



# Geoinformatika

## V - Sběr dat

jaro 2015

**Petr Kubíček**

**kubicek@geogr.muni.cz**

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



# Zdroje prostorových dat pro GIS

- **Naplňování databáze** je v drtivé většině případu jednoznačně nejnáročnějším a nejzdlouhavějším 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, fotogrammetrické snímky a obrazové záznamy DPZ, statistické údaje a další.
- Při pořizování dat je ale důležité vybrat vhodný způsob a vhodná technická zařízení, která mi umožní **získat data ve vhodné přesnosti a za přijatelnou cenu**.
- V zásadě je možné zdroje dat rozdělit na **primární** a **sekundární**.



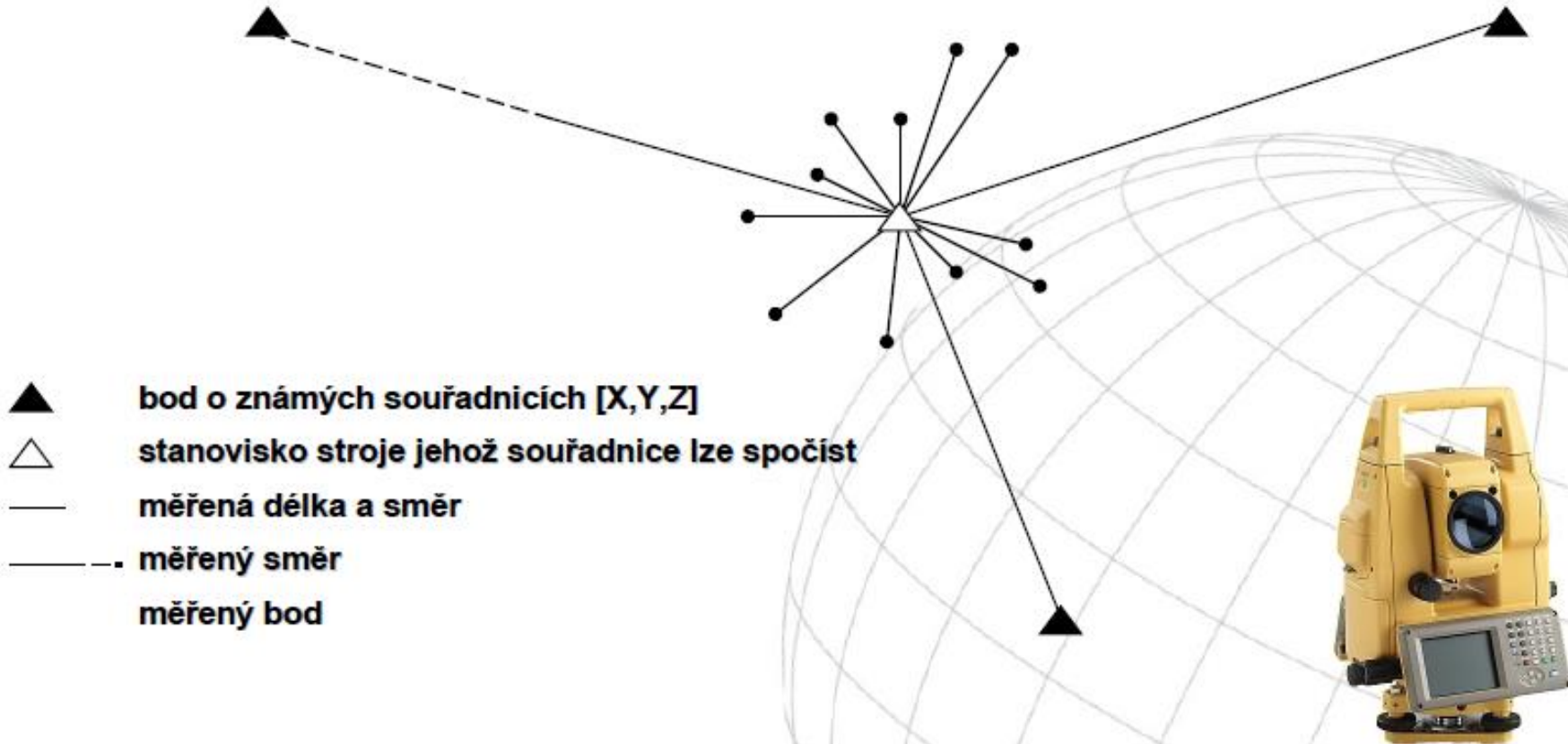
# Zdroje prostorových dat pro GIS

- **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



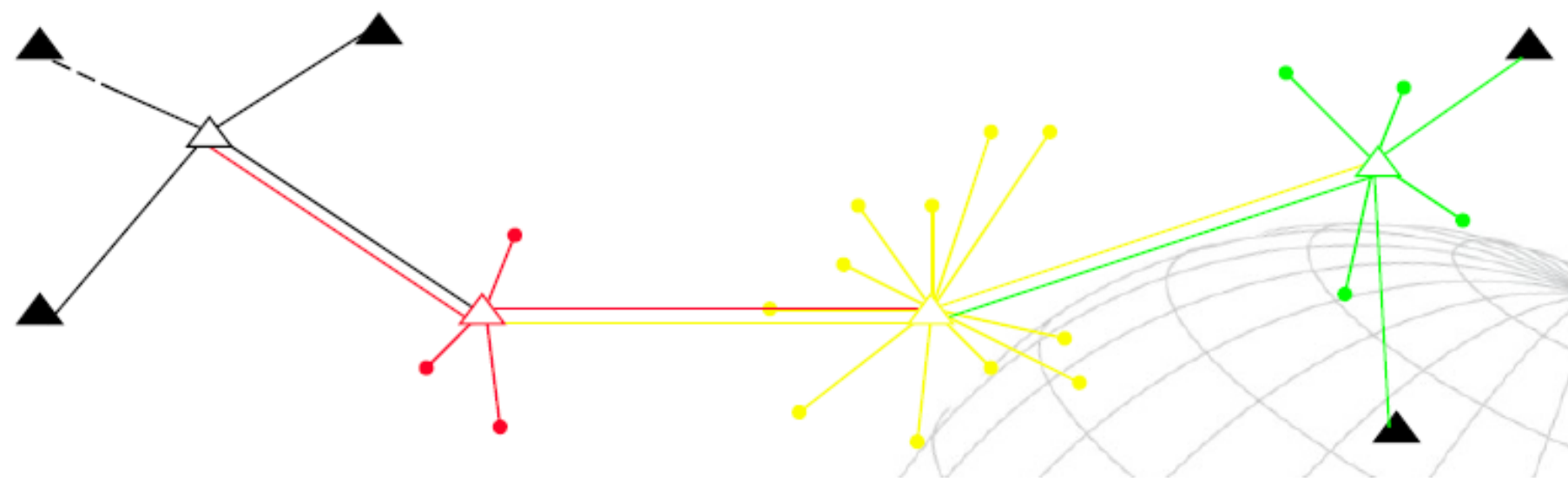
# Zdroje prostorových dat pro GIS

## Geodetická data





# Zdroje prostorových dat pro GIS





# Zdroje prostorových dat pro GIS

## Zpracování obsahu terénních zápisníků údajů pozemních geodetických měření:

- Ruční přepis papírového zápisníku nebo zaznamenání údajů o měření do digitálního zápisníku.
- Zpracování v geodetickém SW (někdy existuje jako modul v GIS).
- Import dat z geodetického SW (většinou CAD based).
- Jednoduchou variantou GIS modulu pro zpracování měřených dat je tzv. **COGO modul** (coordinate geometry ~ souřadnicová geometrie). Základní funkcionalitou je zadání prvního bodu v souřadnicích X,Y a následné zadávání dalších bodů pomocí směru a vzdálenosti od prvního bodu.
- **Používá se hlavně pro mapy velkých měřítek (katastrální mapy, technické mapy, plány, ...).**
- **Produkuje vektorová data, přesnost cm.**



# Zdroje prostorových dat pro GIS - GNSS

## **Global Navigation Satellite System (GNSS) – Globální družicový polohový systém**

- **GPS NAVSTAR (Spojené státy americké)**
- **GLONASS (Ruská federace)**
- **Galileo (Evropská unie)**

