




RESTRUKTURALIZACE GEODAT

KVALITA DAT

Mapové zdroje, 29. 9. 2008

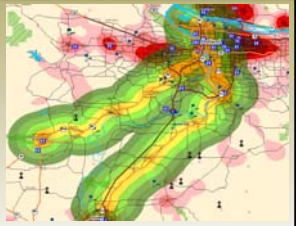

RNDr. Tomáš Řezník, Ph.D.



Restrukturalizace dat

- systémové konverze geodat
 - konverze datových formátů
 - převod datových reprezentací
 - převod typu geometrie
 -
- prostorové operace
- reklasifikace geodat
- kartografické transformace
-

≡ veškeré přípravné práce nezbytné pro vytvoření modelu, provádění analýz a rozhodovacích úloh

Konverze datových formátů geodat

Získal jsem data, jsou v nevhodném formátu, datové reprezentaci, mají špatný typ geometrie, jsou v jiném souřadnicovém systému...

Snaha o převod dat do formy vyhovující použití v projektu se zachováním maxima informace.

shapefile, CAD soubory, coverage, geodatabase, tabulková data, výstupy z databází, rastrová data v nejrůznějších formátech, ...

... to vše (a často ještě více) je nutné dostat do GIS ...

převod speciálně vektorových dat často složitý, pracný, ztrátový...

jednoúčelové nástroje: obvykle pouze pro transformaci do firemního proprietárního software, opačně nikoliv...

univerzální nástroje: GDAL, OGR, FME Suite, ArcGIS Interoperability Extension ...




Konverze datových formátů geodat

různých vektorových i rastrových formátů jsou doslova stovky, od nepsaných standardů až po naprosto exotické – bohužel toto je i případ různých formátů vyvinutých pro použití ve státní správě a rozpočtových organizacích v extrémním případě nutnost vytvoření vlastního konverzního software



FME OGR GDAL

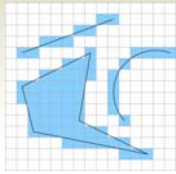


Převod datových reprezentací


Základní typy datových reprezentací v GIS jsou v současné době rastrový a vektorový datový model.

Základní úlohou je vzájemný převod mezi oběma datovými modely.

vektor → rastr
(rasterizace)



- jednodušší úloha, jednoznačná
- tvar vektorové geometrie se aproximuje buňkami o zvolené hodnotě umístěné v pravidelné mřížce
- různé prvky mohou být odlišeny zvolenou hodnotou pixelu odpovídající vybranému atributu
- oblasti bez dat reprezentovány smlouvenou hodnotou (0,-1) nebo logickou nulou (Null)

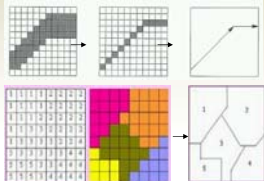


Převod datových reprezentací

Základní typy datových reprezentací v GIS jsou v současné době rastrový a vektorový datový model.

Základní úlohou je vzájemný převod mezi oběma datovými modely.

rastr → vektor
(vektorizace)

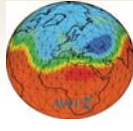
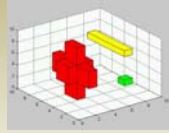


- složitější algoritmy, často nejednoznačná
- složitá detekce hran, uzlů
- použitelný výsledek pouze s kvalitními vstupními daty
- použitelné například pro výsledky klasifikace, různých rastrových analýz
- omezeně použitelné pro skenované podklady
- automatická, poloautomatická, ruční

Převod datových reprezentací

Další méně obvyklé úlohy:

- TIN -> rastr, vektor
- voxel (= 3D rastr) -> 3D vektor, 2D rastr
- převod rastrů s nepravoúhlým gridem
- převody orientovaných sítí,
-



Vektorová data - převod typu geometrie

Základní typy: bod, linie, polygon

+ mnoho odvozených typů

přesná terminologie a výčet typů geometrie závisí na konkrétním GIS softwaru



to / from	point	line	network	polygon	triangulation	surface
point	sampling	smoothing/centering	Delaunay triangulation	buffer operation	Voronoi diagram	horset smoothing
line	node extraction	representative point	line simplification	buffer operation	Voronoi diagram	
network			line graph	buffer operation	Voronoi diagram	
polygon	vertex extraction	representative point	boundary extraction	buffer operation	Voronoi diagram	
triangulation			boundary extraction	buffer operation	Voronoi diagram	
surface	critical point	critical lines	critical lines	centering	triangulation by surface value	

Vektorová data - převod typu geometrie

obecné charakteristiky:

- bod - 0D, žádná měřitelná délka a plocha
- linie - 1D, měřitelná délka, žádná plocha
- polygon - 2D, délka i plocha
- povrch - 3D, plocha včetně třetího rozměru

↑ jednodušší
↓ složitější, obvykle nutná
doplňující data

- bod -> linie ... nutné mít atributy, které určí příslušnost bodu k linii, pořadí
- linie -> polygon ... buď vysoce sofistikovaný algoritmus nebo další vrstvu s centroidy ze kterých lze potom převzít také atributy



polygon -> povrch ... TIN, interpolace, atd...

