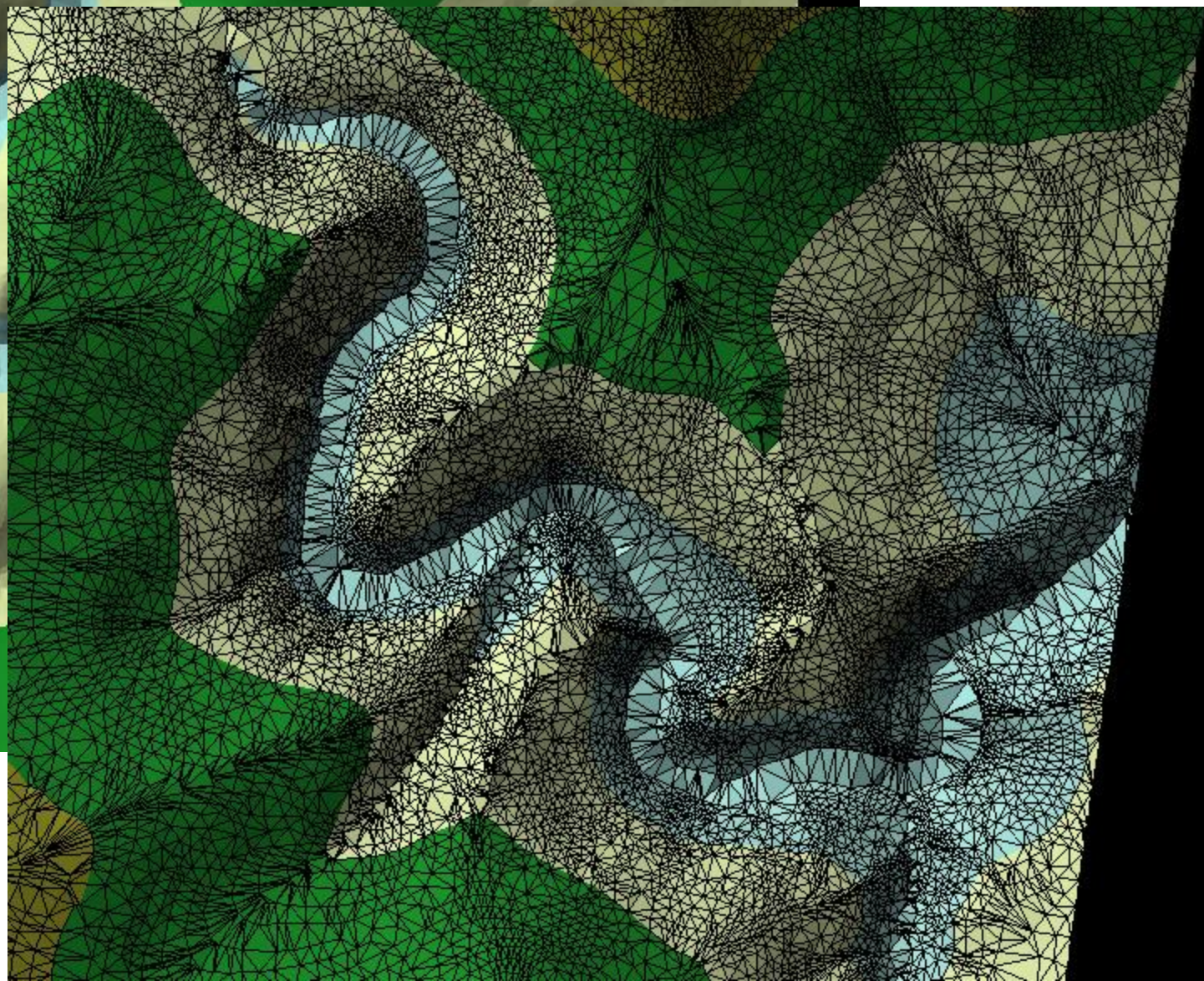
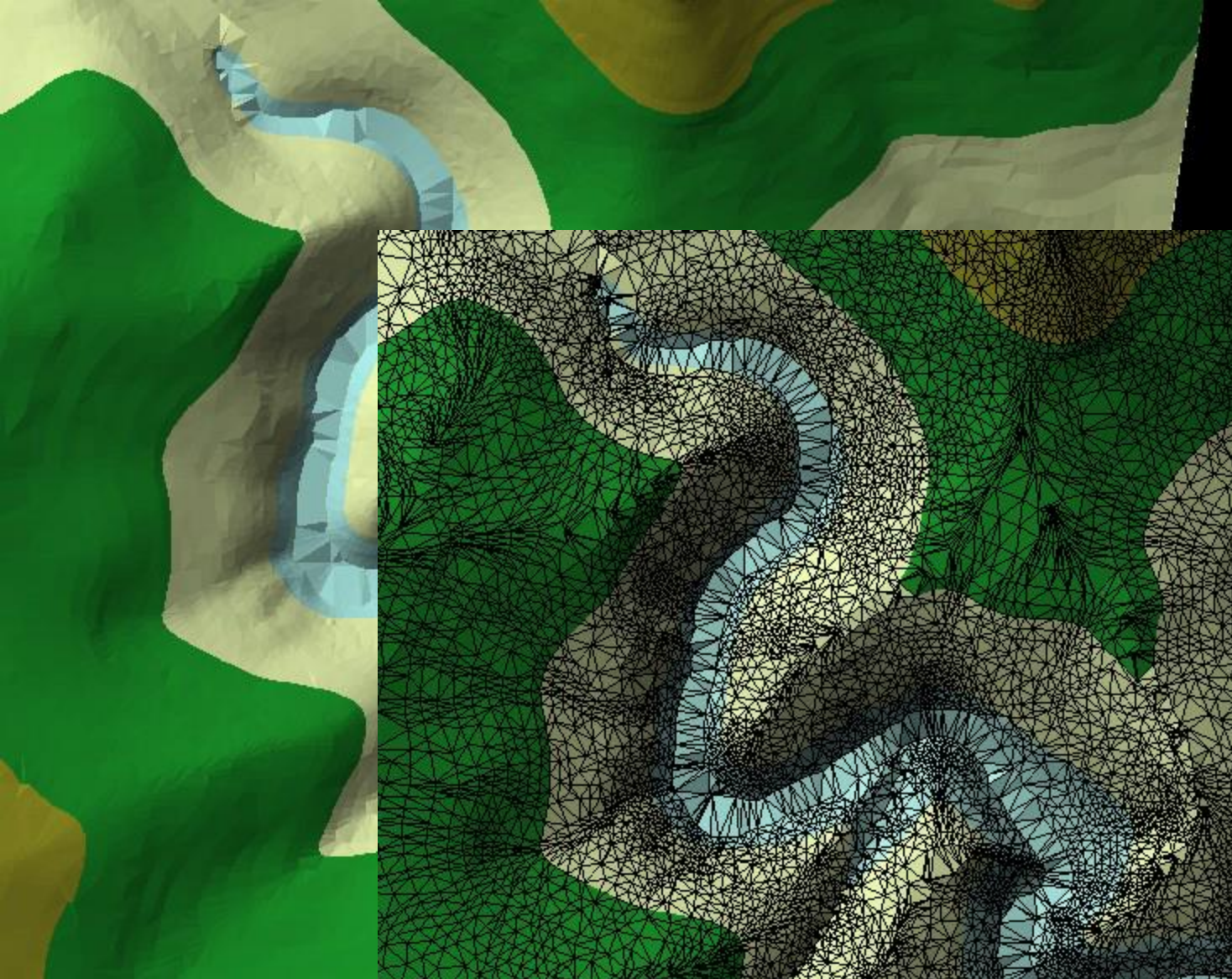
seznam hran

