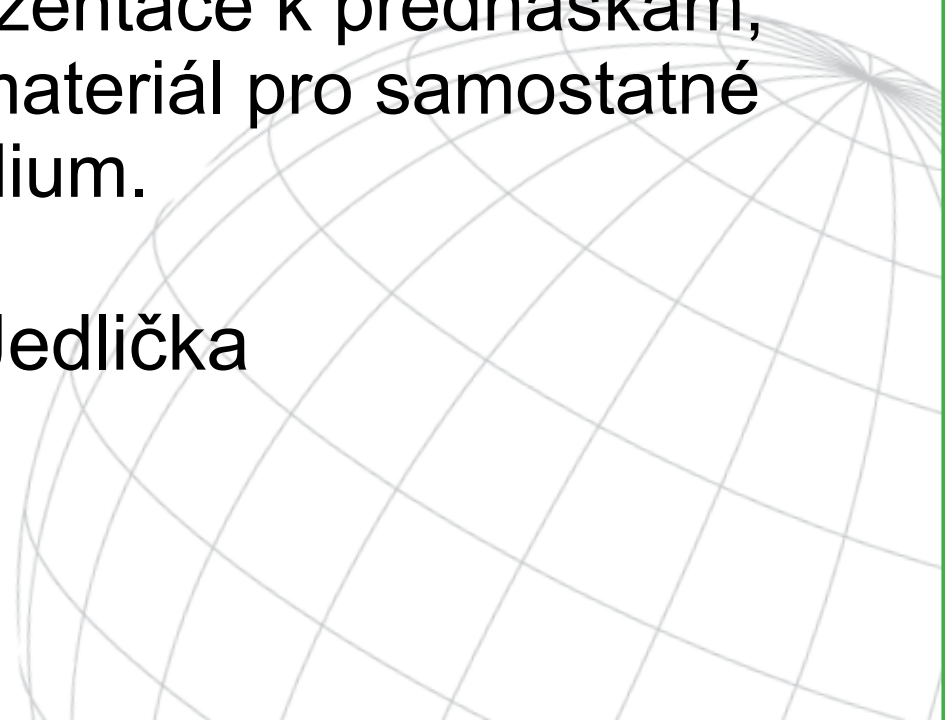


Úvod do GIS

Návrh databáze a vstup geografických dat II

Pouze podkladová prezentace k přednáškám,
nejedná se o studijní materiál pro samostatné
studium.

Karel Jedlička



Zdroje dat, konverze dat, metadata a uchování dat

- Přehled činností v projektu GIS
- Návrh geografické databáze
- Zdroje prostorových dat
 - primární
 - sekundární
- Zdroje atributových dat
- Konverze *geografických dat* z jiných digitálních zdrojů
- Možné chyby při vstupu dat
- Metadata
- Zpracování a uchování dat



Zdroje prostorových dat pro GDB

- 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.**

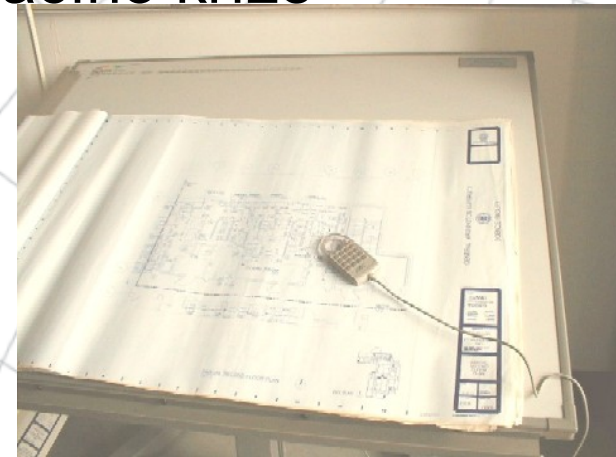
– Možnosti jejich vstupu do GIS:

- manuální vstup přes klávesnici,
- digitalizace,
- scannování a vektorizace.

Zdroje prostorových dat pro GDB

- 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 – 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.