Vektorová data - generalizace

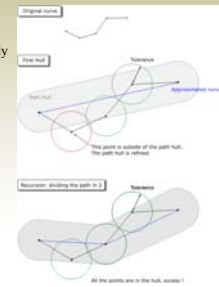
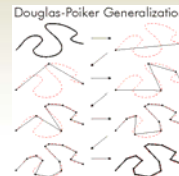
bohužel, obvykle řešíme přesně opačný problém - data nemají potřebnou přesnost

výhody generalizace:

- zmenšení datového objemu
- rychlejší výpočty složitých prostorových úloh
- prezentace v menších měřítkách

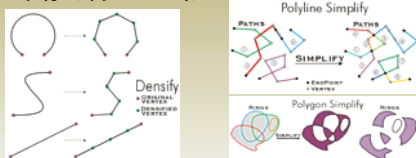
nevýhody generalizace

- v datových sadách s pokročilými topologickými pravidly možnost porušení těchto pravidel
- v případě špatné metody nevhovující výsledek

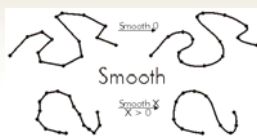


Vektorová data - zhuštění bodů, vyhlazení

V některých případech (nevhovující SW, různé typy prostorových úloh) je nutné nahradit spojitě křivé lomenými čarami, dále například doplnit lomové body v průsečících liniích, polygonů, spojit líniové úseky, které nekonečí v uzlu, provést rozpad polygonů na elementy,



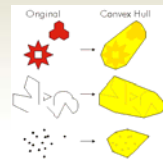
Pro kartografické účely, potřeby prezentace se může hodit opačný postup:



Vektorová data - konvexní obálka, ohrada

Pro různé úlohy může být potřeba řešit prostorové vztahy geometrií, u složitých geometrií, při vytváření prostorových indexů je výhodnější v prvním kroku vyloučit zcela nepravděpodobné případy, kdy dva prvky nemohou mít žádný prostorový vztah, porovnají se pouze obálky (ohrada/bounding box nebo konvexní obal).

Další využití konvexních obalů - jiný způsob generalizace.



Topologické překrytí, geoprocessing, ...

Podle složky geodat ve které dochází ke změnám

- mění se pouze geometrická data
- mění se pouze popisná data
- dochází ke změnám v obou složkách

Podle počtu vrstev, které vstupují do úlohy

- jedna vrstva
- dvě nebo více vrstev

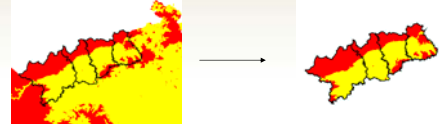
ořiznutí, překrytí, sloučení, rozdíl, průnik, ...

Ořiznutí (Clip, Crop)



Na základě určené vrstvy obsahující hranice zájmového území se ořízne jiná vektorová vrstva.
 Pf.: Zájmové území je definováno 4 okresy, zajímají mne pouze data ze zájmového území. Vrstvou definující ořez musí být geometrie typu polygon, ořezávaná vrstva může být jakéhokoliv typu, ořiznout lze i rastr.
 Plochy ležící uvnitř zájmového území zůstanou beze změny, plochy zcela mimo zmizí, u ploch na hranici ořiznutí se změní geometrie, doplní o lomové body průsečíků s hranicí a lomové body hranice. Atributy se nemění.

Výhody: úspora ve velikosti dat, kompaktní tvar zpracovávané oblasti, konzistentní data pro porovnávání ploch, ...



Spojení (Merge, Append)

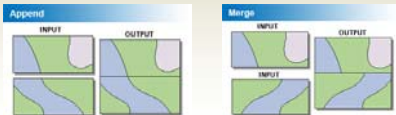
Spojení několika datových vrstev do jedné

Př.: Mám k dispozici vektorová data rozdělená po okresech, potřebuji zpracovat úlohu z území, které leží v několika okresech.

Spojovat lze obvykle pouze vrstvy obsahující geometrii stejného typu, záleží na možnostech software.

Mírně se liší funkce Append a Merge, Append obvykle dokáže spojit přilehlé linie a polygony do jednoho geoprvcu na základě vybraného atributu. Hodnoty atributů pole se převezmou např. podle největšího spojovaného fragmentu nebo nějakou z agregačních funkcí.

Výhody: konzistentní data pro celou zpracovávanou oblast.



Rozpuštění (Dissolve)



Sloučení několika prvků do jednoho na základě shodných hodnot některého z atributů.
 = rozpuštění **vnitřních** hranic, ponechání **vnějších**

Př.: Mám k dispozici mapu hydrologických povodí se členěním na oblasti IV. řádu, potřebuji pracovat s hydrologickými povodími vyšších řádů.

Spojovat lze obvykle pouze vrstvy obsahující geometrii stejného typu, záleží na možnostech software.

Mírně se liší funkce Append a Merge, Append obvykle dokáže spojit navazující linie a polygony do jednoho geoprvcu na základě vybraného atributu.