## **Rádiový dálkoměrný systém**

# Historie GPS

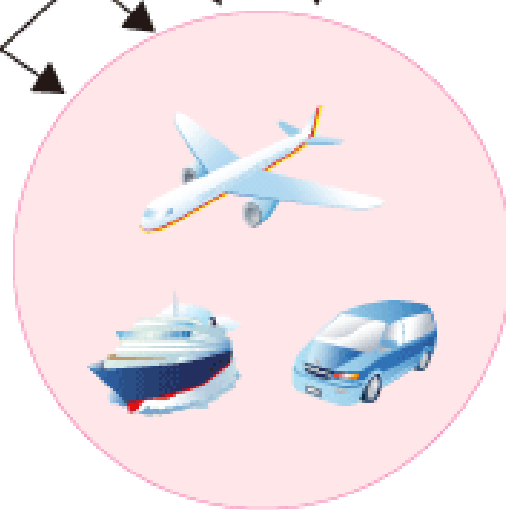
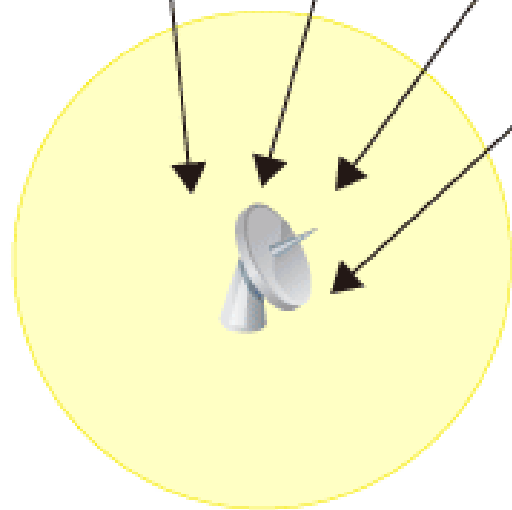
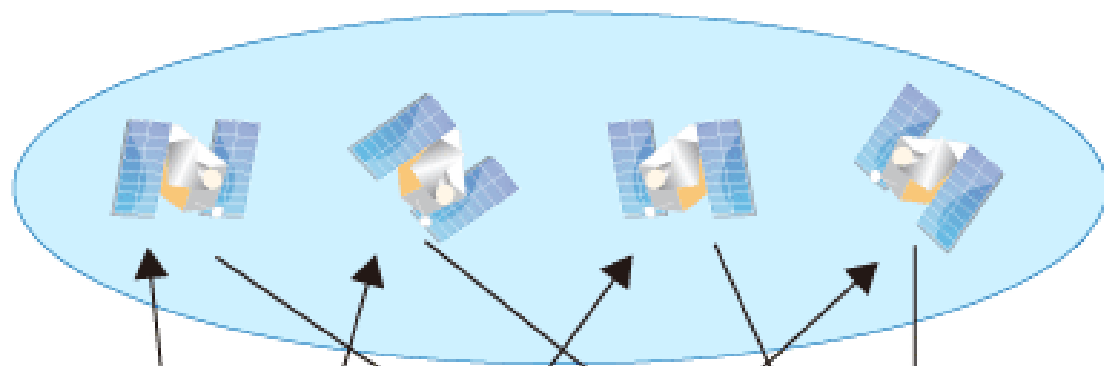
- **60. léta: USA – jak zjistit rychle a přesně polohu svých jaderných ponorek kdekoli na Zemi?**
- **70. léta – nalezení teoretického řešení a následná praktická realizace GPS (NAVSTAR), následně byla vystavěna síť 24 družic.**
- **Květen 2000: zrušení S/A – záměrné chyby zaváděné do GPS signálu.**
- **2005 – budování sítí referenčních stanic v ČR (CZEPOS, VESOG).**





# Segmenty GPS

Space segment



Control segment

User segment

- **Uživatelský**
- **Řídící**
- **Vesmírný**

# GPS přesnost

## – jeden přijímač

- navigační +/- 10m (ale až 40)
- „GIS“ i submetrová přesnost

## – geodetická souprava, či geodetická aparatura v síti referenčních stanic – cm/mm

- statická metoda (více přijímačů, dlouho, postprocessing)
- RTK (Real Time Kinematic) – jeden přijímač, korekce z permanentních referenčních stanic.



## Další charakteristiky GPS dat

- Po zpracování jsou GPS data ve tvaru souboru  $[X,Y,Z]$  souřadnic a ty většina systémů umožní snadno zpracovat.
- GPS se hojně používá pro navigaci, sledování objektů (vozidel, ...) v reálném čase a analýzy v GIS na jejich základě.
- **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).
- **produkuje vektorová data.**



# GPS x terestrické metody

**+**

- 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,
- v poslední době jsou GPS vysoce přesné

