

Datové modely v GIS pro uložení prostorových dat

Uložení atributových dat

Datové modely v GIS pro ukládání prostorových dat

Minule - reprezentace prostoru v GIS

Dnes ukládání prostorových dat:

Špagetový model

Topologický model

Hierarchický model

Špagetový model

nejjednodušší

princip vychází z digitalizace map

objekt na mapě se reprezentuje jedním logickým záznamem v souboru a je definovaný jako řetězec x,y souřadnic, řetězec je seřazený

nevýhody - ačkoli jsou všechny objekty v prostoru definovány, struktura neposkytuje informace o vztazích mezi objekty

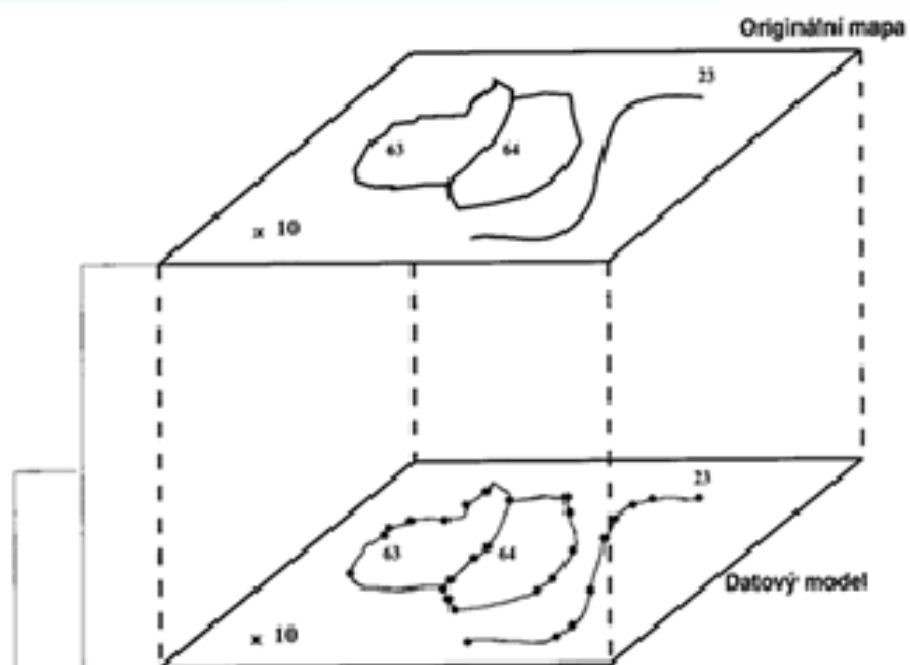
způsob uložení sousedících polygonů. Společná linie je totiž ukládána dvakrát, pro každý polygon zvlášť

Pro většinu prostorových analýz je tento model nevhodný, protože veškeré potřebné prostorové vztahy musí být spočítány před každou analýzou

název špagetový - soubor řetězců souřadnic, nemá žádnou logickou strukturu

neexistence prostorových vztahů činí tento model atraktivním pro použití v jednodušších CAC (Computer Assisted Cartography - počítačová kartografie) systémech, kde je díky jednoduchosti velice výkonný.

Špagetový datový model



Datová struktura

Objekt	Číslo	Poloha
Bod	10	X, Y Jednotlivý bod
Čára	23	$X_1, Y_1, X_2, Y_2, \dots, X_n, Y_n$ Řetězec
Polygon	63	$X_1, Y_n, X_2, Y_2, \dots, X_n, Y_1$ Uzatvřená smyčka
	64	$X_1, Y_1, X_2, Y_2, \dots, X_n, Y_n$

