

DATA

- **prostorová data**
- **atributových data**
- **metadata**

Vstup prostorových dat

Naplňování databáze je v drtivé většině případů jednoznačně nejnáročnějším a nejdélouhavějším krokem v rámci GIS projektu.

Obecně lze pro vstup použít různé zdroje údajů. V úvahu přicházejí zvláště mapy, náčrty v souřadnicovém systému, údaje z geodetických měření, statistické údaje a další. Při pořizování dat je ale důležité vybrat vhodný způsob a vhodná technická zařízení, která umožní získat data ve vhodné přesnosti a za přijatelnou cenu.

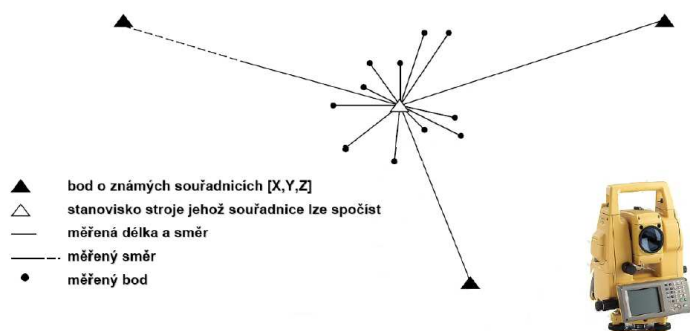
V zásadě je možné zdroje dat rozdělit na:

- **Primární** – přímo měřená data
 - terestrická (pozemní/geodetická) měření
 - Globální polohové systémy (GPS)
 - Fotogrammetrie
 - Dálkový průzkum Země (DPZ)
 - Laserové skenování (LIDAR)
- **Sekundární** – již jednou zpracovaná data
 - manuální vstup přes klávesnici
 - digitalizace
 - skenování a vektorizace
 - konverze dat z jiných datových zdrojů

Primární zdroje

Geodetické měření

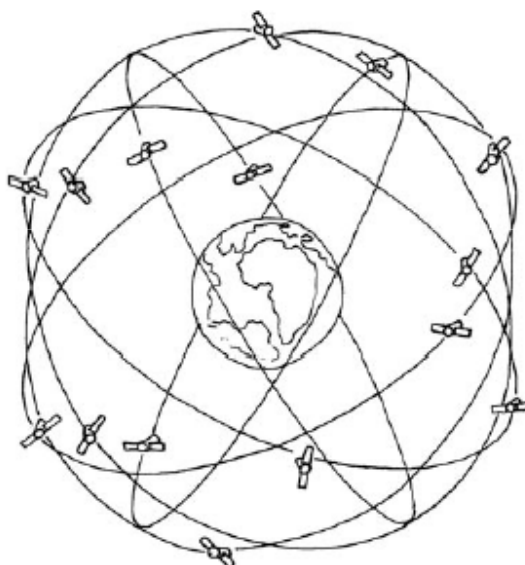
- zpracování obsahu klasických terénních zápisníků údajů pozemních geodetických měření.
- zadají se ručně přes klávesnici a převedou do vektorové podoby. Většina GIS systémů má tzv. COGO systém (souřadnicová geometrie)
- novější systémy umožňují zaznamenávat údaje o měření do digitální podoby rovnou a pak se přenesou do prostorové databáze opět pomocí COGO
- tento způsob je hlavně používán mapy velkých měřítek (katastrální mapy, technické mapy, plány, ...)
- produkuje vektorová data, přesnost na cm.