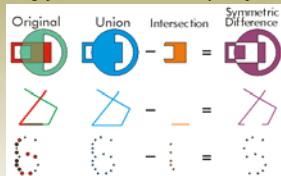
Atributová pole se přebírají podle vybraného prvku nebo se použije např. součet, průměr...

Výhody: konzistentní data, jednodušší pro celou zpracovávanou oblast.



Průnik, rozdíl, spojení (Intersection, Difference, Union)

- funguje na základě množinových operátorů



- některé operace jsou symetrické (nezáleží na pořadí vrstev), jiné ne



Rozdělení (Split)



- hodí se například pro distribuci dat, jejich rozdělení na správní oblasti, po mapových listech...
- často provádíme pro snížení výpočetních nároků, rozdělení práce mezi členy týmu, zajištění důvěrnosti dat (každý zpracovává dílčí část), ...

Prostorová spojení (Union, Identity, Intersection)

geometricky může jít o

- průnik vrstev
- sloučení vrstev
- doplnění jedné vrstvy druhou

atributy mohou být převzaty z

- první z vrstev
- kombinace obou vrstev

SPATIAL JOIN (merging feature attributes)

UNION
Overlay polygons and keep all areas from both coverages

Input coverage	Union coverage	Output coverage

IDENTITY
Overlay points, lines, or polygons on polygons and keep all input coverage features

Input coverage	Identity coverage	Output coverage

INTERSECT
Overlay points, lines, or polygons on polygons but keep only those portions of the input coverage features falling within the overlay coverage features

Input coverage	Intersect coverage	Output coverage

Prostorové spojení – manipule s atributy

Spojení jak geometrie tak atributů dvou původně nezávislých datových vrstev

Př.: Mám k dispozici vrstvu obsahující jednotlivé půdní typy včetně průměrné hodnoty pH, vrstvu s klasifikací vegetačního pokryvu, zajímají mne všechny listnaté lesy, které rostou na silně kyselých půdách

Spojovat lze obvykle pouze vrstvy obsahující geometrii stejného typu, záleží na možnostech softwaru.

Výhody: Společná analýza území na základě dvou původně nezávislých faktorů.

IDENTITY COVERAGE

Ov=1
Ia="A" Ia="B"
Ov=1 Ov=1
Ia="A" Ia="B"
Ov=0 Ov=0

INPUT COVERAGE + **IDENTITY COVERAGE** = **OUTPUT COVERAGE**

Transformace mezi souřadnicovými systémy

Souřadnicový systém (≠ kartografické zobrazení)

- 1: systém, určený údaji o referenční ploše, orientaci sítě na ní, jejím měřítku, referenčním bodu a užitím kartografickém zobrazení
- 2: sada matematických pravidel pro specifikování způsobu, jakým jsou souřadnice přiřazovány k bodům (ČSN ISO 19111)

dva základní typy:

- sférické (WGS-84, GCS-80, ...) + desítky dalších ...
- rovinné (S-JTSK, S-42, UTM, ...) + desítky až stovky dalších ...

Definice nejpoužívanějších souřadnicových systémů v ČR v software ESRI:

- S-JTSK S-JTSK_Krovak_East_North
- S-42 Pulkovo_1942_GK_Zone_3
- UTM WGS_1984_UTM_Zone_33N
- WGS-84 WGS_1984

Transformace mezi souřadnicovými systémy

Ve většině současných GIS lze kombinovat data z různých souřadnicových systémů, přesnost transformace závisí na kvalitě použitých transformačních vztahů a kvalitě souřadnicového systému (odpovídající době vzniku)

S-JTSK je souřadnicový systém starý téměř století využívající měření v triangulační síti staré téměř jedno a půl století! – velké lokální deformace, v závislosti na použité metodě transformace odchylky až několik metrů!

U moderních souřadnicových systémů odchylky max. v řádu decimetrů.

Transformace mezi souřadnicovými systémy

Transformace mezi sférickými systémy přímo na základě prostorové transformace (např. 3D sedmiprvková Helmertova transformace)

Transformace z rovinného systému na sférický a naopak přechodem na referenční plochu a pak viz výše.

Transformace mezi rovinnými systémy přechodem přes referenční plochy.

Transformace mezi souřadnicovými systémy

Při současném zobrazení dat z různých souřadnicových systémů se interně tyto transformace provádějí. Pro většinu běžných úloh přesnost dostává, pro přesnější práce je nutné data nejprve převést do společného souřadnicového systému přesnější metodou, provést topologickou kontrolu, teprve potom pracovat.

Explicitní datové konverze jsou nutné při předávání dat v rámci týmů nebo spolupracujících organizací, při spojování dat získaných v různých souřadnicových systémech atd...

Chyba při transformaci [cm] z WGS84 do S-JTSK i při použití jedné z nejhodnějších metod transformace.



REKLASIFIKACE RASTROVÝCH DAT

Manipulace s rastrovými geodaty

Většina systémových konverzí je shodná nebo analogická jako u vektorových dat.