Zdroje prostorových dat pro GDB

- 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.
 - Pozn: tyto metody byly následně převzaty m.j. i pro měření GPS.

Zdroje prostorových dat pro GDB

- Konkrétní 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.

Výsledná přesnost digitalizace závisí na měřítku a přesnosti původní mapy a také na preciznosti registrace. Více viz sekce **transformace!**

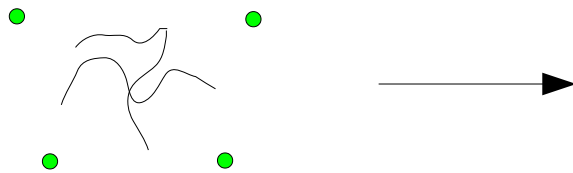
3. **Vlastní digitalizace** mapy (proudová/bodová metoda, zadávání id).

4. **Editace chyb** – nespojené čáry, nedotahy a přetahy, vícenásobné zaznamenání – souvisí s topologickým čištěním (viz. dále).

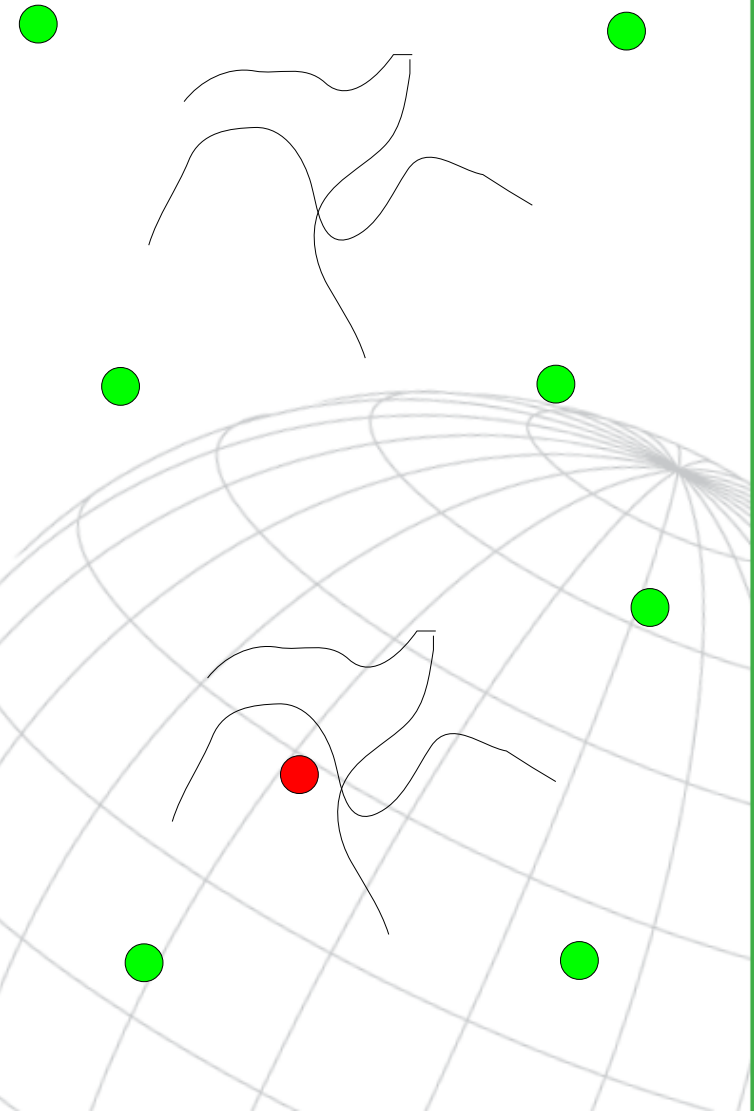
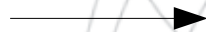
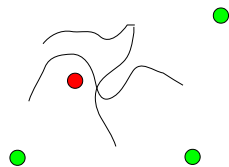
Zdroje prostorových dat pro GDB