**-**

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

# Fotogrametrie

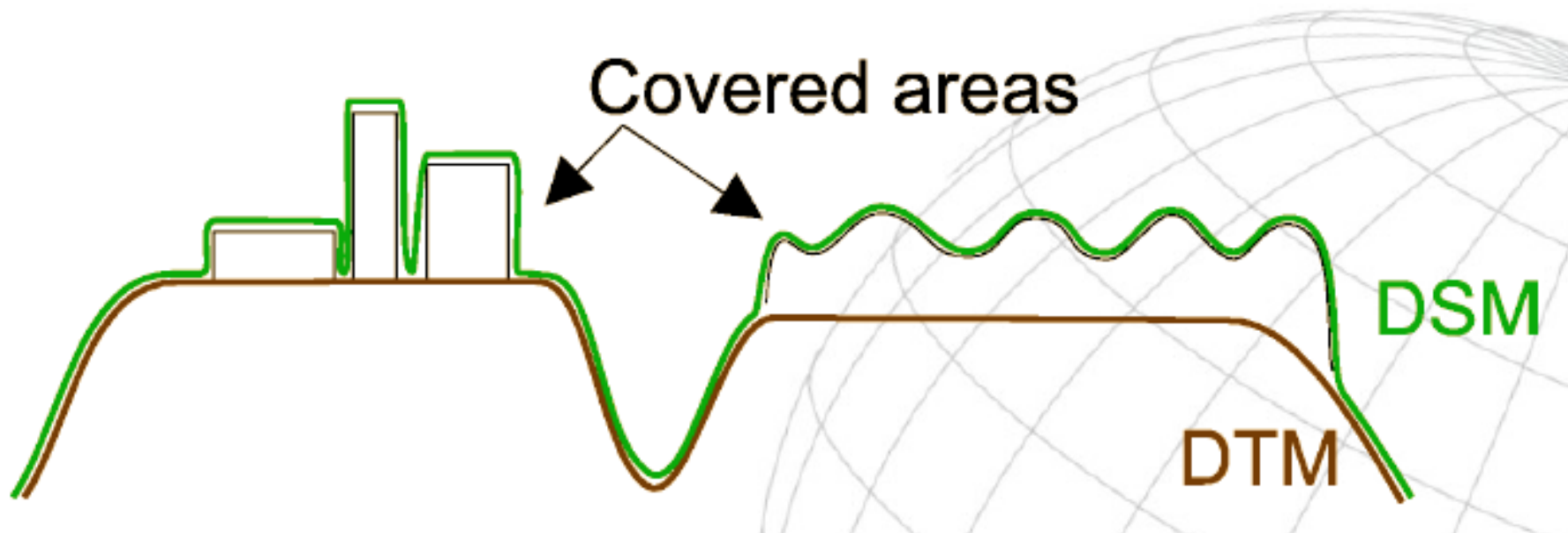
- Fotogrametrie (FGM) je věda zabývající se rekonstrukcí **tvaru, velikost a polohy předmětů** zobrazených na fotogrammetrických snímcích.
- Měření se uskutečňuje na fotografii, ne na objektu, jedná se tedy o **bezkontaktní** (nepřímou) metodu sběru dat.
- Existuje fotogrametrie letecká a pozemní, a také jednosnímková a dvousnímková, analogová a digitální.
- Výstupy z fotogrametrie - digitální model reliéfu (DMR), digitální ortofoto.
- Přesnost závisí na velikosti pixelu, v současnosti cca 0,17 – 0,5 m.
- Data jsou k dispozici v rastrové podobě.