- změna formátu uložení dat
- změny rozlišení – převzorkování
- změna datové hloubky
- u vektorových dat analogie například generalizace
- rastrová algebra
 - plní podobnou funkci jako topologická překrytí u vektorových dat
- reklasifikace
 - podobné principy jako u vektorových dat, bude probráno společně
- filtrování rastrových dat
 - specifické pro rastrová data
- zvýrazňování obsahu, metody vizualizace
 - v některých ohledech shodné s vektorovými daty, v některých odlišné

Převzorkování

Při práci s několika rastry v různých souřadnicových systémech nebo v různém rozlišení je nutné interně nebo na vyžádání provést převzorkování dat na stejné rozlišení ve stejném gridu, aby je bylo možné použít pro společnou analýzu. Měnit rozlišení lze oběma směry – zvyšovat i snižovat.

- zvýšením rozlišení převzorkováním **nezískáme** více informací o území, pouze zvýšíme objem dat
- snížením rozlišení dokonce část informace o území **ztratíme**.

Několik různých metod pro převzorkování, výběr vhodné metody závisí na charakteru dat a dalším plánovaném postupu. Stejně tak je vhodné zvážit ve které etapě projektu převzorkování provedeme (klíčové analýzy je vhodnější provádět na originálních datech)

Základní metody:

- metoda nejbližšího souseda (nearest neighbour)
- bilineární
- bikubická

Převzorkování

porovnání různých metod převzorkování

metoda nejbližšího souseda je nejjednodušší, nejrychlejší a je vhodná pro rastrová data diskrétního typu – přesně ohraničené oblasti s buňkami stejného typu

Převzorkování

Převzorkování diskrétních rastrových dat jinou metodou než pomocí nejbližšího souseda by bylo chybné – zanikly by původně jednoznačné hranice mezi plochami různých tříd

Převzorkování spojitých dat metodou nejbližšího souseda není vyloženě chybné, ale v případě velkých změn ve velikosti rozlišení vstupních a převzorkovaných dat by výsledek nebyl optimální. Proto se hodnota každého elementu počítá jako vážený průměr s přihlédnutím k okolí každé elementární plochy => bilineární nebo bikubická interpolace

příklad spojitých dat:

Převzorkování

bilineární interpolace – (bilinear interpolation)

nejbližší body v okolí

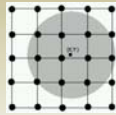
bikubická interpolace – (cubic convolution)

nejbližší body v okolí + jejich okolí

Přezorkování

další metody – složitější, výpočetně pomalejší, pro specializované využití:

- cubic spline
 - přímá
 - FFT
- radiální funkce
- Gaussovská i. s použitím FFT
-

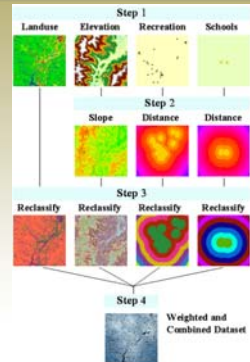


Reklasifikace

Reklasifikační funkce mění hodnoty buněk na alternativní za použití řady metod.

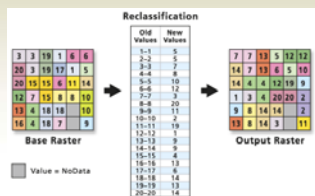
- záměna konkrétních hodnot
- klasifikaci hodnot do tříd
- změna hodnot funkčním vztahem

Všechny reklasifikační metody zpracovávají na základě jednotných pravidel každý pixel celé plochy rastru.



Reklasifikace - Záměna konkrétních hodnot

- přifažení na základě reklasifikační tabulky
- vytvoření tabulky obvykle ručně
- jedné hodnotě vstupního rastru přiřadíme **právě jednu hodnotu** výsledného rastru
- tabulku lze uložit pro případné další použití



Reklasifikace - Klasifikaci hodnot do tříd

- přifažení na základě reklasifikační tabulky
- vytvoření tabulky obvykle ručně nebo poloautomaticky s využitím základních statistických metod (automatické rozdělení do intervalů na základě četnosti, rozptylu, ...)
- tabulku lze uložit pro případné další použití pouze pokud bude mít další překlasifikovány rastru obdobné rozdělení četností hodnot
- jedné hodnotě ve výsledném rastru odpovídá **jedna nebo více** hodnot vstupního rastru



Reklasifikace - Změna hodnot funkčním vztahem

- předchozí dva typy reklasifikace vyžadují specializovaný nástroj, který pracuje s reklasifikační tabulkou, reklasifikace změnou hodnoty funkčním vztahem je záležitost nástroje pro rastrovou algebru
- stejně použití jako v přednášce o rastrové algebře, s jediným rozdílem že je zde pouze jeden vstupní rastr

- [vystup] = 10*[vstup]
- [vystup] = [vstup]*[vstup]
- [vystup] = sin([vstup])
- [vystup] = abs([vstup]-avg([vstup]))
-



Změna datové hloubky

Rastrová data mohou obsahovat jedno pásmo, případně více pásem analogie:

- černobílý snímek ~ jedno pásmo
- barevný snímek ~ tři pásma
- multispektrální snímek

V každé buňce může být uložena hodnota v různé číselné reprezentaci o různé přesnosti

- celočíslné datové typy - diskrétní hodnoty, kategorie, ...
 - INTEGER (obvykle 32 bitů)
 - LONG (obvykle 64 bitů)

- datové typy s plovoucí desetinnou tečkou - spojité hodnoty, výšky, teploty, ...
 - FLOAT (REAL) (obvykle 32 bitů)
 - DOUBLE (obvykle 64 bitů)

Nelze bez ztráty informace vzájemně převádět.