- Volba identických bodů

– dobře



– špatně



Zdroje prostorových dat pro GDB

- Digitalizace

+

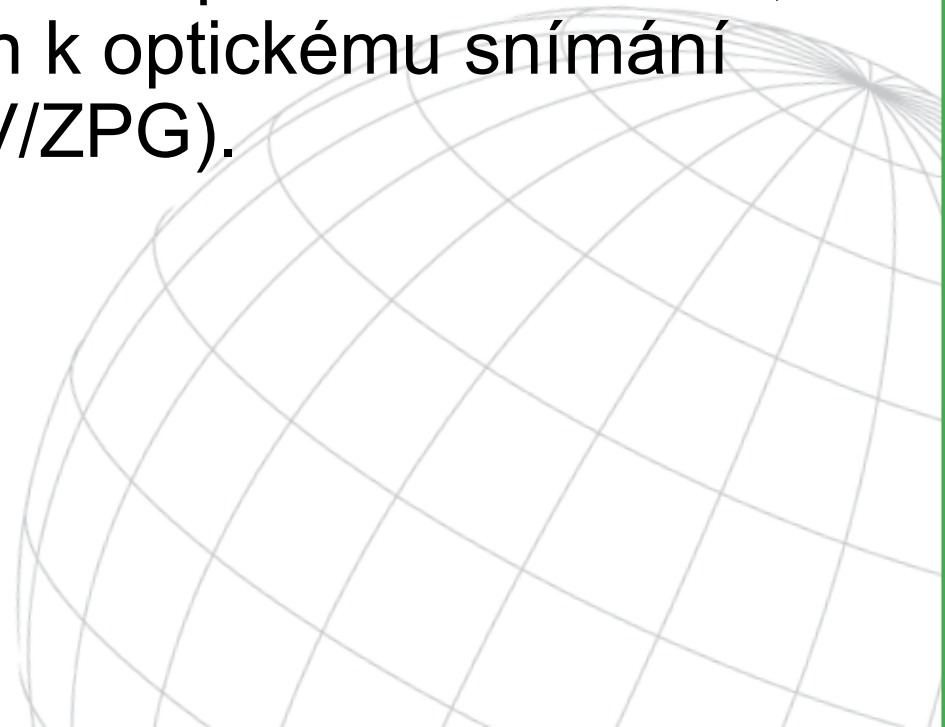
- **Malé finanční nároky**; digitizéry jsou relativně levné, pracovní síla je také levná.
- Flexibilita a adaptibilita na různé zdroje dat.
- Technika je snadno zvládnutelná v krátkém čase – lze se snadno naučit.
- Kvalita výstupů je víceméně vysoká.
- Digitizéry jsou velice přesné (přesnější než zdrojová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

-

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

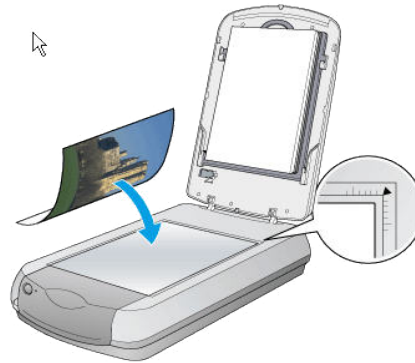
Zdroje prostorových dat pro GDB

- Scannování a (následná) vektorizace
 - Dnes nejrozšířenějším způsobem převodu dat z analogové do digitální (rastrové) formy je scannování. Vykonává se pomocí scannerů, zařízeních sloužících k optickému snímání dokumentů (více KIV/ZPG).



Zdroje prostorových dat pro GDB

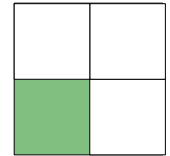
- Scannery
 - Bubnové
 - Deskové (stolní)
 - Posuvné velkoformátové
 - 3D ...



Zdroje prostorových dat pro GDB

- Hodnocení scannerů
 - Nejdůležitějšími hodnotícími ukazateli jsou:
 - **optické rozlišení** (body na palec – Dots Per Inch, DPI, 1" = 25,4 mm),
 - **přesnost** – souvisí s tím, jak precizně je vyroben snímací senzor, tj. jak pravidelně jsou na něm umístěny snímací prvky,
 - **barevnost** či šedotónovost.
- Poznámka: v GIS se používají jak scannery monochromatické (dvojúrovňové) nebo šedotónové, ale i barevné.