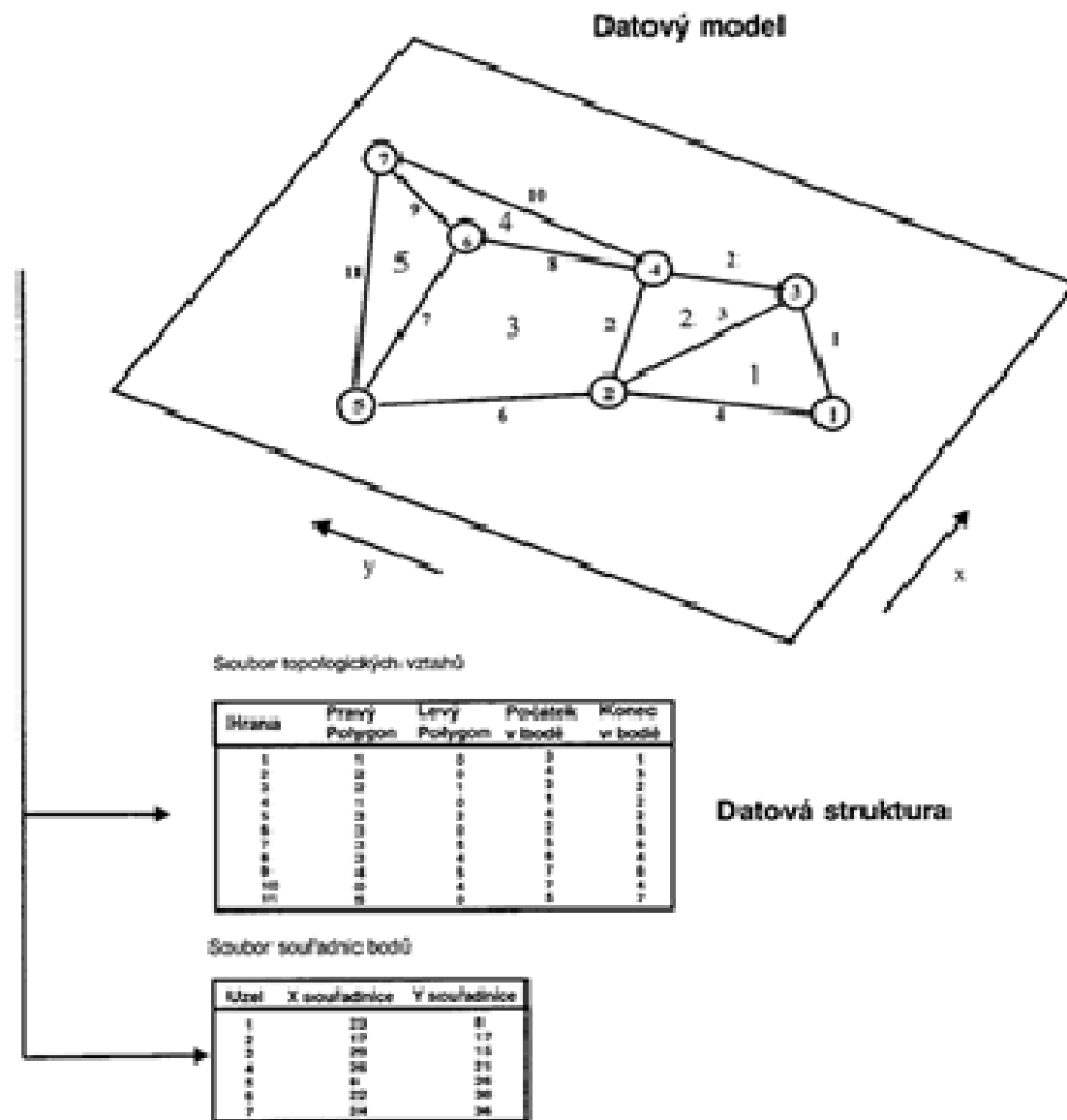
Topologický model

V tomto modelu každá linie začíná a končí v bodě nazývaném **uzel - node**. Dvě linie se mohou protínat opět jenom v uzlu. Každá část linie je uložena s odkazem na uzly a ty jsou uloženy jako soubor souřadnic x,y. Ve struktuře jsou ještě uloženy identifikátory označující pravý a levý polygon vzhledem k linii. Tímto způsobem jsou zachovány základní prostorové vztahy použitelné pro analýzy. Navíc tato topologická informace umožňuje body, linie a polygony uložit v neredundantní podobě.

Jak špagetový, tak topologický formát mají nevýhodu v neuspořádanosti jednotlivých záznamů. K vyhledání určitého liniového segmentu je třeba sekvenčně projít celý soubor. K vyhledání všech linií ohraničujících polygon je třeba tento soubor projít několikrát!

Existuje ještě i tzv. rozšířený topologický model. Jedná se o vylepšení hierarchického modelu (bude za okamžik) o **regiony (regions)** a **cesty (routes)**.

Topologický datový model



Hierarchický model

Tento model odstraňuje neefektivnost při vyhledávání v jednodušším topologickém modelu pomocí ukládání v logicky hierarchické podobě.

Vzhledem k tomu, že polygony se skládají z linií, které odpovídají jejich hranicím, a linie se skládají ze souboru bodů, jsou do modelu zahrnuty odkazy mezi jednotlivými druhy objektů (polygony, liniemi a body).

Tyto odkazy pak umožňují mnohem snadnější vyhledávání jednotlivých objektů než v případě topologického modelu. Hierarchický model obvykle také obsahuje topologickou informaci.

Příkladem hierarchického modelu může být datová struktura **arc-node**, kterou mimo jiné využívá v malé modifikaci i geografický informační systém ArcGIS.

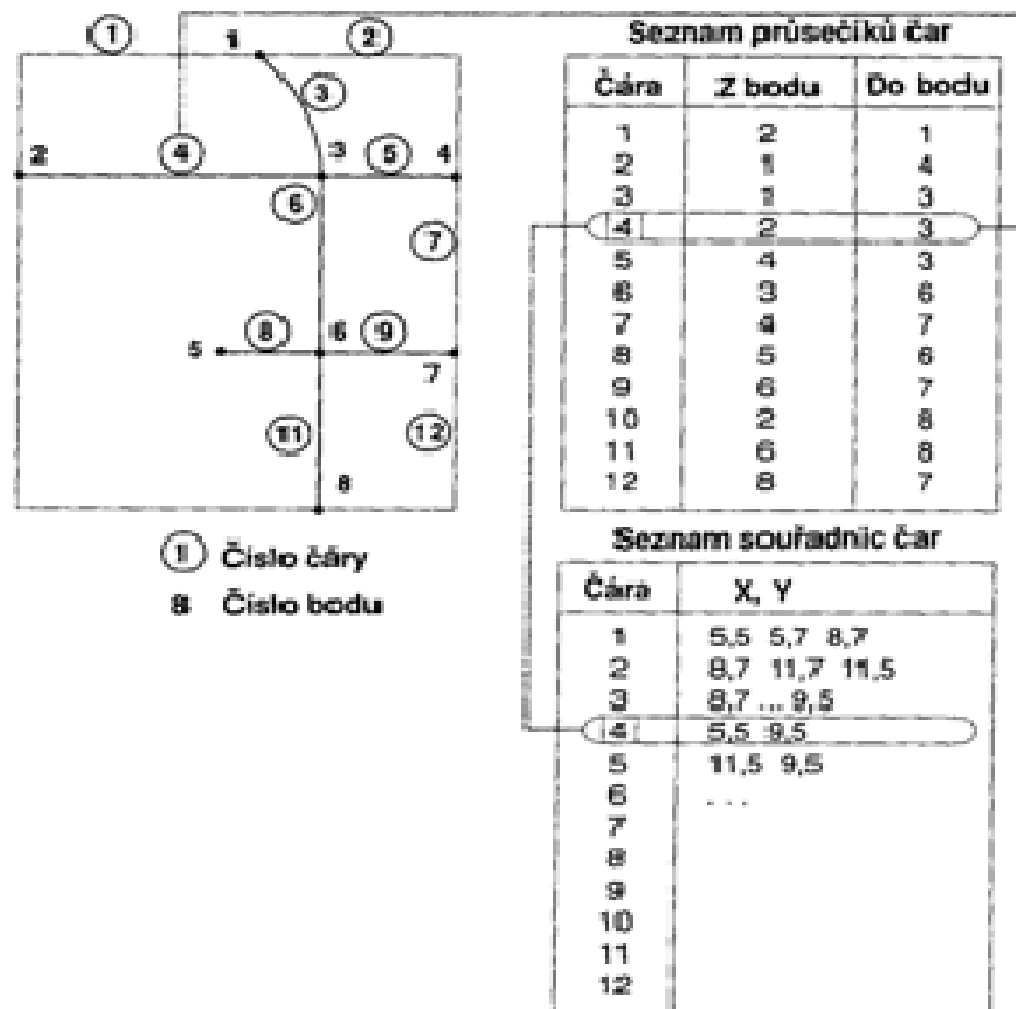
roli stavebního kamene zde, stejně jako v případě topologického modelu, hrají **body**.