Každý typ datové reprezentace má své jasné využití.

Filtrování rastrových dat

Doposud probírané metody pracovaly vždy s hodnotou konkrétní elementární plochy, respektive s hodnotami elementárních ploch různých datových vrstev se stejným prostorovým umístěním.

Filtrace obrazu = operace s digitálním obrazem, které slouží ke **zvýraznění** určité informace.

- průměrování
- vyhlazení obrazu
- potlačení šumu
- zvýraznění kontrastu
- detekce hran
- postklasifikační zpracování obrazu
-

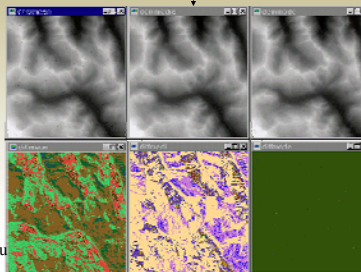
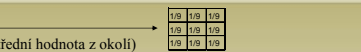


Vzhledem k rozsahu většiny digitálních dat je z technického hlediska nevhodné řešit podobné úlohy najednou v celém obrazu. Daný filtr je tak definován jako šablona rastrové matice (tzv. "moving window", v české literatuře se často využívá termín "kernel") - tedy pohybující se (plošoucí) okno. Má obvykle rozměry (px) 3x3, 5x5, 7x7, ...

Filtrování rastrových dat - vyhlazení dat (low pass filtry)

- průměr
- medián (střední hodnota z okolí)
- mód (nejčastější hodnota z okolí)
- ...

průměr,
medián,
mód



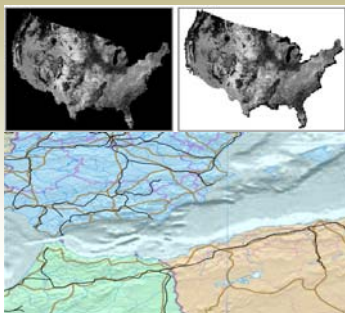
rozdíl původního
a
filtrováného rastrového

Vizualizace a zvýraznění rastrových dat

Aniž bychom měnili obsah dat, můžeme jednoduchými prostředky změnit jejich vzhled a usnadnit si tak práci. Lze potom snáze identifikovat jevy na snímku nebo rastrová data lépe prezentovat.

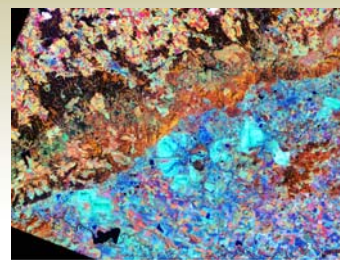
Úpravy se mohou týkat:

- výběru barevné škály
- roztažení histogramu
- => zvýšení/snížení kontrastu a jasu
- nastavení průhlednosti
- nastavení transparentního okolí
- ...



Vizualizace a zvýraznění rastrových dat - multispektrální

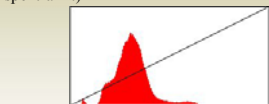
U multispektrálních rastrových dat (obsahují více než jeden barevný kanál) musíme zvolit pro každý ze 3 (vybraných) kanálů (pásem) odpovídající základní barevnou složku (R G B). Multispektrální snímek je potom zobrazen v nepřevných barvách. Volba kanálů a přiřazení některého ze 3 základních barevných kanálů se řídí obecnými zvyklostmi, spektrálními charakteristikami pásem nebo požadavkem na zvýrazněný jev.



Vizualizace a zvýraznění rastrových dat - monochromatická data

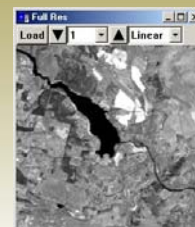
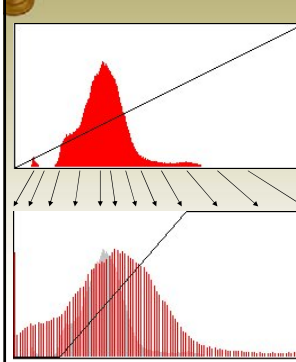
Obsahují pouze jeden barevný kanál. V některých rysech podobné reklasifikaci, na rozdíl od reklasifikace nedochází k faktickým změnám v rastrových datech, mění se jen způsob zobrazení dat, data zůstávají stále stejná.

(Stejný postup zvýraznění lze použít nezávisle na každém barevném kanále a zvýraznit tak i data multispektrální.)



Histogram udává četnost každé hodnoty v rastrové matici. Je vidět že není dostatečně využito celého dynamického rozsahu zvolené barevné škály.