Zdroje prostorových dat pro GDB



- Konkrétní postup při scanování
 1. **Výběr rozlišení** – to je docela důležité rozhodnutí, jelikož platí, že dvakrát větší rozlišení vede ke čtyřnásobné velikosti výsledného souboru!
 2. **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.
 3. **Příprava mapy ke scannování** – očištění od mechanických nečistot, identifikace vlíčovacích bodů, případně úpravy nečitelných částí.
 4. Vlastní **scannování**.
 5. **Registrace** (transformace rastru) pomocí vlíčovacích bodů, probíhá analogicky jako u digitalizace.

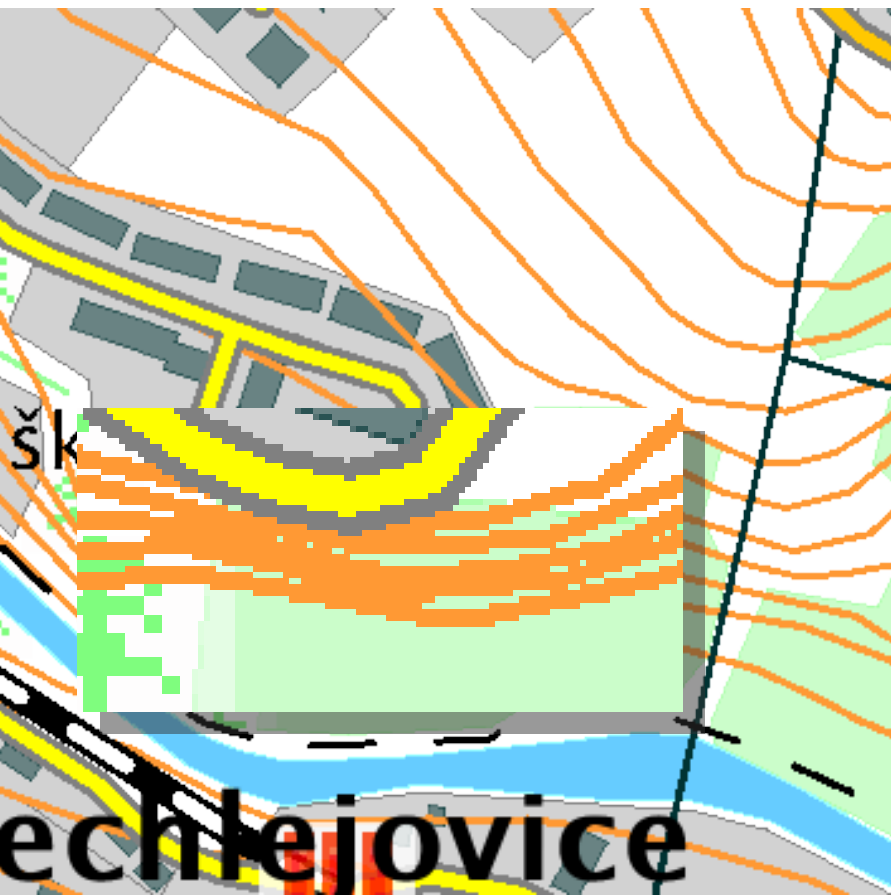
Zdroje prostorových dat pro GDB

- Volitelné kroky při skenování
 6. úprava obrazu (jas, kontrast, prahování, vyvážení histogramu).
 7. čištění rastrového podkladu.
 8. vektorizace.
- Volba rozlišení