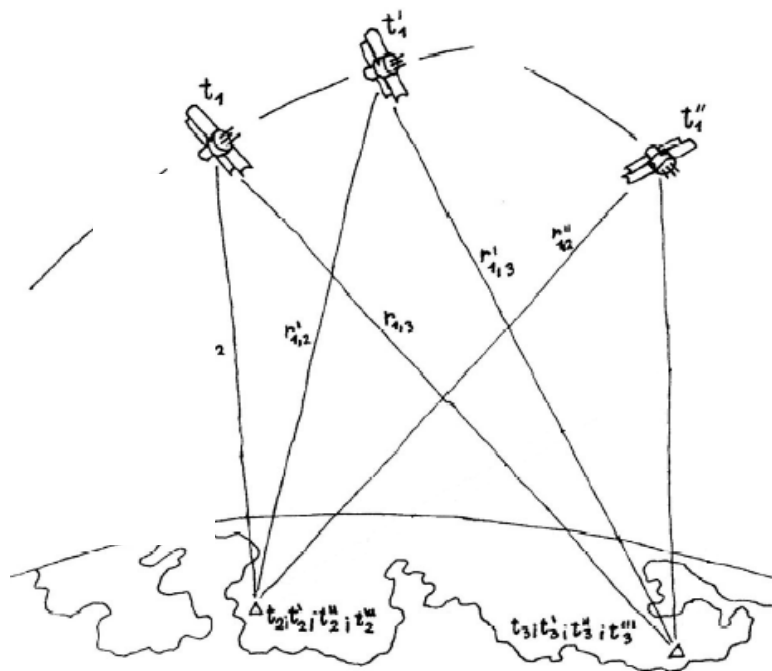
Primární zdroje

GPS – globální polohový systém

- GPS NAVSTAR (Spojené státy americké)
 - GLONASS (Ruská federace)
 - Galileo (Evropská unie)
- Je vystavěna síť 24 družic, které jsou rozmístěny na orbitální dráze ve výšce 22000 km tak, aby bylo z každého místa na zemi dostupných alespoň 6 satelitů.
 - Každá družice zná svoji přesnou polohu na orbitu, tudíž pozemní přijímač je schopen po zaměření několika (3 pro zem. šířku a výšku, 4 i pro výšku nadmořskou) satelitů rekonstruovat jeho přesnou polohu (z polohy satelitů a vzdálenosti od nich).



- Přesnost měření závisí na typu GPS (např. turistická GPS – přesnost cca 5 m) a případně další úpravy dat (post processing – zpřesní data až na desítky cm)
- Po naměření dat pomocí GPS se musí převést do GIS – jednoduché (soubor X,Y a Z souřadnic v textovém tvaru lze snadno importovat).
- GPS vrací geografické souřadnice nebo WGS 84, tudíž pro použití v ČR je nutné u získaných dat většinou převést data do jiného souřadného systému.



Obr. 15 Trojitá diference

Výhody GPS:

- levný a rychlý sběr dat (až o 50%),
- 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

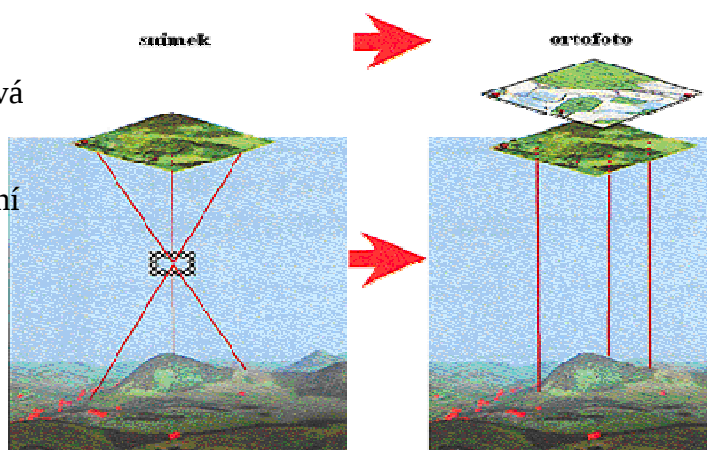
Nevýhody GPS :