Řetězce linií jsou složeny z jednotlivých **linií**, které jsou definovány jako seřazená množina **vrcholů-vertices/vertexes** o souřadnicích **x,y**.

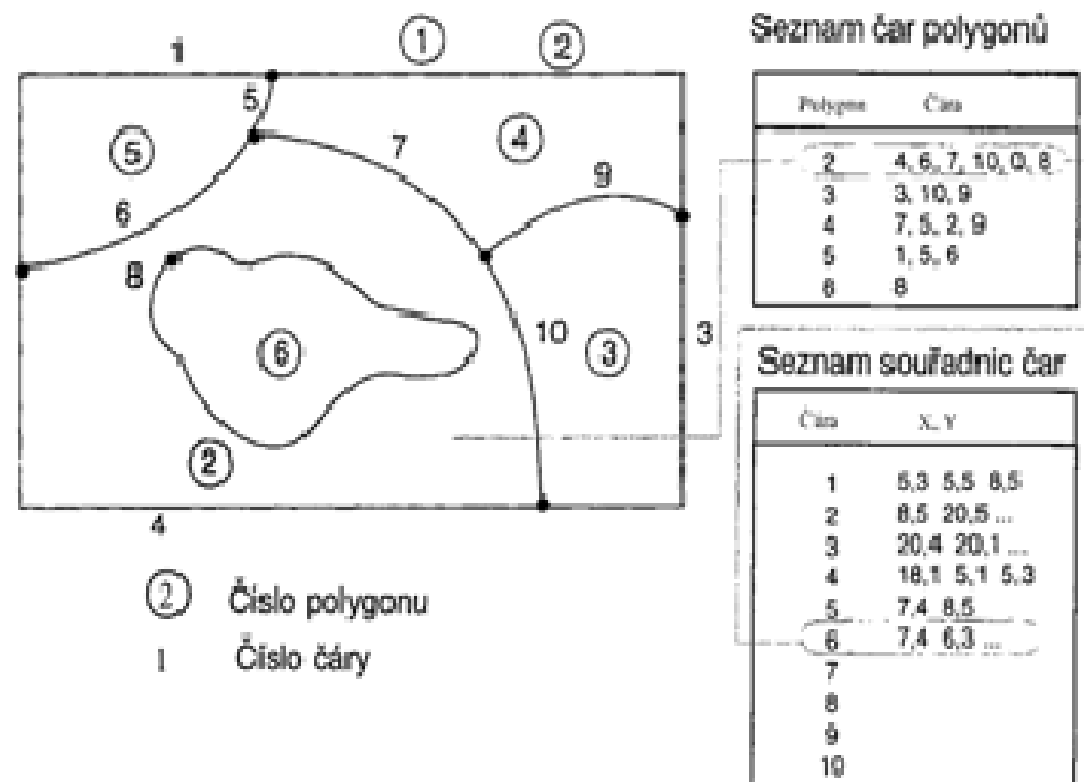
Důležitou roli pak hrají **uzly (nodes)**, což jsou body vyskytující se na konci, na začátku a v místech, kde se jednotlivé **linie** protínají. **Polygony** jsou definovány jako plochy ohraničené pomocí **řetězců linií**.

Hierarchický vektorový model - výhody oproti topologickému modelu především při vyhledávání a manipulaci. Rozdělení polygonů, linií a bodů do různých souborů (nebo tabulek) umožní při vyhledávání použít pouze část datových struktur a tím urychluje práci.

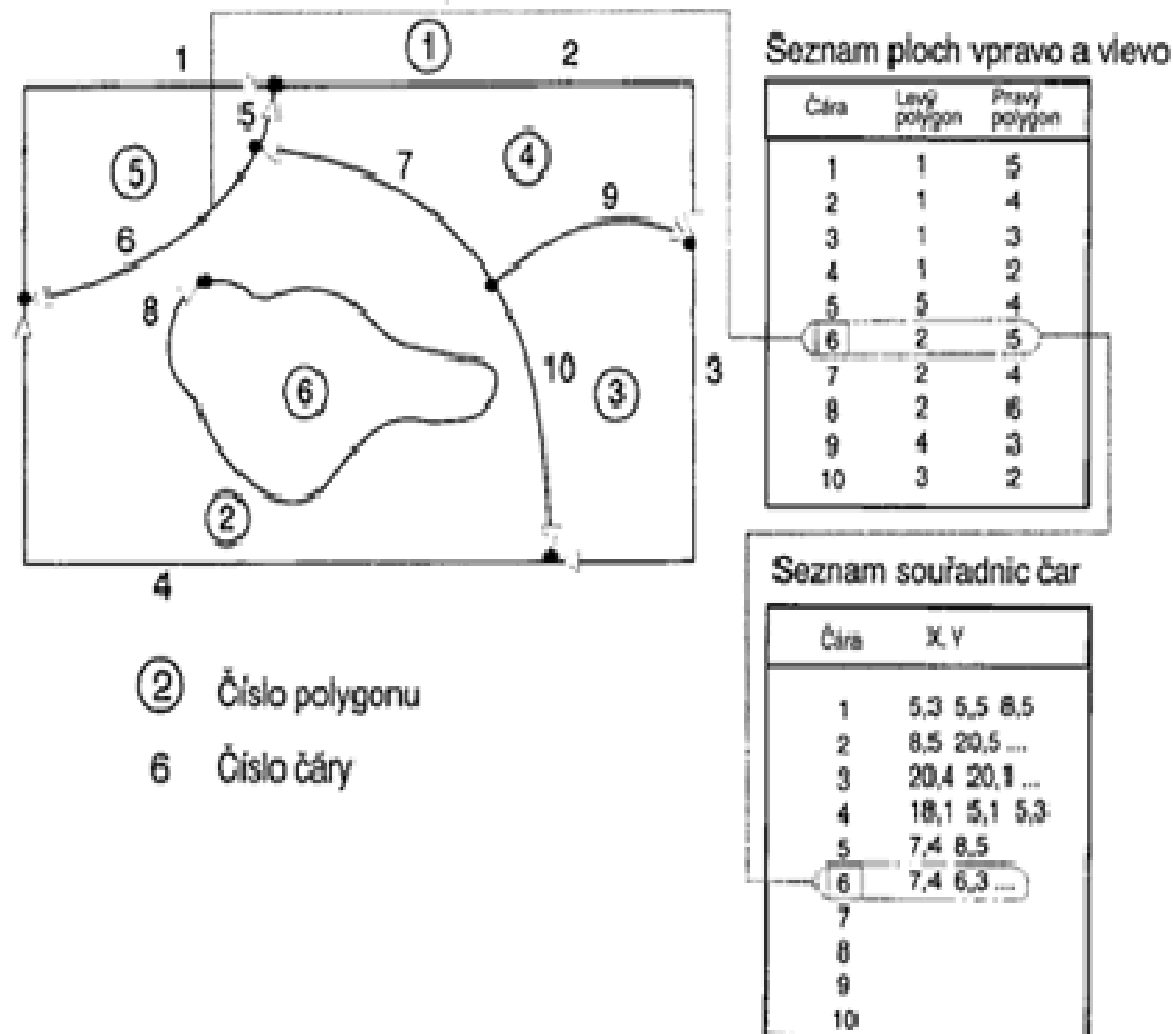
Hierarchický datový model - způsob uložení linie