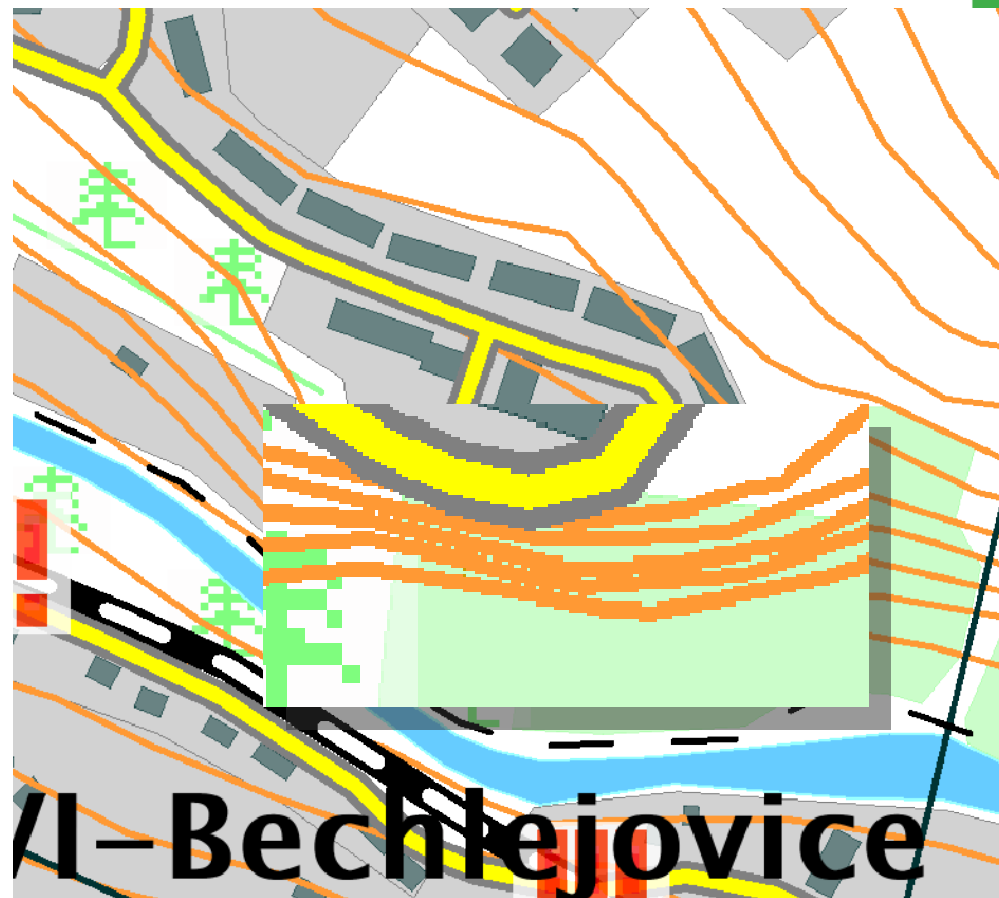
Rozlišení (DPI)	Typ scanovaných dat
800-1000	Ryté nebo leptané kontury (vrstevnice, isolinie). Linie, které mají mezeru mezi jednotlivými objekty menší než je jejich šířka. Linie s šířkou menší než 0,01 mm.
500-600	Ryté nebo leptané liniové objekty (řeky, silnice). Kreslené kontury.
400-500	Kreslené liniové objekty. Polygonové vrstvy (vegetace, druh půdy, ...).
300-400	Liniové objekty kreslené tužkou. Mapy na papírovém podkladu.
200-300	Šedotónové fotografie.
180-200	Barevné fotografie.