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

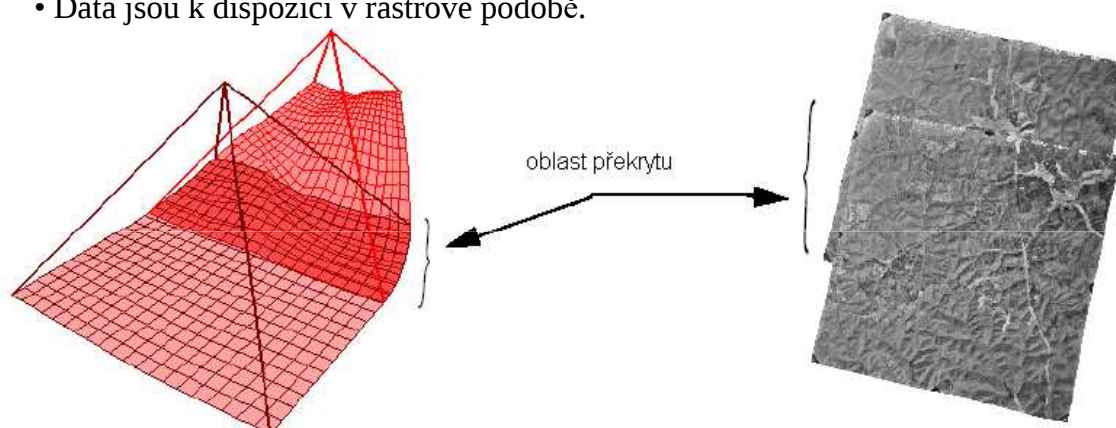
Primární zdroje

Fotogrammetrické zpracování leteckých či pozemních snímků

- fotogrammetrie je věda zabývající se rekonstrukcí tvaru, velikost a polohy předmětů zobrazených na fotogrammetrických snímcích.
- měření se uskutečňuje na fotografii, ne na objektu
- fotogrammetrie je:
letecká a pozemní,
jedno- a dvoj-snímková, analogová
či digitální
- umožňuje převod objektu
z centrální projekce do ortogonální
- produkuje rastrová data

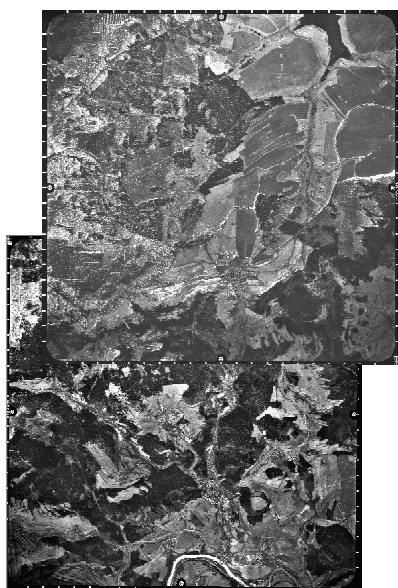


- Výstupy z fotogrammetrie - digitální model povrchu, digitální ortofoto.
- Přesnost závisí na velikosti pixelu, v současnosti cca 0,17 – 0,5 m.
- Data jsou k dispozici v rastrové podobě.



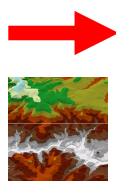


Překreslení snímků na ortofoto

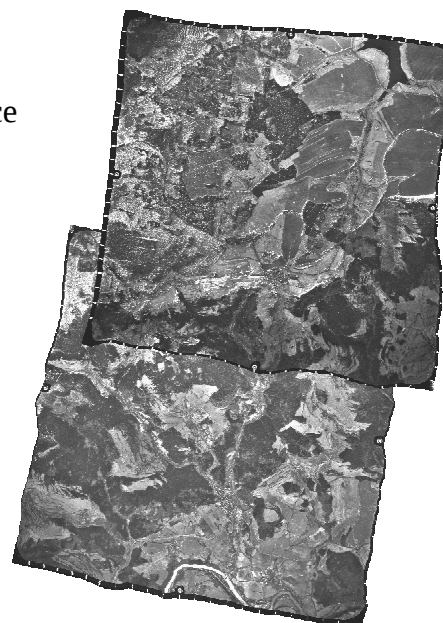


surový snímek

výsledek
triangulace



DTM

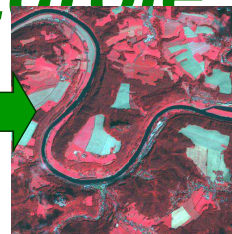


ortofoto

Primární zdroje

DPZ – dálkový průzkum Země

- data z leteckých a družicových nosičů,
- objekty mohou být identifikovány z velké vzdálenosti, jelikož vyzařují nebo odrážejí elektromagnetickou energii (zdroj je hlavně slunce). Spektrální charakteristika pak identifikuje jednotlivé objekty.
- využití v životní prostředí (sledování ozónové vrstvy, olejových skvrn, stavu napadení lesů škůdci a další), biologie, hydrologie, geologie, využití země a další.
- data jsou k dispozici v rastrové podobě.