Hierarchický datový model - způsob uložení polygonu



Hierarchický datový model - definování principů sousednosti



Rastrová data – čtvercová mřížka

Definice pro dvourozměrný prostor (2D): Čtvercová mozaika (mřížka) je v souřadnicovém systému jednoznačně definována souřadnicemi počátečního bodu, velikostí buňky a počtem buněk ve směru X a Y

V **geometrii** potřebujeme metriku:

Metrika je způsob definice vzdálenosti dvou buněk. V GIS jsou používány nejčastěji následující typy metrik:

Euklidovská metrika - znáte

Bloková metrika - vzdálenost středů dvou buněk je definována jako minimální počet překonaných hran (diagonální vzdálenost, tj. vzdálenost buněk sousedících rohem, je 2).

Šachovnicová metrika - vzdálenost středů dvou buněk je definována jako minimální počet překonaných hran i rohů (diagonální vzdálenost je 1)

Použití dvourozměrných matic je výhodné, jelikož jsou:

dobře definovatelné matematickým aparátem,

jednoduše implementovatelné jako základní datový typ většinou programovacích jazyků (2D pole),
obecně použitelné, jelikož na každý pixel může být použita jakákoli definovaná operace.

Nevýhody použití matic pro reprezentaci rastrových dat jsou:

detail - celý obraz je reprezentován stejným způsobem, tudíž uložení velké oblasti, skládající se z pixelů stejného typu, není nijak optimalizováno,

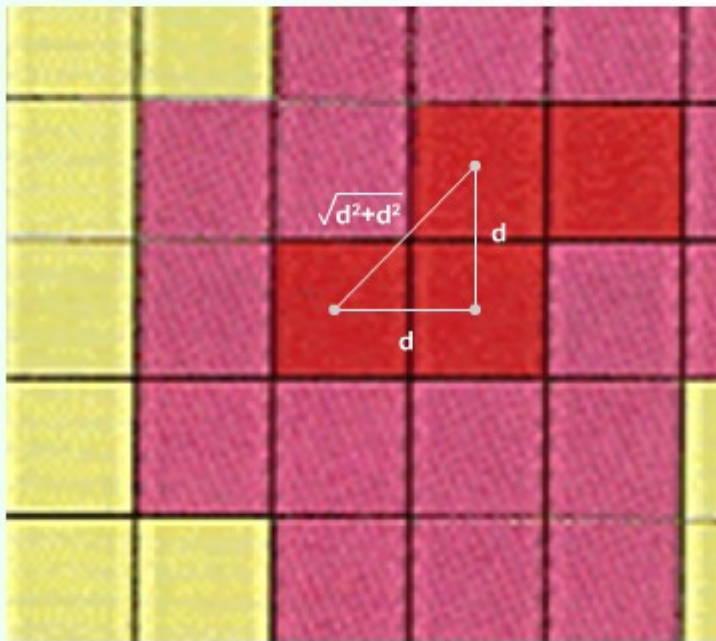
abstrakce - k získání méně detailního obrazu je nutné použít všechna data,

náročnost na úložný prostor - zbytečné plýtvání místem,

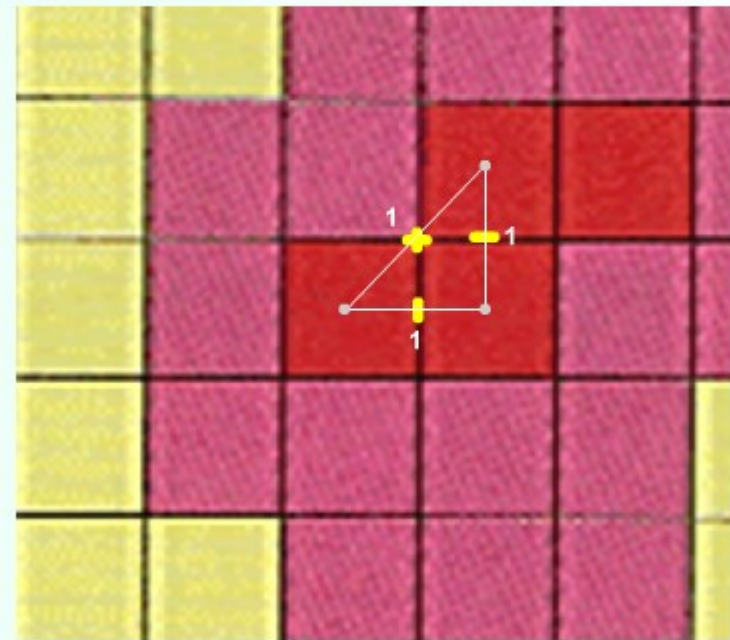
plýtvání pamětí - celý datový soubor musí být v paměti, což může být rozhodující u velkých obrazů.