[  $h_1, ( U_1, U_2),$   
 $h_2, ( \dots ),$   
 $h_3, ( \dots ), \dots ]$

seznam uzlů

[  $U_1, ( X_1, Y_1, Z_1),$   
 $U_2, ( \dots ),$   
 $U_3, ( \dots ), \dots ]$





# TIN - porovnání s rastry

+

- **zmenšení objemu uložených údajů při reprezentaci nehomogenních povrchů,**
- **větší přesnost a věrnost pro nehomogenní povrchy**
- **struktura automaticky obsahuje informace o sklonu a směru tohoto sklonu.**
- **kompatibilita s moderními grafickými kartami .**

-

- **složitost datové struktury a tím i algoritmů s ní pracujících.**



# Datové modely - shrnutí

## Vektorová data

- geometrie prostorových objektů je vyjádřena za použití geometrických elementů;
- základními geometrickými elementy jsou: bod, linie, polygon;
- je možné pracovat s jednotlivými objekty jako se samostatnými celky;
- atributy prostorových objektů jsou připojeny pomocí tabulky;
- vztah mezi prostorovými objekty je zajištěný pomocí topologie;

## Rastrová data

- rovinný prostor je rozdělen pravidelnou mřížkou na jednotlivé dílky, zvané buňky (pixely);
- poloha pixelu je dána jeho souřadnicemi (umístění v rastru);
- každý pixel má v sobě jedinou hodnotu atributu;
- prostorové vztahy mezi objekty jsou implicitně obsaženy v rastru.