Vizualizace a zvýraznění rastrových dat - monochromatická data



Vizualizace a zvýraznění rastrových dat

- neroztažený histogram
- lineárně roztažený histogram
- vyrovnaný histogram
- nelineárně roztažený histogram
-

Pyramidy

vytvoření pyramid

- pro běžné zobrazování v řádově menším rozlišení se použije pyramidový náhled
- náhledy jsou zapouzdřeny uvnitř rastrových dat
- objem dat se zvětší o třetinu ($1 + 1/4 + 1/16 + 1/64 + \dots$)
- metoda vytváření náhledů:
 - nejbližší soused
 - bilineární
 - bikubická
 - ...

3 levels of a pyramid

Prostorové operace s rastry – spojování (Merge)

Plně se překryjí pouze oblasti s daty, oblasti bez dat (hodnota Null, NoData) jsou brány jako transparentní. Záleží na pořadí překrytí.

Prostorové operace s rastry – ořez (Clip, Crop)

- zcela shodné s vektorovými daty
- v případě pravouhelníkové oblasti rovnoběžně se souř. s. prostý ořez
- v případě nepravidelné oblasti se rozdíl mezi ohradou a polygonem pro ořez vyplní hodnotou NoData (Null)

Prostorové operace s rastry – změny prostorového umístění

- v odůvodněných případech možné použít
- obvykle není při kvalitně provedeném georeferencování rastru není potřeba

Original data
Affine
Second order polynomial
Third order polynomial

KVALITA DAT

Kvalita dat

- nepodložená očekávání – „digitální data jsou kvalitnější než analogová“
- ne vždy, pouze intuitivní spojení kvality dat s pokročilejší technologií jejich uložení a zpracování
- digitální systémy pouze umožňují přesnější zpracování, uložení a kontrolu
- jsou limitovány úrovní kvalifikace obsluhy, kvalitou zdrojových dat
- digitální data musí být použita pouze pro předpokládaný účel pokud nemá uživatel zkušenosti a znalosti s posouzením vhodnosti pro účely jiné

Přesnost – čím vyšší přesnost dat, tím větší důvěra k nim

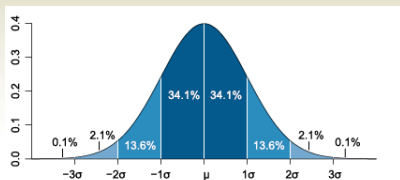
- míra pravděpodobnosti s jakou lze charakterizovaný číselný údaj považovat za správný
- polohová přesnost
- atributová přesnost

Kvalita dat – polohová přesnost

- pravděpodobnost že poloha bodu určená jeho polohovými daty je „skutečná“
- **odchylka** - rozdíl mezi udávanou a skutečnou hodnotou
- **rozptyl odchylek** – chyby jednotlivých naměřených hodnot
- **směrodatná odchylka** (= střední chyba) – jednotný ukazatel chyb
- prakticky každý soubor geodat je zatížen chybou (výjimky: např. klady mapových listů, souřadnicová síť, ...) => můžeme hovořit o přesnosti
- matematický model – stanovení předpovědi odchylky určité velikosti
 - zvolí se empiricky, na základě zkušenosti nebo statistickými metodami
 - nejčastěji **normální rozdělení** – u prostorových dat mnohonásobně ověřeno
 - hladina spolehlivosti – stanovení míry jistoty pro určité tvrzení „80% hodnot souboru má odchylku menší než...“
 - v praxi obvykle násobek (2x 2,5x 3x) směrodatné odchylky -> polohová přesnost geodatabáze

Kvalita dat – polohová přesnost

- dodržení předpokladů daného modelu (např. u normálního rozdělení)
 - rozdělení chyb je zcela náhodné
 - vybraný vzorek dat tvoří reprezentativní soubor
 - není zatíženo systematickou chybou
 - nevhodně provedená transformace souřadnicových systémů
 - část dat pořízená jinou technologií,
 - část dat zpracována jinou metodikou, ...
 - ...



Kvalita dat – polohová přesnost

- ověření polohové přesnosti
 - srovnání vzorku dat s přesnějšími daty (například nákupem vzorku geodat vyšší přesnosti nebo vyžádáním vzorku)
 - přímé měření v terénu (i řádově) přesnější metodou
 - mírou polohové přesnosti je univerzálně platná **střední kvadratická chyba (RMS – Root Mean Square)** = součet čtverců (druhých mocnin) odchylek vydělený počtem měření
- přesnost analogových podkladů je nepřímo úměrná mapovému měřítku „mapa 1 : 1 000 je přesnější než mapa v měřítku 1 : 10 000“
- u digitálních dat je hodnocení složitější – data jsou fakticky uložena v měřítku 1:1 a lze z nich vytvořit mapy jakéhokoliv měřítka (což neznamená že je to vhodné)
- použití dat neodpovídající kvality může stát v důsledku víc než pořízení kvalitnějších dat – zavádějící výsledky, špatná rozhodnutí, ...