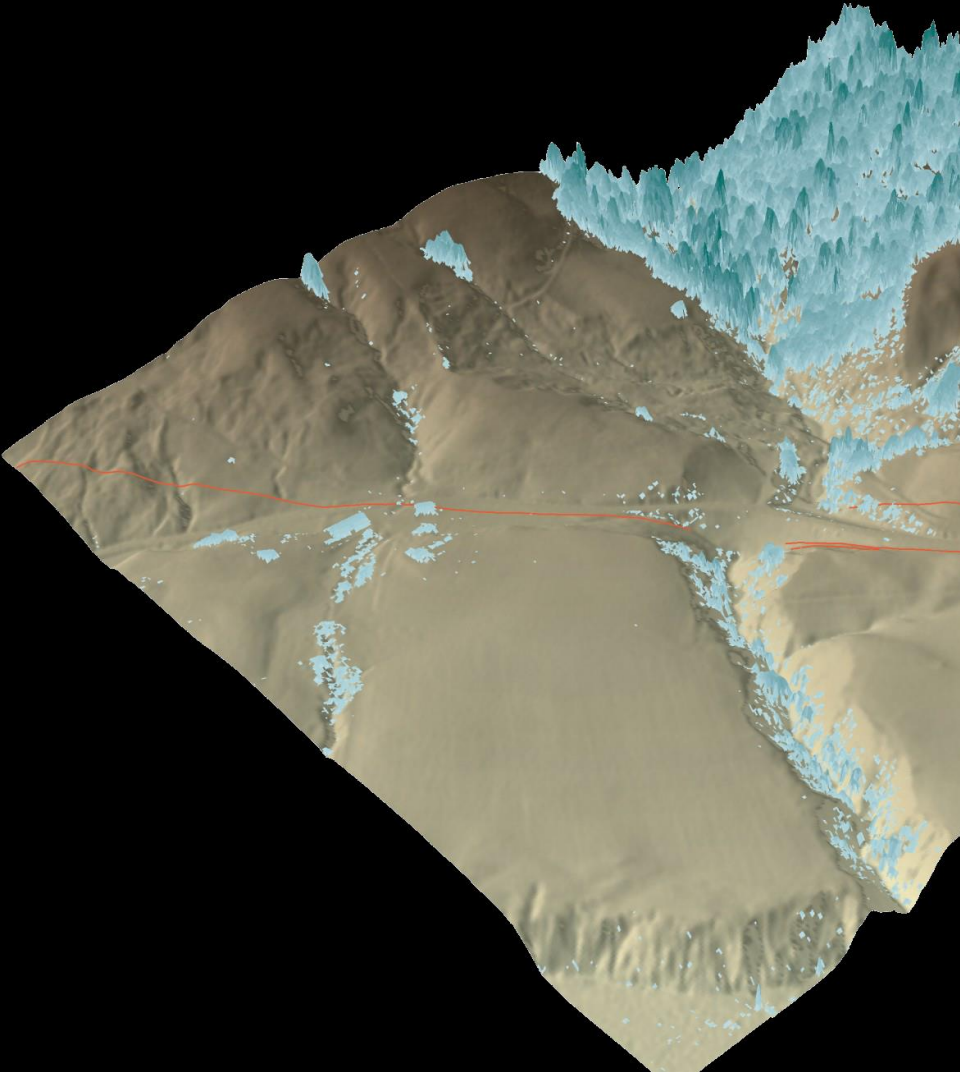
# Laserové skenování Lidar

- **Light Detection and Ranging (LIDAR)**
- **Princip LIDAR je postaven na aktivním senzoru, který vyšle laserový paprsek a zaznamená jeho zpětný odraz.**
- **Odrasů může být několik, první je považován za digitální model povrchu (DMP, DSM – digital surface model). Poslední odraz je pak považován za digitální model reliéfu (DMR, DTM – digital terrain model).**
- **Teprve potom se senzor otočí a zaznamenává další bod!**
- **Existují letecké a pozemní scannery!**

- **Digitální model povrchu x model reliéfu**



View to the northeast



**Geoinformatika**



# Sekundární zdroje dat

- **Sekundární zdroje dat jsou již jednou zpracované primární zdroje**
- **=> jsou v nich obsaženy chyby získané již během prvního zpracování dat, tudíž nemohou být přesnější než zpracovávané primární zdroje.**
- **Možnosti jejich vstupu do GIS**
  - manuální vstup přes klávesnici (pracné, zdlouhavé)
  - digitalizace
  - skenování a vektorizace
  - import dat.

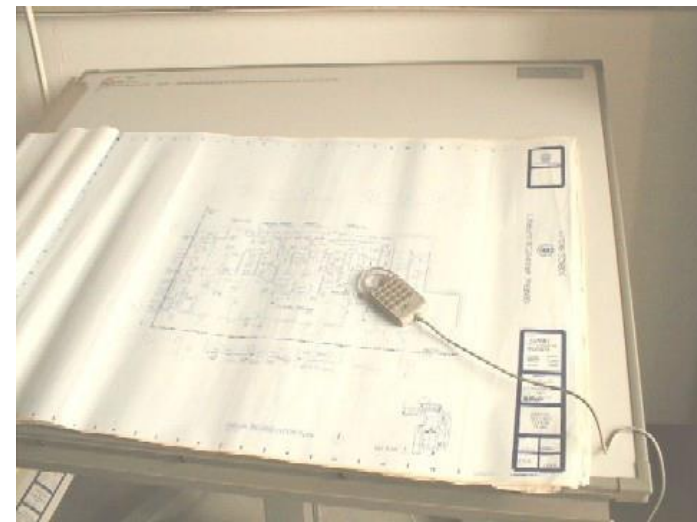
# Manuální digitalizace

- Využívá se tablet-digitizér, což je zařízení na snímání souřadnic s různě velkou pracovní plochou (obvykle A3-A0) a různou rozlišovací schopností a přesností .
- Je třeba kalkulovat i s měřítkem podkladu!
- Princip digitalizace – snímaný 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.

dvě základní metody digitalizace:

- **bodová** (point) - kliká se na každém vrcholu, který je třeba zaznamenat.
- **proudová** (stream) - počítač automaticky zaznamenává sekvence bodů v zadaném časovém nebo vzdálenostním intervalu.

**Geoinformatika**



# Postup digitalizace

1. **Definování oblasti** - definování minimálních a maximálních hodnot souřadnic.
2. **Registrace mapy** - zadání nejméně 4 kontrolních (**identických**) 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.
3. **Vlastní digitalizace** mapy.
4. **Editace chyb** - nespojení čar, nedotahy a přetahy, vícenásobné zaznamenání - souvisí s topologickým čištěním (viz. dále).