Systémy DPZ používají ke snímání dva druhy senzorů:

- **pasivní** – zaznamenává vyzařené nebo odražené elektromagnetické vlny (od slunce)
- **aktivní** – používá svůj vlastní zdroj elektromagnetického vlnění, tudíž je možné ho používat jak ve dne, tak v noci. Výhodou aktivních senzorů je také schopnost monitorovat mnohem delší vlnové délky, než pasivní. Výhodou delších vln je schopnost lépe pronikat atmosférou, mraky a dokonce i mělkou vodou. Nevýhodou aktivních systémů je nutnost poskytovat energii senzoru.

Na snímcích DPZ jsou důležité následující parametry:

spektrální rozlišení (popsáno počtem a šířkou spektrálních pásem, zaznamenávaných snímačem)

příklad šedotónové obrazy mají jen jedno pásmo, barevné jsou lepší a mají 3

(R G B): Například Thematic Mapper (TM) zaznamenává 7 pásem

prostorové rozlišení – nejmenší plocha, která je rozlišitelná.

TM má 15x15 m v panchromatickém a 30x30 m

v multispektrálním módu, SPOT má 2,5x2,5 m

v panchromatickém a 10x10 m v multispektrálním módu,

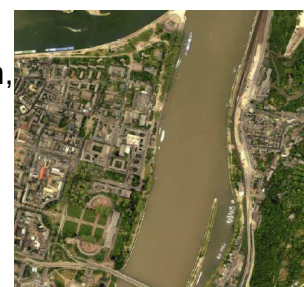
QuickBird 0,6x0,6 m v panchromatickém a 2,5x2,5 m

v multispektrálním módu.

radiometrické rozlišení – kolik různých hodnot (počet rozlišitelných úrovní – citlivost detektoru) mohou dostat v každém spektrálním pásmu (např. TM a SPOT mají 256, MSS má pouze 64)

časové rozlišení – což je interval mezi dvěma úspěšnými přelety nad tím samým územím. (zde může hrát roli i počasí).

Landsat-5 asi 16 dní, SPOT má asi 26dní (lze i po 2.5 dnech)



Ortofoto (RGB)



Ortofoto (NIR)

Rozlišení družicových dat



SPOT- rozlišení 5m (kombinace pásem zelené, červené a infračervené barvy) → **dobře vynikají oblasti porostlé zdravou vegetací, které se jeví červenou barvou.**

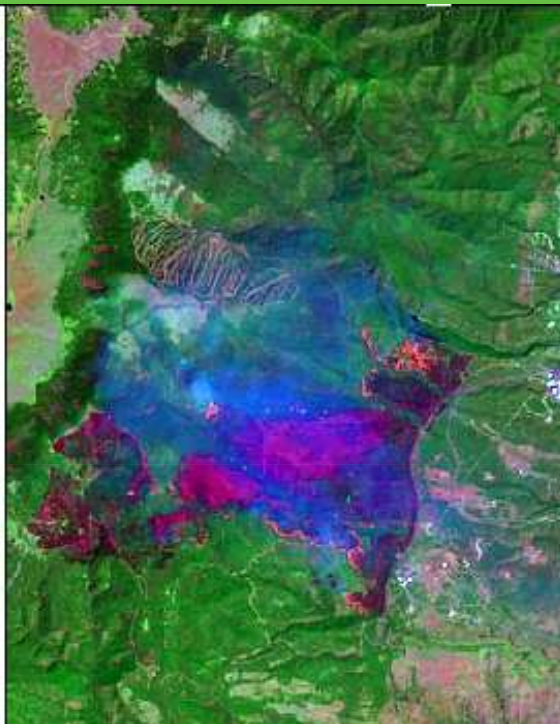
QuickBird

-rozlišení 2,44 – 2,88 m





pravé barvy



nepravé barvy: červená: krátkovlnné infračervené, zelená blízké infračervené, modré: zelená

Výhody DPZ:

- V porovnání s leteckými snímky poskytují komplexní obraz rozsáhlého území (tisíce km²) na jednom obrazovém záznamu
- Pravidelný sběr údajů
- Možnost rychlého zpracování (obvykle několik hodin po jejich naměření)
- Opakovatelnost aplikace stejných metod digitální interpretace – je možné snadno sledovat např. změny krajiny