Pravidelná čtvercová mřížka – nejpoužívanější rastrová reprezentace

Metriky používané pro pravidelnou čtvercovou mřížku

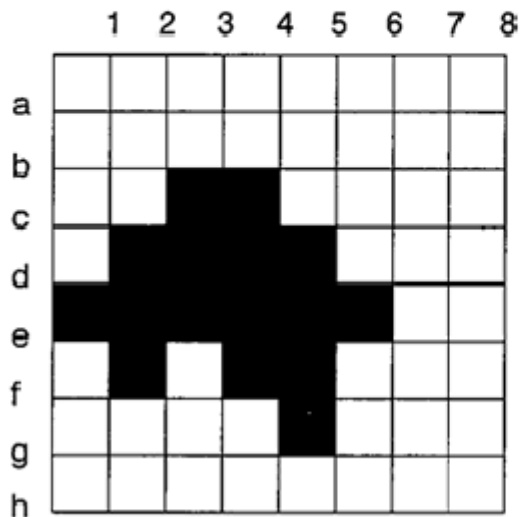


Metriky používané pro pravidelnou čtvercovou mřížku



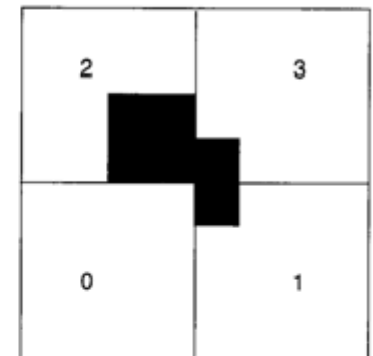
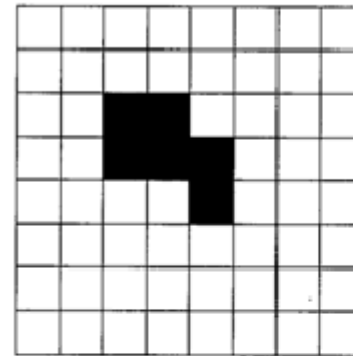
Kompresní techniky používané pro ukládání pravidelných rastrů

Kódování úseků řádků (Run Length Codes - RLC)



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

Čtyřstrom (QuadTree)



Komprese rastrových dat

Metoda délkových kódů (Run Length Encoding – RLE)

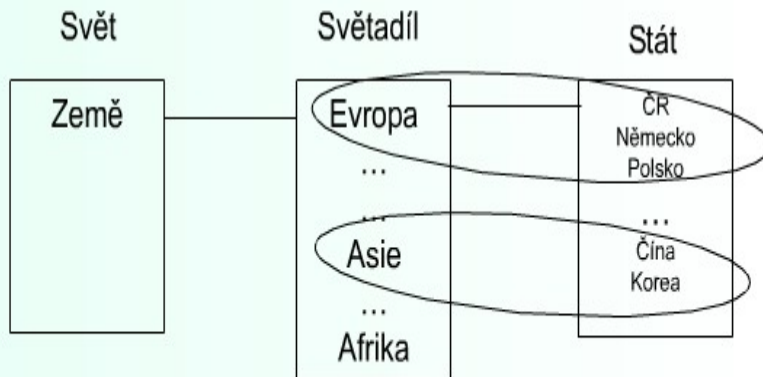
Kódování úseků řádků (Run Length Codes – RLC)

Čtyřstrom (QuadTree)

Adaptivní komprese

Datové modely pro uložení atributových dat na počítači

Hierarchický datový model



Hierarchický datový model

Tento model organizuje data ve stromové struktuře, tj. jeden rodič má hodně dětí, vztahy 1:1 a 1:N.

výhoda - umožňuje rychlé vyhledávání

nevýhoda - je vhodný pouze pro aplikace se stabilní strukturou, tam kde se primární vztahy mezi daty mění jen velice málo. Proto se v GIS příliš nepoužívá.

Pokud například do modelu, který je zobrazen, budeme chtít přidat organizace, které sdružují více států, např: EU, OSN, NATO, dostáváme se v tomto datovém modelu do problémů hned ze dvou důvodů. Za prvé třeba pro světadíl Antarktida žádná takováto organizace neexistuje, protože tam nejsou žádné státy. Druhý problém je, že výše jmenované organizace mohou působit na více světadílech a dostáváme se k otázce, zda má být ve vyšší hierarchické úrovni typ entity světadíl, či organizace.

Sít'ový model

výhoda - není omezen vazbou dítě - jeden rodič

nevýhoda - časem velice složité (vazby jsou velice komplikované), snadno se ztrácí přehled nad databází, pro udržení integrity DB je nutný velký výkon.

Podobně jako hierarchický model, se tento model příliš v GIS neosvědčil

Relační model

Organizuje data do tabulek

Každá tabulka má položku primární klíč, která je unikátní tj. jednoznačně identifikuje položku (entitu).

Díky klíčům je možné logicky spojovat více tabulek - vytvářet relační vztah (relaci) 1:1, 1:N, M:N.

Pro odstranění redundance (duplicity dat) se při návrhu struktury používá tzv. normalizace – problém databází (my nyní neřešíme)

Výhody:

- velice flexibilní struktura (lze ji měnit i za běhu, pouhým odebráním či přidáním tabulky a relace)
- snadný výběr z několika tabulek najednou pomocí relačního vztahu
- neredundantní uložení (při zachování pravidel normalizace)
- relační SŘBD (RDBMS) jsou velice rozšířené a podporované
- pro přístup k datům se používá jazyk SQL, který je v relačních databázích standardem
- široce podporován dalšími nejrůznějšími produkty (Excel, statistický software, ...)
- díky výkonu, standardizaci a rozšíření se dnes jedná o **nevyužívanější model v GIS**.

Objektový model

Vychází z objektově orientovaného programování, kde jsou data spravována jako objekty, což více přibližuje model reálnému světu. Pro každý objekt, jsou popisovány nejen jeho vlastnosti (atributy), ale i způsob jeho chování (metody).

Příklad: meteorologická stanice má atribut teplotu a rekordní teploty (tmin, tmax). Když teplota překročí nějaký rekord - mezní hodnotu, jsou automaticky aktualizovány i atributy rekordních teplot - informace o chování objektu (metoda).