# Digitalizace – výhody a nevýhody

+

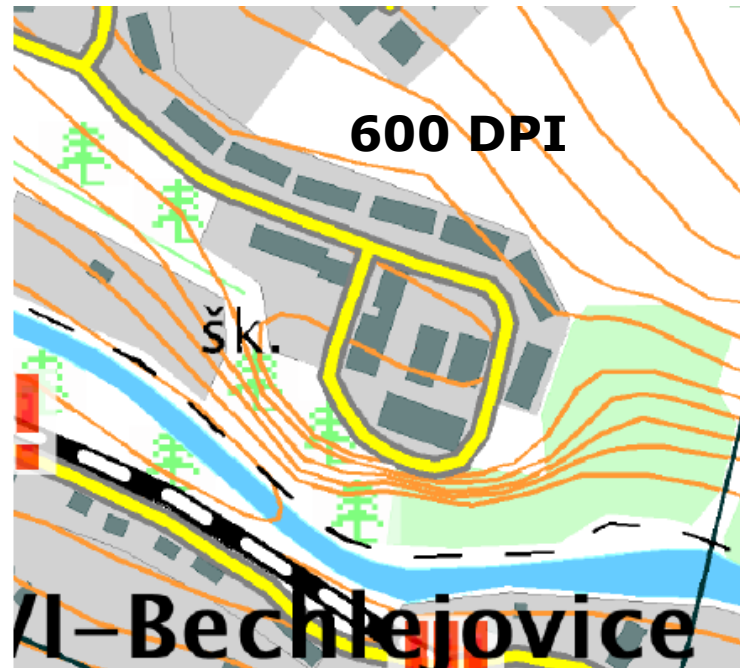
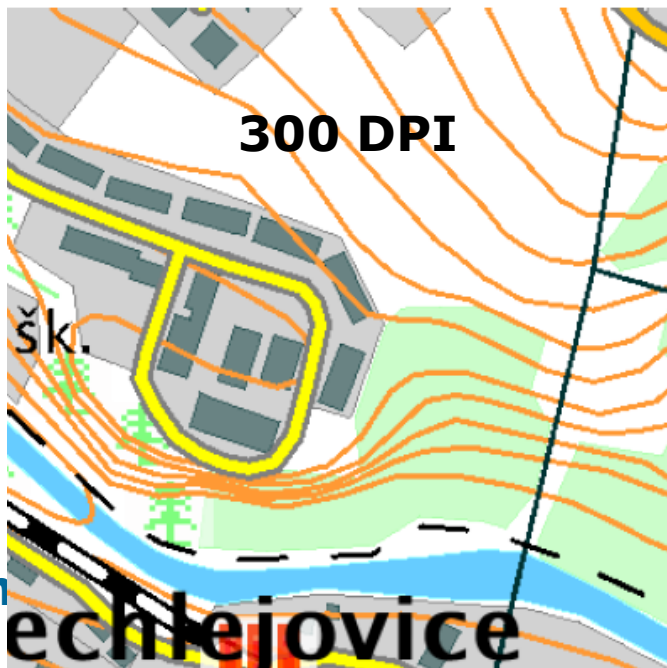
- **Malé finanční nároky**; digitizéry jsou relativně levné, pracovní síla je také levná.
- Flexibilita a adaptabilita na různé zdroje dat.
- Technika je snadno zvládnutelná v krátkém čase - lze se snadno naučit.
- Kvalita výstupů je víceméně vysoká.
- Digitizéry jsou velice přesné (přesnější než zdrojová data).
- Snadné úpravy digitalizovaných dat.

-

- Přesnost je limitována stabilitou vstupního média.
- Digitalizace je únavná a nudná, tudíž velice **náchylná k operátorovým chybám**.

- Stále rozšířenějším způsobem převodu dat z analogové do digitální (rastrové) formy.
- zařízeních sloužících k optickému snímání dokumentů.
- Typy skenerů:
  - Bubnové
  - Deskové (stolní)
  - Posuvné velkoformátové
  - 3D

- **Nejdůležitějšími hodnotícími ukazateli jsou:**
  - **optické rozlišení** (body na palec - Dots Per Inch, dpi),
  - **přesnost** - souvisí s tím, jak precizně je vyroben snímací senzor, tj. jak pravidelně jsou na něm umístěny snímací prvky,
  - **barevnost** či šedotónovost.



# Vektorizace

- **Automatická** - vše dělá počítač. Je to velice rychlé (co se týče nároků na uživatele), ale je nutné provádět čištění vektorových dat.
- **Polautomatická** - interaktivní metoda, s tím že počítač sám vektorizuje, ale uživatel jej koriguje na sporných místech (ArcScan).
- **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).



# Import geometrických dat

- **Soubory – binární/textové**
  - Souřadnice
  - CAD
  - Vektorová grafika
  - Rastrové soubory
  - GIS výměnné formáty
- **Databázové připojení**
- **Webové služby**
- **Senzory**



- **Textové soubory**
  - Oddělení mezerou, tabelátorem, čárkou
  - S hlavičkou nebo bez
  - Oddělovače řádků (závislé na OS)
  - Znaková sada (kvůli atributům) UTF/ASCII a rozšíření
- **Databázové výměnné soubory (.dbf)**
- **Spreadsheetové výměnné soubory (.xls)**

# Atributová data

## Způsoby vstupu do GIS:

- **Manuální**
- **Skenování + rozpoznávání textu (OCR)**
- **Převod z externích digitálních zdrojů**



# Manuální zadávání atributů

- Nejběžnější způsob zadávání atributových dat je manuálně, **pomocí klávesnice**, na což stačí pouze jednoduchý hardware.
- Možné problémy s **integritou** dat – lze kontrolovat.
- Atributy se následně navazují na prostorovou část pomocí **unikátního identifikátoru**, který prostorové prvky již obsahují (vytváří se obvykle již při jejich tvorbě).
- Kontrola **správnosti** zadaných údajů.
  - **Single Key Data Entry** - jeden operátor zadává atributová data a druhý operátor již zadaná data kontroluje (porovnává originál s vytištěnými výpisy, ...).
  - **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. Metoda se používá spíše na větší projekty, u kterých velice záleží na správnosti zadaných údajů.