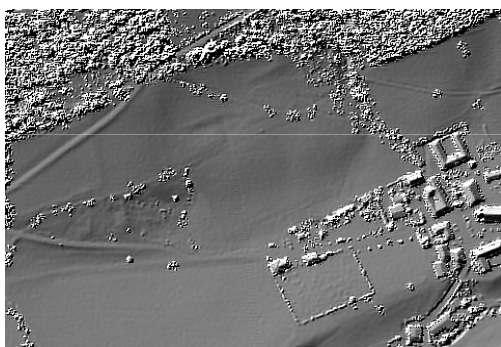
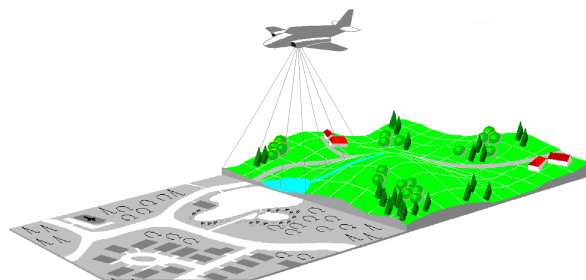
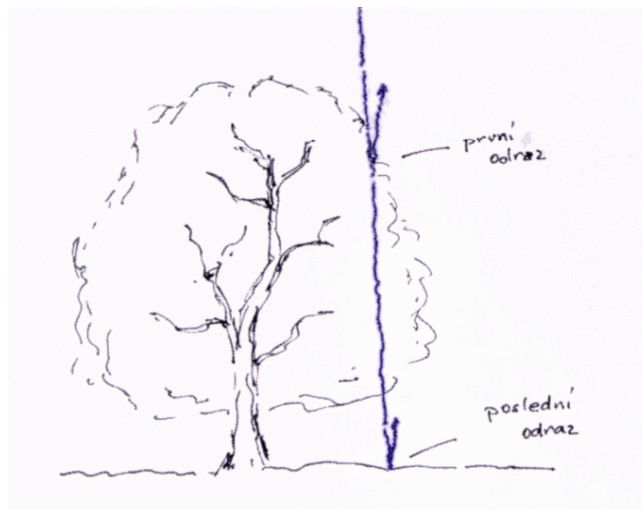
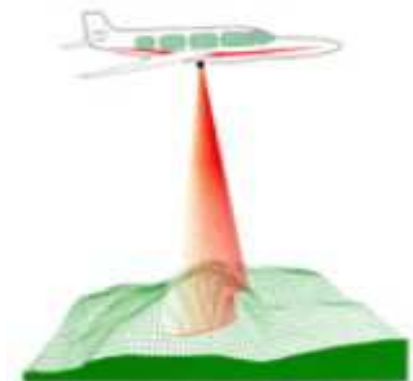
Nevýhody DPZ:

- Obvykle pro měřítko 1:25 000 a menší (nyní zlepšení)
- Náročné na SW a HW i školený personál

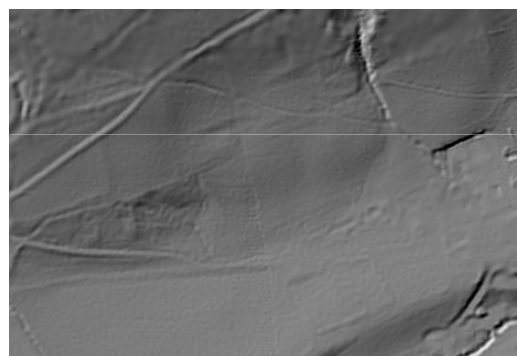
Laserové skenování – Light Detection And Ranging (LIDAR)

- technologie umožňující sběr digitálního modelu reliéfu a modelu terénu a to zvláště v zalesněných oblastech
- vhodný pro tvorbu 3D modelu města, analýzu vegetačního pokryvu, sledování nadzemních vedení
- vektorová data ve 3D