Je možné vytvářet složitější objekty z jednodušších, říkáme, že potomek dědí vlastnosti a chování po rodiči + přidává svoje vlastní.

Příklad: z meteorologické stanice pro teploty můžeme vytvořit (děděním) meteorologickou stanici pro srážky a teploty. V relačním modelu je tohle obtížné a časově náročné.

Jednotlivé objekty mezi sebou komunikují pomocí zpráv.

Příklad: meteorologická stanice v Plzni zjistí rekordní teplotu pro Plzeň (a nastaví podle toho své atributy). Zároveň zašle zprávu centrální meteorologické stanici pro ČR o této události. Centrální stanice vyhodnotí, jestli jde o rekordní teplotu v celé republice a podle toho nastaví své vlastní atributy.

Objekty stejných vlastností (např. meteorologické stanice v Krkonoších, na Šumavě) jsou popsány jako třída objektů (class). Konkrétní meteorologická stanice - individuální objekt se pak nazývá instancí této třídy.

Výhody:

Není třeba definovat primární klíč - každý objekt má tzv. Object Identification Descriptor (OID), který má po celou dobu jeho existence (mezi tím je možné objekt modifikovat, přidávat atributy i chování). Pomocí tohoto OID na sebe jednotlivé objekty ukazují.

Model je velice flexibilní (je možné vytvářet a modifikovat objekty za běhu).

Snadná podpora časových dat (meteorologická stanice ví kdy byla postavena a zbořena).

V databázi je často obsažena i podpora verzí (versioning) - důležité pro GIS.

Nevýhody:

nedostatek zkušeností a standardů (neexistuje obdoba SQL),

složitost - jelikož má větší možnosti, je takové řešení dražší a složitější,

relativně malý výkon (OODB jsou dnes méně výkonné než RDBMS).

Po vyhodnocení všech výhod a nevýhod objektového a relačního modelu nakonec vznikl objektově-relační model, který zachovává všechny výhody relačního modelu a přidává výhody objektového modelu. Dnes je objektově-relační model používán u většiny velkých databází (Oracle 8i, Informix, ...).

Činnosti v rámci projektu GIS

Jednotlivé činnosti v GIS můžeme rozdělit podle jeho základních funkčních komponent.

1. Návrh databáze a vstup dat.
2. Zpracování a uchování dat.
3. Vykonávání analýz a syntéz z využitím prostorových vztahů - jádro GIS, tedy to co nejvíce odlišuje GIS a jiné IS.
4. Prezence výsledků (výstupy grafické - mapy, negrafické - zprávy, souhrnné tabulky).
5. Interakce s uživatelem (desktop GIS, Web GIS).

Návrh databáze a vstup dat

Základní kroky při návrhu databáze GIS:

Modelování uživatelského pohledu - co vlastně je potřeba, jaká data budou potřeba, ...

Definice objektů (entit) a jejich vztahů - jaké objekty budou popisovány, jaké jsou mezi nimi vztahy, co může konkrétní objekt obsahovat (jakých v DB nabývá hodnot)...

Identifikace reprezentace jednotlivých objektů (entit) - jakým způsobem (prostorovým či atributovým datovým typem) budou reálné objekty reprezentovány v GIS (jako body, linie, polygony, text, rastry, atributy, ...), jaké analýzy budou potřeba,...

Přizpůsobení dat konkrétnímu GIS - zde je nutné prozkoumat, co který GIS SW poskytuje, jaká jsou jeho omezení (maximální počet bodů, liniových segmentů, možnost ukládání do souborů či do DB,), jak je možné mé objekty přizpůsobit konkrétní aplikaci - v tomto bodě by měl také probíhat výběr konkrétního SW (v závislosti na prvním bodě).

Organizace dat do logických (geografických) celků - v podstatě se jedná o tématickou organizaci dat v závislosti na konkrétním GIS SW (po mapových listech x po zájmových oblastech, po vrstvách x objektově, ...).

Modelování reálného světa: konceptuální model, logický model

Vstup prostorových dat

Naplňování databáze - nejnáročnějším a nejzdlouhavějším krokem v rámci GIS projektu.

Vstup - různé zdroje údajů: mapy, náčrty v souřadnicovém systému, údaje z geodetických měření, fotogrammetrické snímky a obrazové záznamy DPZ, statistické údaje a další.

Při pořizování dat je důležité vybrat vhodný způsob a vhodná technická zařízení, která mi umožní získat data ve vhodné přesnosti a za přijatelnou cenu.

Dělení zdrojů dat: primární a sekundární

Naplňování GDB prostorovými daty – primární zdroje

Primární zdroje = přímo měřené

Vstup z geodetických měření

Zpracování obsahu klasických terénních zápisníků údajů pozemních geodetických měření.

Zadání ručně přes klávesnici a převedou do vektorové podoby. Většina GIS systémů má tzv. COGO modul (coordinate geometry - souřadnicová geometrie).

Velice stručně řečeno, jedná se o modul, který po zadání prvního bodu v souřadnicích X,Y umožňuje zadávat další body pomocí směru a vzdálenosti od prvního bodu.

Geodetické přístroje umožňují zaznamenávat údaje o měření do digitální podoby rovnou a pak se přenesou do prostorové databáze opět pomocí COGO.

Tento způsob je hlavně používán pro mapy velkých měřítek (katastrální mapy, technické mapy, plány, ...).

Produkuje vektorová data.

Vstup dat z GPS

Po naměření dat pomocí GPS - dostanu soubor [X,Y,Z] nebo [B,L,H] souřadnic v textovém tvaru a ty většina systémů umožní snadno zpracovat.

Poznámka: V poslední době se GPS hojně používá pro navigaci, sledování objektů (vozidel, ...) v reálném čase a analýzy v GIS na jejich základě.

Poznámka: GPS udává geografické souřadnice v souřadnicovém systému WGS 84, tudíž pro použití v ČR je nutné u získaných dat většinou převést data do jiného souřadnicového systému (S-JTSK, S 42).