# Skenování + 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).
- Problémy s diakritikou.
- Další nevýhodou je obvyklá **nutnost ručního navazování atributů** na prostorovou část, podobně jako u ručního zadávání dat.

# Převod z jiných zdrojů

- **Kritéria pro volbu vhodnosti či nevhodnosti 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.
  - **Tematický 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)?
  - **Kompatibilita datových modelů** - např. problematika převodu křivek při převodu z CAD do GIS nebo i z GIS do GIS, převod formátu atributů.
  - **Cena** - ...

# ZDROJE DAT?

# Chyby v datech

- 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

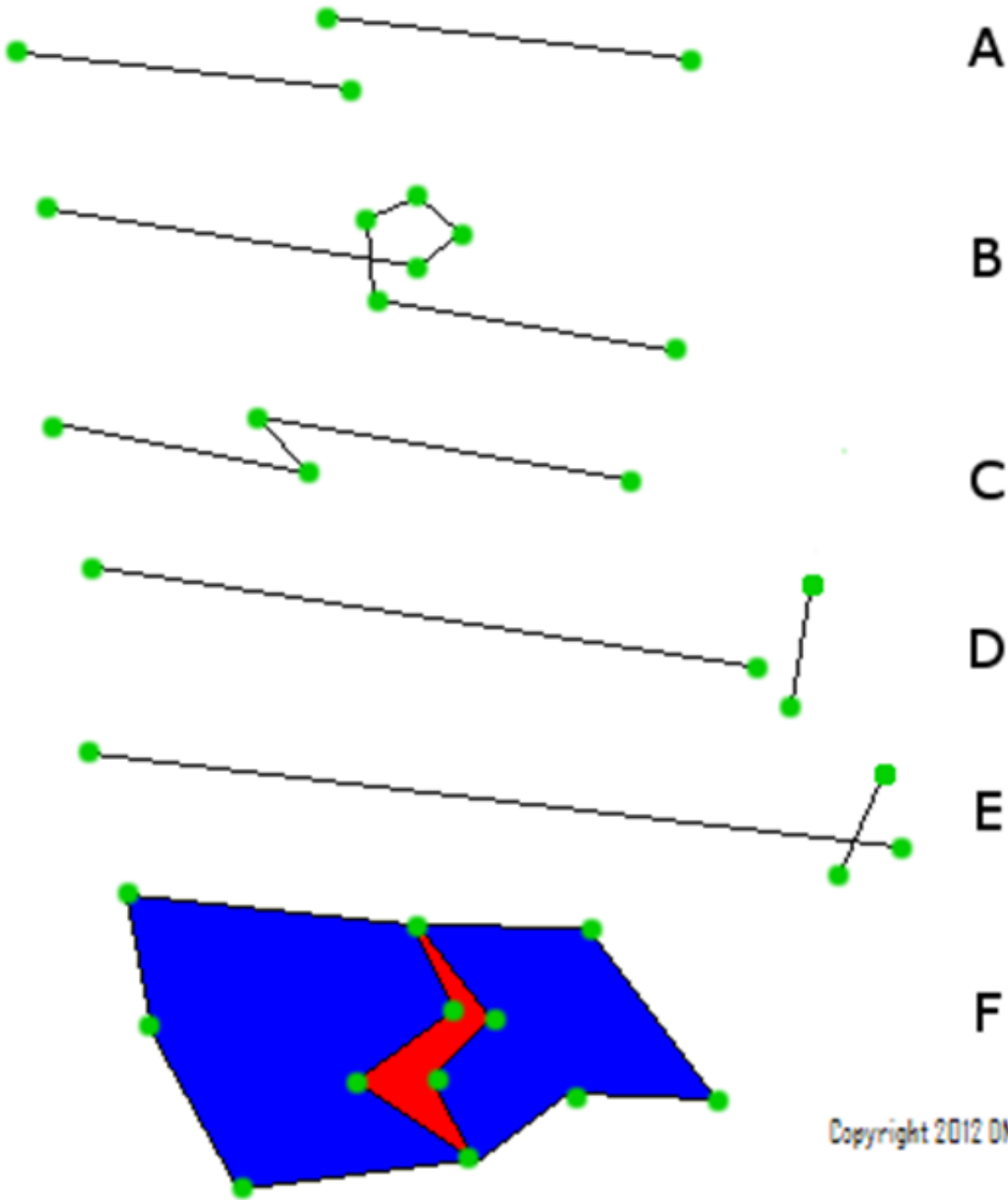
- **Nekompletnost dat** - scházejí body, linie, polygony.
- **Chybné umístění** prostorových dat - chyby vycházející ze špatné kvality vstupních dat nebo z nedostatečné přesnosti při digitalizování.
- **Zkreslení prostorových dat** - chyby z nepřesností vstupních dat (deformace podkladových dat, zkreslení 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.