Zdroje prostorových dat pro GDB

- Volba rozlišení
300 dpi



- 600 dpi

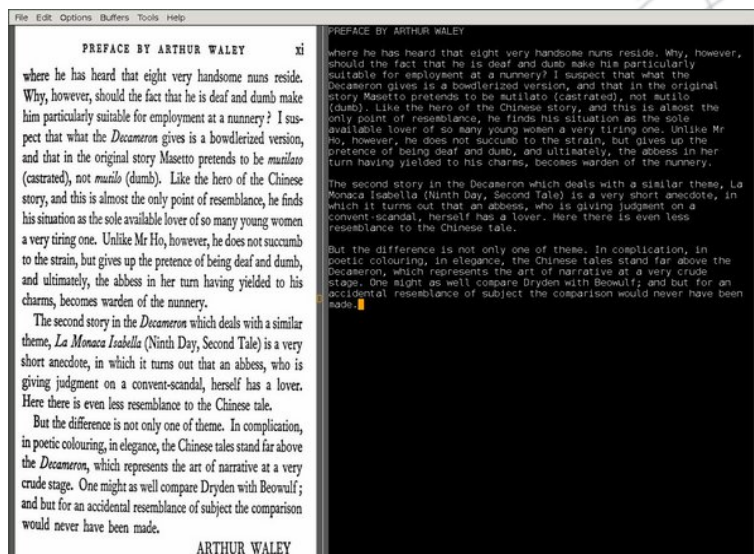


Zdroje prostorových dat pro GDB

- Princip vektorizace (více o ní až v části zabývající se zpracováním dat v GIS):
 - **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.
 - **Ruční** (on screen digitizing) – interaktivní, kdy uživatel provádí sám vektorizaci na základě rastrového podkladu. Některé systémy umožňují automatizovat alespoň přichycení na rastr.

Zdroje atributových dat

- Rozlišujeme podle způsobu vstupu
 - Manuální
 - Scannování + rozpoznávání textu (OCR)
 - Převod z externích digitálních zdrojů





Zdroje atributových dat

- Manuální

- Nejběžnější způsob zadávání atributových dat je manuálně, pomocí klávesnice, na což stačí pouze jednoduchý hardware.

Poznámka: při zadávání atributů může nastat **problém s doménovou integritou** (např. zadaný věk = -1), ta se ale většinou kontroluje již během zadávání.

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



Zdroje atributových dat

- **Manuální**

- 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.



Zdroje 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.



Zdroje atributových dat

- OCR

A screenshot of a text editor window showing the OCR output of a preface by Arthur Waley. The window has a menu bar with 'File', 'Edit', 'Options', 'Buffers', 'Tools', and 'Help'. The text is displayed in two columns. The left column shows the OCR output, and the right column shows the original text. The OCR output is slightly blurred and has some artifacts, while the original text is clear and well-formatted. The status bar at the bottom shows 'ocr-display All L1 (Fundamental)-----** ocr-edit All L26 (OCR)-----'.

PREFACE BY ARTHUR WALEY xi

where he has heard that eight very handsome nuns reside. Why, however, should the fact that he is deaf and dumb make him particularly suitable for employment at a nunnery? I suspect that what the *Decameron* gives is a bowdlerized version, and that in the original story Masetto pretends to be *castrated*, not *dumb*. Like the hero of the Chinese story, and this is almost the only point of resemblance, he finds his situation as the sole available lover of so many young women a very tiring one. Unlike Mr Ho, however, he does not succumb to the strain, but gives up the pretence of being deaf and dumb, and ultimately, the abbess in her turn having yielded to his charms, becomes warden of the nunnery.

The second story in the *Decameron* which deals with a similar theme, *La Monaca Isabella* (Ninth Day, Second Tale) is a very short anecdote, in which it turns out that an abbess, who is giving judgment on a convent-scandal, herself has a lover. Here there is even less resemblance to the Chinese tale.

But the difference is not only one of theme. In complication, in poetic colouring, in elegance, the Chinese tales stand far above the *Decameron*, which represents the art of narrative at a very crude stage. One might as well compare Dryden with Beowulf; and but for an accidental resemblance of subject the comparison would never have been made.

ARTHUR WALEY

PREFACE BY ARTHUR WALEY

where he has heard that eight very handsome nuns reside. Why, however, should the fact that he is deaf and dumb make him particularly suitable for employment at a nunnery? I suspect that what the *Decameron* gives is a bowdlerized version, and that in the original story Masetto pretends to be *castrated*, not *dumb*. Like the hero of the Chinese story, and this is almost the only point of resemblance, he finds his situation as the sole available lover of so many young women a very tiring one. Unlike Mr Ho, however, he does not succumb to the strain, but gives up the pretence of being deaf and dumb, and ultimately, the abbess in her turn having yielded to his charms, becomes warden of the nunnery.