Výhody GPS proti klasickým geodetickým metodám měření:

levný a rychlý sběr dat zejména bodových polí a měření v extravilánu (mimo zastavěnou část obce),
dá se měřit kdykoliv (v noci) a za každého počasí,
snadná konverze do GIS systémů,
v poslední době jsou GPS systémy vysoce přesné (dosahují přesnosti až cm) , také ale drahé
(obvykle platí čím dražší, tím přesnější - pro přesnost v cm až mm je nutné do aparatury investovat okolo 1-2 mil. korun).

Nevýhody:

vysoké budovy a stromy (v lese) blokují signály satelitů,
relativně složitá konfigurace systému (pořízení, přeškolení klasických měřičů, ...),
špatně se měří nedostupné objekty.

GPS měření produkuje vektorová data.

Fotogrammetrická data

Fotogrammetrie (FGM) je věda zabývající se rekonstrukcí tvaru, velikost a polohy předmětů zobrazených na fotogrammetrických snímcích.

Měření se uskutečňuje na fotografii, ne na objektu, jedná se tedy o bezkontaktní (nepřímou) metodu sběru dat.

Existuje fotogrammetrie letecká a pozemní, a také jednosnímková a dvousnímková, analogová a digitální.

Problematika převodu centrální projekce do ortogonální.

Problematika mozaikování, diferenciálního překreslování, ...

Výstup fotogrammetrie - digitální model reliéfu, digitální ortofoto.

Produkuje rastrová data.

Data jsou k dispozici v rastrové podobě.

Vstup z DPZ (Dálkový průzkum Země)

Remote Sensing (družicové snímky a obrazové záznamy)

Výhody DPZ z kosmického nosiče:

V porovnání s leteckými snímky poskytují komplexní obraz rozsáhlého území (tisíce km²) na jednom obrazovém záznamu.

Pravidelný sběr údajů.

Možnost rychlého zpracování (obvykle za několik hodin po jejich naměření).

Opakovatelnost aplikace stejných metod digitální interpretace - je možné snadno sledovat např. časové změny krajiny.

Nevýhody DPZ:

Obvykle pro měřítko 1:25 000 a menší (dnes již existují i data z DPZ téměř srovnatelné podrobnosti jako fotogrammetrická data, problém je ovšem v tom, že jsou prodávána pro velké oblasti).

Náročné na SW a HW i školený personál.

Sekundární zdroje – již jednou zpracovaná prostorová data

Obecně lze o sekundárních zdrojích říci, že jsou v nich obsaženy chyby získané již během prvního zpracování dat, tudíž **nemohou být přesnější** než zpracovávané primární zdroje.

Manuálně přes klávesnici
prakticky nepřichází v úvahu, velice pracné, např. zadávat souřadnice, ...

Manuální digitalizace
využívá se tablet-digitizér, což je zařízení na snímání souřadnic s různě velkou pracovní plochou (obvykle A3-A0) a různou rozlišovací schopností a přesností (maximální přesnost jsou řádově setiny milimetru). Zde je ovšem třeba kalkulovat i s měřítkem podkladu!
Princip digitalizace: snímáný podklad se upevní na pracovní plochu a pomocí zaměřovacího kříže (kurzoru) je snímána poloha zaměřovaných bodů a z klávesnice nebo pomocí kurzoru se zadává identifikátor objektu.

Existují dvě základní metody digitalizace:

bodová (point) - kliká se na každém vrcholu, který je třeba zaznamenat - je to nejčastější způsob použití,

proudová (stream) - počítač automaticky zaznamenává sekvence bodů v zadaném časovém nebo vzdálenostním intervalu.

Postup digitalizace

Konkrétní postup digitalizace:

Definování oblasti - definování minimálních a maximálních hodnot souřadnic.

Registrace mapy - zadání nejméně 4 kontrolních (vlícovacích) bodů (co možná nejvíce po obvodu). Jedná se například o rohy mapových listů, od kterých známe souřadnice v souřadnicovém systému. Nejprve se do GIS zadají souřadnice těchto bodů v cílové soustavě, např.: S-JTSK a potom se tyto body identifikují (kliknutím) na mapě.

Výsledná přesnost digitalizace záleží na měřítku a přesnosti původní mapy.

Vlastní digitalizace mapy

Editace chyb - nespojení čar, nedotahy a přetahy, vícenásobné zaznamenání - souvisí s topologickým čištěním

Výhody ruční digitalizace:

Malé finanční nároky; digitizéry jsou relativně levné, pracovní síla je také levná.

Flexibilita a adaptibilita na různé zdroje dat.

Technika je snadno zvládnutelná v krátkém čase - lze se snadno naučit.

Kvalita výstupů je víceméně vysoká.

Digitizéry jsou velice spolehlivé a přesné (přesnější než zdrojová data).

Snadné úpravy digitalizovaných dat.

Nevýhody:

Přesnost je limitována stabilitou vstupního média.

Digitalizace je únavná a nudná, tudíž velice náchylná k operátorovým chybám.

Skenování, vektorizace

Konkrétní postup při scanování:

Výběr rozlišení - to je docela důležité rozhodnutí, jelikož platí, že dvakrát větší rozlišení vede ke čtyřnásobné velikosti výsledného souboru. V manuálech ke GIS doporučeny hodnoty pro různé typy podkladů.

Výběr přesnosti scanneru a také přesnosti vstupních dokumentů (nikdy nedostanu přesnější výstup než vstup, vždy je to naopak). V této části je také nutné uvažovat zkreslení vstupních dokumentů (papír se roztahuje a smršťuje - je lepší scannovat z nějakých nesrážlivých materiálů). Vyhodnocení nároků na rozlišení a přesnosti vede k výběru konkrétního scanneru.

Příprava mapy ke scannování - očištění od mechanických nečistot, identifikace vlíčovacích bodů, případně úpravy nečitelných částí.

Vlastní scannování.

Registrace (transformace rastru) pomocí vlíčovacích bodů, probíhá analogicky jako u digitalizace.

Volitelně **úprava obrazu** (jas, kontrast, prahování, ekvalizace histogramu).

Volitelně **vektorizace.**

Automatická vektorizace - vše dělá počítač. Je to velice rychlé (co se týče nároků na uživatele), ale je pak nutné provádět čištění vektorových dat, což je velice zdlouhavé (záleží na podkladu).

Polautomatická vektorizace - interaktivní metoda, s tím že počítač sám vektorizuje, ale uživatel jej koriguje na sporných místech (ArcScan, Descartes).