Letecké laserové skenování

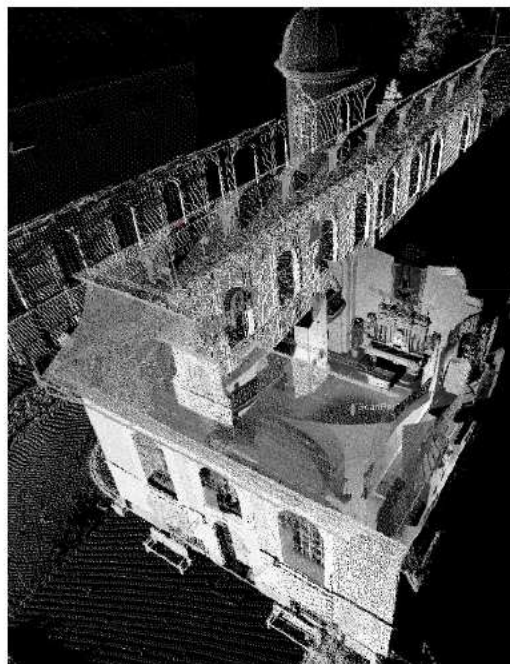
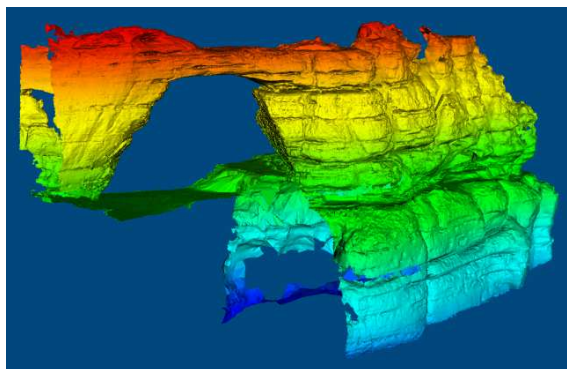


Digitální model povrchu



Digitální model reliéfu

Pozemní laserové skenování



Sekundární zdroje

- již jednou zpracované primární zdroje
- 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 (nutné zadávat souřadnice, ...)

■ Digitalizace

- přímo nad obrazovkou – dnes velmi rozšířené
- digitalizace se provádí pomocí tablet-digitizér (je to zařízení na snímání souřadnic s různě velkou pracovní plochou (obvykle A3-A0) a různou rozlišovací schopností a přesností (maximální přesnost jsou řádově setiny milimetru)).

Sekundární zdroje

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.



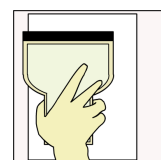
Sekundární zdroje

■ Skenování

- převod dat z analogové do digitální (rastrové) formy
- vykonává se pomocí scannerů, zařízeních sloužících k optickému snímání dokumentů.

Existují tři různé typy scannerů:

- Bubnové – vhodné pro velké plakáty, přesné, ale nedrahé, nejstarší typy skenerů.
- Deskové - stopni – nejběžnější, ale mají malou snímatelnou plochu (asi tak do A2-A3).
- Posuvné velkoformátové skenery - nejpoužívanější typ pro GIS.
- Ruční (hand-held) - nutno ručně přejíždět po snímané předloze, rychlé pořízení, ale nízká kvalita.



Konkrétní postup při skenování:

- **Výběr rozlišení** – to je docela důležité rozhodnutí, jelikož platí, že dvakrát větší rozlišení vede ke čtyřnásobné velikosti výsledného souboru.
- **Výběr přesnosti** (zkreslení $\pm 0,04\%$) a také přesnosti vstupních dokumentů (nikdy nelze dostat 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 skenovat z nějakých bezesrážkových materiálů). Vyhodnocení nároků na rozlišení a přesnosti vede k výběru konkrétního scanneru.
- **Příprava mapy ke skenování** – očištění od mechanických nečistot, identifikace vlíčovacích bodů, případně úpravy nečitelných částí.
- **Vlastní skenování.**

Další postupy zpracování:

- Registrace (transformace rastru) pomocí vlíčovacích bodů.
- Volitelně úprava obrazu (jas, kontrast, prahování, ekvalizace histogramu).
- Volitelně čištění rastrového podkladu.
- Volitelně vektorizace.

Vektorizace

Sekundární zdroje

- Stejný princip jako digitalizace, ale zákres prvků se provádí přímo nad obrazovkou – dnes velmi rozšířené

Princip vektorizace:

- **Automatická** - 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** - 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).