Kvalita dat – atributová přesnost

- oproti geometrické (polohové) přesnosti bývá zanedbávána, přitom je stejně důležitá
- špatně definované popisné údaje mohou způsobit chyby při klasifikaci, tedy i při následné analýze
- atributy jsou vyjádřeny diskrétní nebo spojitou veličinou – diskrétní nabývají konečného počtu, typicky z číselníku, spojitě mohou nabývat hodnot jakýchkoliv
- metodika hodnocení přesnosti je víceméně shodná jako hodnocení polohové přesnosti
- určení přesnosti diskrétních atributů = ohodnocení přesnosti klasifikace objektivita ovlivněna
 - počtem tříd
 - typickým tvarem a velikostí oblastí
 - způsobem výběru testovacích ploch
 - vzájemnou podobností tříd

Kvalita dat – atributová přesnost

- př.: křoviny podél potoků jsou typicky dlouhé úzké polygony, zabírají i méně než 1% plochy, při náhodně vybraných souvislé oblasti testovacích dat se do výběru nedostanou
- velikost ploch jednotlivých tříd může kolidovat s prostorovou přesností
- ostré hranice v geodatech nemusí korespondovat s typicky nejednoznačnými rozhraními (les x louka, jehličnatý x listnatý les) => velký podíl individuální interpretace při klasifikaci
- náhodné chyby – určení rozhraní, omyl v klasifikaci, omyl při vkládání dat...
- systematické chyby – chyba metodiky, přístroje (kyselost půdy, intenzita odraženého záření, ...)
- ohodnocení přesnosti atributů by mělo vyjadřovat obsah správně nebo nesprávně zadaných atr. dat, např. „95% všech atributů je správných“ – tento údaj často z neznalosti postupu nebo z důvodů dalších nákladů chybí

Kvalita dat – vnitřní provázanost

- udává kvalitu logických vztahů mezi jednotlivými objekty v geodatabázi
 - topologická kvalita dat
 - sousedící polygony mají společnou hranici
 - linie se spojují v uzlech
 - polygony jsou uzavřené
 - ⇒ kontroly, automatické opravy
 - provázanost prostorové a atributové složky
 - atributová data jsou vztažena ke správnému geom. objektu
 - př.: výsledky rozboru vody ve studni odpovídají konkrétní studni v mapě
 - při aktualizaci aktualizovat obě složky geodat najednou
- ⇒ důsledná kontrola například při vkládání dat z ručních vstupů – kontrolní součty, identifikační kódy...

Kvalita dat – vnitřní provázanost

- zdánlivě levná geodata (i pokud mají vyhovující polohovou a atributovou přesnost) mohou v důsledku špatné vnitřní provázanosti vyjít v konečném důsledku velmi drah
- neexistuje objektivní veličina pro posouzení vnitřní provázanosti, omezeně je možné použít různé kontroly topologie a testy na dodaném vzorku dat před jejich pořízením například použitím nástrojů, u nichž je předpoklad že budou v projektu použity (topologická překrytí, síťové analýzy, ...)

Kvalita dat – rozlišení

- rozlišení = prostorově nejmenší jednotka, ke které jsou vztaženy prostorové informace
 - u rastru velikost buňky (pixelu)
 - u vektoru maximální přesnost uložení souřadnic lomového bodu (nepřímě úměrně rozsahu zájmové oblasti)
- odpovídá pojmu nejmenší mapovaná jednotka u tematických map

Kvalita dat – shrnutí

- technická snadnost s jakou lze použít geodat v libovolném vztažném měřítku zvyšuje potřebu objektivního určení přesnosti a kvality dat
- digitalizaci geologické mapy 1 : 500 000 lze vytvořit digitální mapu ve vztažném měřítku 1 : 50 000, ta ovšem nebude mít takové kvality jaké bychom od mapy v takovém měřítku očekávali

- typicky:
 - „V jaké přesnosti chcete data dodat?“
 - „V té největší možné.“

Cena geodat a (klesající) měřítko mají obvykle kvadratickou (nebo dokonce exponenciální) funkční závislost.

Databázová úroveň dat používaných v GIS

- kvalitativní standardy na úrovni databáze se týkají datového souboru jako celku
- nelze je zjistit testováním vzorkem jejich hodnota se stanoví subjektivním posouzením nebo pomocí doprovodných údajů (metadata)
- **úplnost databáze** = stav kdy všechny objekty reálného světa dané třídy z daného zájmového území jsou obsaženy jako objekty v příslušné třídě geodatabáze
- **úplnost územního pokrytí** = jaká část polohových dat v geodatabázi se týká zájmového území
- **úplnost obsahového pokrytí** = vhodnost mapové legendy, posouzení, zda příslušný číselník pokrývá všechny objekty reálného světa

Databázová úroveň dat používaných v GIS

- problémy:
 - hierarchický strom není vyvážený

- př.:
- jehličnatý les
 - smrk
 - jedle
 - borovice
 - ostatní
 - listnatý les
 - smíšený les

- definice tříd se překrývají
- definice tříd z různých zdrojů jsou nejednoznačné, neporovnatelné
- některé třídy nejsou definovány – obvykle „sběrný koš“ nerozlišeno, ostatní, ...