Ruční (on screen digitizing) - interaktivní, kdy uživatel provádí sám vektorizaci na základě rastrového podkladu. Některé systémy umožňují automatizovat alespoň přichycení na rastr (Kokeš, GeoMedia Pro).

Naplňování GDB atributovými daty

Vstup atributových dat

Manuální

Nejběžnější způsob zadávání atributových dat je manuálně, pomocí klávesnice, na což stačí pouze jednoduchý hardware. Při zadávání atributů může nastat problém s doménovou integritou (např. zadaný věk = -1), ta se ale většinou kontroluje již během zadávání.

Atributy se pak navazují na prostorovou část pomocí unikátního identifikátoru, např. primárního klíče, který prostorové prvky již obsahují (vytváří se obvykle již při jejich tvorbě).

U ručního zadávání je poměrně velký problém kontroly správnosti zadaných údajů. Mohou se používat dvě základní metody kontroly:

Single Key Data Entry - jeden operátor zadává atributová data a druhý operátor již zadaná data kontroluje (porovnává originál s vytištěnými výpisy, ...). Tato metoda se používá při limitovaném čase a financích a je vhodná spíše pro menší projekty.

Double Key Data Entry - atributová data jsou zadávána dvěma na sobě nezávislými operátory (každý zadává stejná data) a poté se obě varianty v počítači porovnají. Při nalezení rozdílných hodnot se zadaný atribut překontroluje a opraví. Tato metoda se používá spíše na větší projekty, u kterých velice záleží na správnosti zadaných údajů.

Vstup atributových dat

Scannování + rozpoznávání textu

Další možností je scannování textu obsahující žádané atributy a poté jeho automatizované rozpoznávání pomocí nějakého OCR (Optical Character Recognition - nástroje na rozpoznávání písma) software.

Tato metoda, ačkoli relativně velice rychlá, je stále úspěšná jen z části a je možné ji aplikovat většinou pouze na již tištěný text (i z psacího stroje). Po automatickém převodu je navíc nutné vše pečlivě zkontrolovat (podobně jako u manuálního zadání pomocí metody Single Key Data Entry). Další nevýhodou je obvyklá nutnost ručního navazování atributů na prostorovou část, podobně jako u ručního zadávání dat.

Převod z externích digitálních zdrojů

Poslední možností je načítání atributových údajů z jiných, již digitálních, zdrojů.

Naplňování GDB konverzí dat z jiných digitálních zdrojů

Posledním způsobem získávání dat je jejich konverze z jiných systémů/formátů. Tento způsob, ačkoli vypadá zcela nevinně, může přinést celou řadu problémů se kterými je třeba počítat.

Před použitím dat z jiných digitálních zdrojů je třeba pečlivě prozkoumat několik kritérií, která rozhodnou o vhodnosti či nevhodnosti konkrétního zdroje:

Formát souboru - mám možnost ho použít/importovat, případně existuje konverzní program?

Přenosové médium - na čem budu data přenášet? (CDROM, disketa, DAT pásek, síť). Toto kritérium je důležité hlavně v případě přenosu dat velkých objemů, například letecké snímky.

Tématický obsah dat - jsou v datech obsaženy všechny prvky co potřebuji?

Měřítko a přesnost - jsou data v požadovaném měřítku a přesnosti ?

Časový interval pořízení - kdy byla data pořízena a k jakému časovému intervalu se vztahují?

Souřadnicový systém - v jakém SS byla data pořizována? Mohu takový souřadnicový systém využít (případně mohu provést transformaci do mnou používaného souřadnicového systému)?

Problematika compatibility datových modelů - např. problematika převodu křivek při převodu z CAD do GIS nebo i z GIS do GIS, převod formátu atributů.

Cena - ...

Možné chyby, které se mohou vyskytnout při vstupu dat

Možné chyby při zadávání:

Nekompletnost dat - scházejí body, linie, polygony.

Chybné umístění prostorových dat - chyby vycházející ze špatné kvality vstupních dat nebo z nedostatečné přesnosti při digitalizování.

Zkreslení prostorových dat - chyby z nepřesností vstupních dat (deformace podkladových dat, zkreslení již existující analogové kresby).

Špatná vazba mezi prostorovými a atributovými daty.

Atributy jsou chybné nebo nejsou kompletní - velice častá chyba zvláště pokud jsou atributy pořizovány z různých zdrojů v různých časech.

Nejčastější chyby odhalené při vytváření topologie:

třísky a mezery (Sliver and gaps) - jev nastává, když jsou dvě hranice digitalizovány z různých zdrojů, ačkoli v terénu představují jednu a tu samou. V takovém případě jsou linie představující tutéž hranici neidentické (nepřerývají se)

Mrtvé konce (dead ends) - nedotahy a přetahy.

Duplikátní linie reprezentující stejný objekt.

Pokud se používá pro reprezentaci polygonů metoda hranic a centroidů, tak i **přiřazení více centroidů jednomu polygonu**.

Metadata

data o datech, tzn. informace co popisovaná data obsahují a kde se nacházejí.

Tyto informace jsou zvláště důležité, pokud je zpracováváno několik druhů dat či velký objem dat. Pomáhají pak lépe organizovat a udržovat přehled na daty. Problematika tvorby metadat je úzce spjata s tvorbou digitálních dat a jejich převody.

Informace obsažené v metadatach:

Obsah dat (tématická složka, co data obsahují).

Rozlišení - prostorové (rozsah území), popisné (popis možných hodnot atributů a jejich význam) a časové (jakou dobu zahrnují - kdy byly aktuální).

Formát dat (typ souboru, předávací médium).

Datum pořízení dat (případně aktualizace).

Kontakt na pořizovatele a správce.

Pro metadata je vhodné vyvinout standard, který by byl široce akceptovaný (aby bylo možné snadněji převádět data z jiných GIS systémů). Dnes již existuje několik standardů, v ČR se o metadata stará například Česká asociace pro geoinformace (ČAGI). Více informací a odkazů je možné nalézt na serveru ČAGI (www.cagi.cz).

Příklady metadatových serverů:

<http://www.geographynetwork.com>

<http://gis.vsb.cz/midas/>