Sekundární zdroje

Konverze dat z jiných digitálních zdrojů

- načítání dat již digitálních zdrojů a jejich následná konverze z jiných systémů/formátů.

důležité faktory:

- **Formát souboru** - mám možnost ho použít/importovat, případně existuje konverzní program
- **Přenosové médium** – na čem budu data přenášet ? (CDROM, disketa, DAT pásek) Tohle kritérium je důležité hlavně v případě přenosu dat velkých objemů (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 byla data pořizována? Mohu takový souřadný systém využít (případně mohu provést transformaci souřadnicového systému)?
- **Problematika kompatibility datový modelu.**
- **Cena**

Vstup atributových dat

▪ Manuální

- zadávání dat pomocí klávesnice (nejběžnější způsob)
- atributy se pak navazují na prostorovou část pomocí primárního klíče

Nevýhody:

- 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.

■ Scanování

- scanování textu s využitím OCR (Optical Character Recognition – nástroje na rozpoznávání písma) pro automatizovaný převod dat

Nevýhody:

- nelze převést všechna data
- nutná pečlivá kontrola (podobně jako u manuálního zadání pomocí metody Single Key Data Entry).
- nutnost ručního navazování atributů na prostorovou část, podobně jako u ručního zadávání dat

■ Konverze dat z jiných digitálních zdrojů

- načítání atributových údajů z jiných, již digitálních zdrojů a jejich následná konverze z jiných systémů/formátů.

důležité faktory:

- **Formát souboru** - mám možnost ho použít/importovat, případně existuje konverzní program
- **Přenosové médium** – na čem budu data přenášet ? (CDROM, disketa, DAT pásek) Tohle kritérium je důležité hlavně v případě přenosu dat velkých objemů (letecké snímky, ...)
- **Tématický obsah dat** – jsou v datech obsaženy všechny prvky co potřebuji ?
- **Problematika kompatibility datový modelu.**
- **Cena**

Metadata

- jsou to data o datech, tzn. informace, co popisovaná data obsahují. Tyto informace jsou zvláště důležité, pokud je zpracováváno několik druhů dat. Pomáhají pak lépe organizovat a udržovat přehled na daty. Problematika tvorby metadat je úzce spjata s tvorbou digitálních dat a jejich převody.

V metadatach by měla být obsaženy následující informace:

- Co je obsahem dat (tématická složka).
- Rozlišení (prostorové (jaké území zabírají), 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.



Pozemky

popis: plocha MCHÚ s rozkreslenými pozemky
geometrický typ: polygon
formát: ArcView shapefile
název: Pozemky
zdroj: Linie.shp
vstup: zapločování vrstvy Linie.shp
stav k datu: 2003
atributy:

<i>jméno</i>	<i>popis</i>	<i>typ</i>	<i>počet míst</i>	<i>nabývané hodnoty</i>
Id	druh pozemku	N	15	2 – orná půda 3 – chmelnice 4 – vinice 5 – zahrada 6 – ovocný sad 7 – tr. trav. porosty 10 – lesní pozemky 11 – vodní plochy 13 – zastavěné plochy 14 – ostatní plochy

Kontrola údajů a odstraňování chyb

- Při vkládání dat do systému není možné zabezpečit správnost 100% zadání dat.
- Identifikace chyb je velice obtížná. Obvykle se data kontrolují vizuálně. Dalším způsobem kontroly chyb prostorových dat je proces vytváření topologie neboli topologické čištění dat.
- GIS mají většinou schopnosti procházet místa s potenciální chybou a umožní uživateli interaktivně odstranit případné chyby.

Možné chyby při zadávání:

- Nekompletnost dat – schází 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 daty a atributy.
- Atributy jsou chybné nebo nejsou kompletní – velice častá chyba zvláště pokud jsou atributy pořizovány z různých zdrojů v různých časech..

Identifikace chyb je velice obtížná.

Kontrola dat:

- Vizuálně
- topologického čištění při vytváření topologie

Nejčastější chyby odhalené při vytváření topologie (více viz topologické čištění dat):

- 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 (hlavně v CAD, ale i u některých GIS, které z toho vyrobí regulární polygon) reprezentující stejný objekt.
- Pokud se používá pro reprezentaci polygonů metoda hranic a centroidů, tak i přiřazení více centroidů jednomu polygonu.

Většina systémů obsahuje funkce, které umožní automaticky vyřešit některé z těchto uvedených chyb.

Chyby právního charakteru:

- při pořizování dat je nutné brát v potaz i právní souvislosti problematiky, kdo má na data obchodní práva, zda je možné data využívat pro akademické, soukromé, či obchodní účely.
- Zvláště v ČR je tato oblast velice problematická!