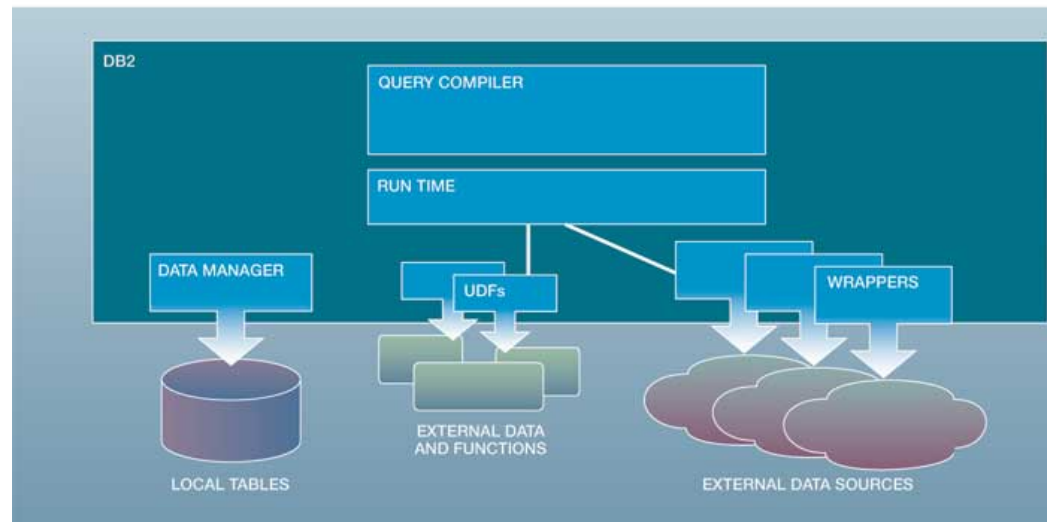
The second story in the *Decameron* which deals with a similar theme, *La Monaca Isabella* (Ninth Day, Second Tale) is a very short anecdote, in which it turns out that an abbess, who is giving judgment on a convent-scandal, herself has a lover. Here there is even less resemblance to the Chinese tale.

But the difference is not only one of theme. In complication, in poetic colouring, in elegance, the Chinese tales stand far above the *Decameron*, which represents the art of narrative at a very crude stage. One might as well compare Dryden with Beowulf; and but for an accidental resemblance of subject the comparison would never have been made.

Zdroje atributových dat

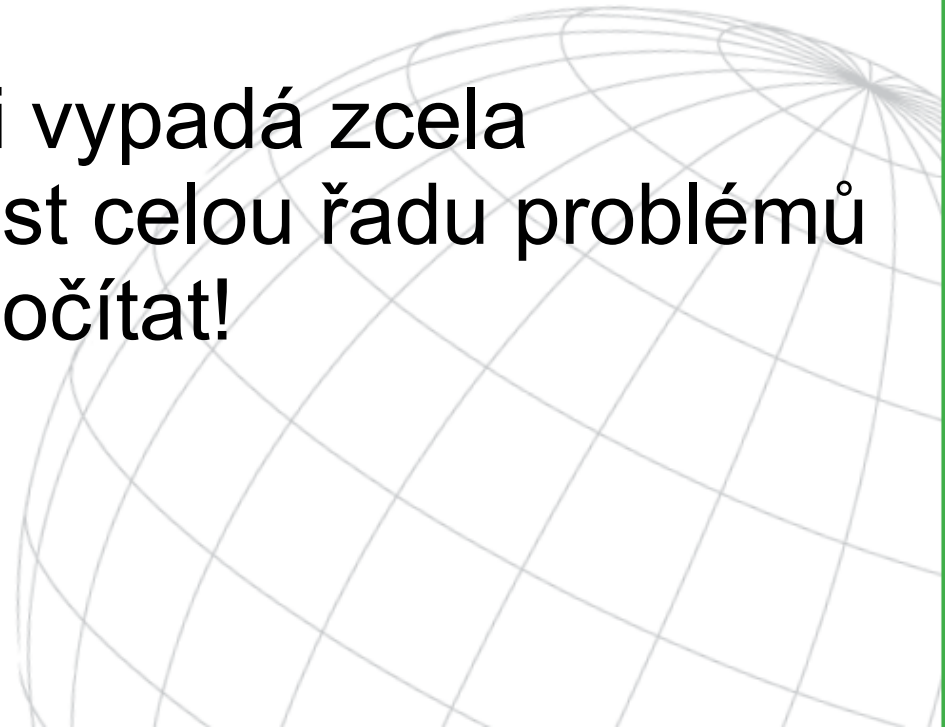
- Převod z externích digitálních zdrojů
 - Poslední možností je načítání (**nejen**) atributových údajů z jiných, již digitálních, zdrojů. Tato problematika se úzce dotýká konverze geografických dat z jiných digitálních zdrojů a je probrána podrobněji v následující části.