Databázová úroveň dat používaných v GIS

• stáří dat

- časový faktor je pro řadu aplikací životně důležitý
- změna polohových a popisných dat stávajících objektů
- vznik nových a zánik stávajících objektů

- kvalitativní parametr, záleží na typech objektů – např. klimatické oblasti, hranice států se budou měnit podstatně pomaleji než například infrastruktura, výstavba
- zastaralé údaje způsobují potíže – nutnost dodatečných měření, často duplicitních, protože uživatelé stejných dat dělají paralelně totéž
- na stáří dat se často nebere zřetel – spojování různých starých dat z několika databází – problémy
- srovnávací analýzy – nutno zachovávat časovou jednotu pro celé území – raději starší, ale konzistentní data než novější a různě kombinovaná

Databázová úroveň dat používaných v GIS

• původ dat

- toto kritérium hodnotí historii databáze
 - zdroje dat
 - zpracovatelské postupy
- každý datový zdroj a každý postup vnáší do dat jisté chyby
- původ databáze rozhodující pro výběr dat pro danou aplikaci
- nejen z těchto důvodů je nutné trvat na co nejobsáhlejších metadatech
- všechny údaje by měly obsahovat ukazatele správnosti

• uživatelská úroveň

- cena dat – pro různé uživatele relativní
- specializovaná data obvykle dražší než univerzální (základna uživatelů, distribuce výrobních nákladů)
- přístupnost dat – utajení (z hlediska bezpečnosti státu, z hlediska konkurence)

Databázová úroveň dat používaných v GIS

• přímé x nepřímé náklady

- nákupní cena
- + čas, materiál vynaložený na práci s daty – prohlížení a seznamování s daty, převod formátů, reklasifikace, transformace, v extrémních případech i vývoj sw pro tyto úlohy

- **přístupnost** – bezpečnost, ochrana osobních údajů – RČ, příjmy, další citlivé údaje využitelné například pro marketing
 - umělá degradace atributových dat, například náhodným rozdělením jednotlivých identifikátorů, pronásobení náhodným koeficientem, ...
 - u geometrických dat warping, translace, rotace, ...

Zdroje chyb v datech GIS

- chyba je neodstranitelnou součástí výsledku každého měření
- velikost může být zásadní nebo zcela zanedbatelná
- určující faktor kvality

• chyby nezávislé na operacích GIS

- měření původních dat
 - nepřesnost přístroje
 - měření různými přístroji
- chyby vzniklé při vytváření převzatých databází
 - tvorba databáze (chyby výpočtů, chyby v geodetických základech, ...)
 - editace dat
- změny v krajině
- nedostatečné pokrytí území

Zdroje chyb v datech GIS

• chyby vznikající při práci v GIS

- chyby při pořizování dat
 - nepřesnost digitalizace (chyby zařízení, chyby operátora)
 - nepřesnost při ukládání atributů (chyba operátora, nedostatečná kontrola)
- chyby při ukládání dat
 - nedostatečná numerická přesnost formátu dat
 - chyby média, hardware
- chyby při manipulaci
 - konverze rastr <-> vektor
 - generalizace
 - spojování tříd
 - překrytí vrstev
 - interpolace
 - klasifikace satelitních snímků
 - ...
- chyby při prezentaci dat (v současnosti zanedbatelné – minimum výskytu)
 - nepřesnost kreslicího zařízení
 - nepřesnost vlivem změn papíru, fólie

Zdroje chyb v datech GIS

• chyby metodické

- chyba metodiky použité pro sběr dat (např. nevhodný počet vzorků)
- chybně definované objekty a třídy
- nesprávně použitá komprese dat
- neurčitě hraniční linie
- datově ztrátové operace v nevhodném pořadí degradující výsledek

• chyby převzatých dat

- generalizace, odsazení liniových tvarů (železnice, řeka, silnice, ...)
- kartografická reprezentace – silnice svou osou, plošný objekt bodovou značkou

Zdroje chyb v datech GIS

- **nesprávné použití výsledku**
 - nedodržení kvalitativních standardů
 - zanedbání omezení určeného přesnosti dat
 - nevhodné použití některé metody, nástroje
 - nekompetentní rozhodnutí na základě správného výsledku – špatná interpretace
- chyby v rozhodování jsou součástí celého procesu využití prostorové informace
- nutné provést objektivní rozbor, protože snáze se chyba přisoudí zařízení, metodice než například obsluze nebo zdrojovým datům

Sledování kvality dat v GIS

- stanovení kvality často problematické
- často pravidlo „zlaté střední cesty“
- vždy posoudit efekt navýšení nákladů oproti výhodám dat vyšší kvality
- některé metody kontroly dat příliš nákladné vzhledem k rozpočtu projektu nebo ceně kontrolovaných dat – kontrolu dat s menší důležitostí v terénu lze například nahradit interpretací leteckých fotografií, porovnáním se vzorkem dat pořízených nezávislým postupem, ...
- snažit se vyhnout tradičním argumentům
 - „takhle se data pořizovala vždycky“
 - „všechny třídy na mapě musejí mít stejnou přesnost“
- technicky lze hranici přesnosti stále posouvat, pragmaticky je ale omezena cenou pořízení
- čím jsou požadavky na přesnost a kontrolu vyšší, tím je získání dat časově náročnější, tím je obtížnější zajistit aktuálnost dat, ...