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

# Obvyklé chyby v GIS geometrii



IIGC



# Topologické čištění dat

- **Jednotlivé úlohy**
  - **Eliminace duplikátních linií** (stejných i podobných).
  - **Odstraňování nedotahů** a přetahů.
  - **Nalezení průsečíků** dvou nebo více liniových prvků s následující segmentací.
  - **Odstranění mezer** (souvisí s nedotahy).
- **Topologicky čistá data** jsou taková data, nad kterými **je možné vytvořit topologii**, aniž by se jakkoli změnila jejich poloha.
- Pro tvorbu topologicky čistých dat se používají **topologické koncepty** (konektivita, definice plochy, sousednost).



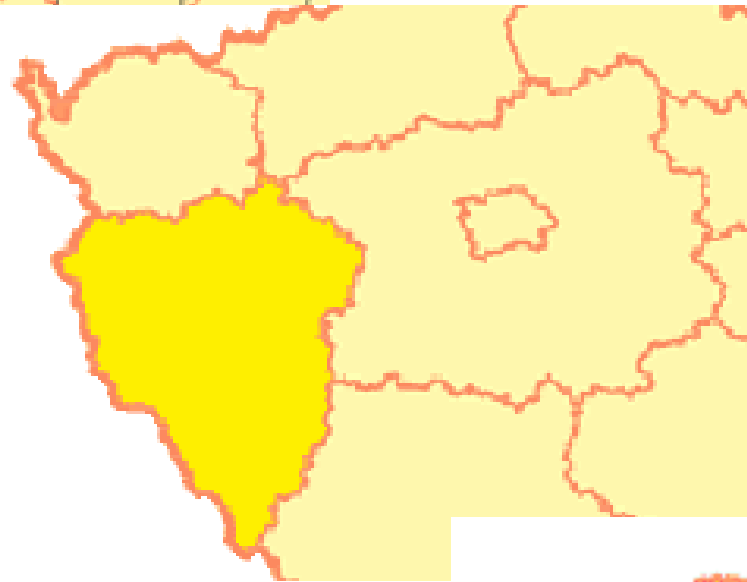
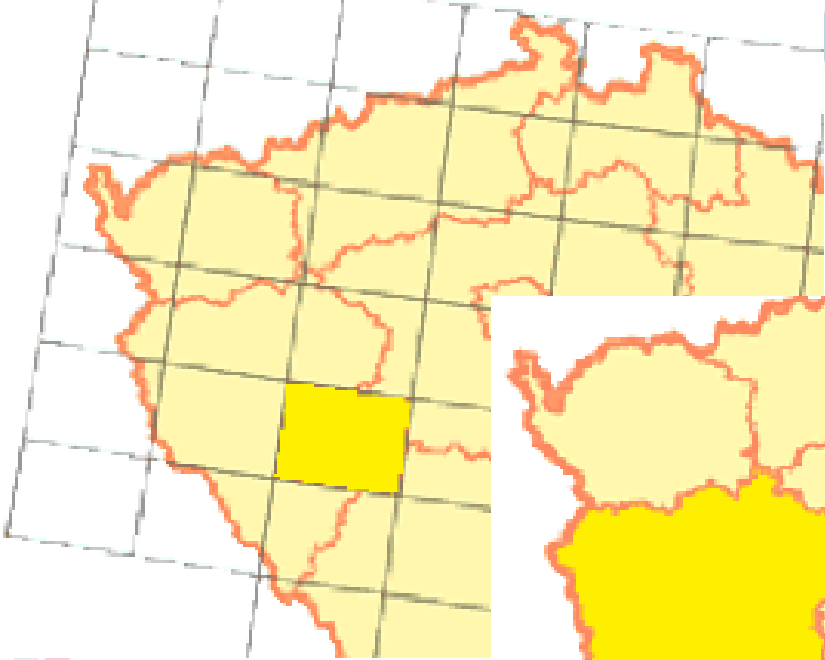
# Chyby právního charakteru

- **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.**
- **Zdroje obvykle přesně popisují možnosti využití a omezují zejména komerční či veřejné použití dat (i jako podkladu).**
- **Ochrana dat (vodotisk, záměrné chyby).**



# Uchovávání a zpracování dat

- **Pravidelné (např. mapové listy).**
  - Na disku je každý mapový list v jednom souboru (resp. ve více souborech se stejným jménem, lišících se pouze příponou) či adresáři.
- **Nepravidelné (mapové listy, zájmové území - katastrální území, území národního parku, okresu, kraje ...).**
  - Na disku je každé zájmové území v jednom souboru (resp. ve více souborech se stejným jménem, lišících se pouze příponou) či adresáři.
- **Bezešvé (Seamless)**
  - Celé zájmové území je uloženo v jednom souboru, adresáři či databázi).



**Geoinformatika**