Figure 1 DB2 architecture for database federation



Konverze geografických dat

- 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!



Konverze geografických dat

- 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, datový 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** – ...

Konverze geografických dat

- Zdroje dat:

- Ve světě

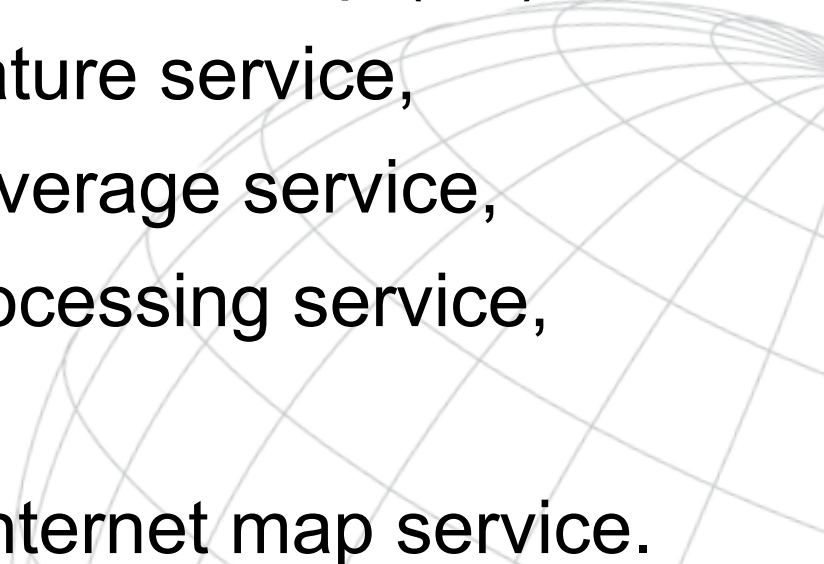
- <http://www.arcgis.com>
 - United States Geological Survey – <http://www.usgs.gov>
 - ...

- V ČR

- ČÚZK – <http://geoportal.cuzk.cz>
ZABAGED, DKM, KM-D
 - Armáda – DMÚ 25, DMÚ 200, DMR 1 a DMR 2
<http://izgard.cenia.cz/dmunew/viewer.htm>
 - <http://geoportal.cenia.cz>
 - Další – ÚHUL, VÚMOP, Geodézie ČS, ČGS, ARCDATA, ...
 - ... <http://mapy.kr-plzensky.cz>



Konverze geografických dat

- Klasické souborové formáty (viz dříve).
 - Formáty on-line dat (OGC OWS):
 - WMS a WMTS – web map (tile) service,
 - WFS – web feature service,
 - WCS – web coverage service,
 - WPS – web processing service,
 - ...
 - IMS – (ESRI) internet map service.
- 

Možné chyby při vstupu dat

- Při vkládání dat do systému **není možné zabezpečit 100% správnost zadání dat**. Následující řádky pouze konstatují možné chyby, které mohou nastat při tvorbě geografických dat. V části zabývající se zpracováním bude tato problematika rozebrána podrobněji a to zejména z hlediska opravování chyb.
- **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 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.

Možné chyby při vstupu dat

- 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**.

Možné chyby při vstupu dat

- 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!

