

Úvod do GIS

Atributy a jejich vztah k prostoru

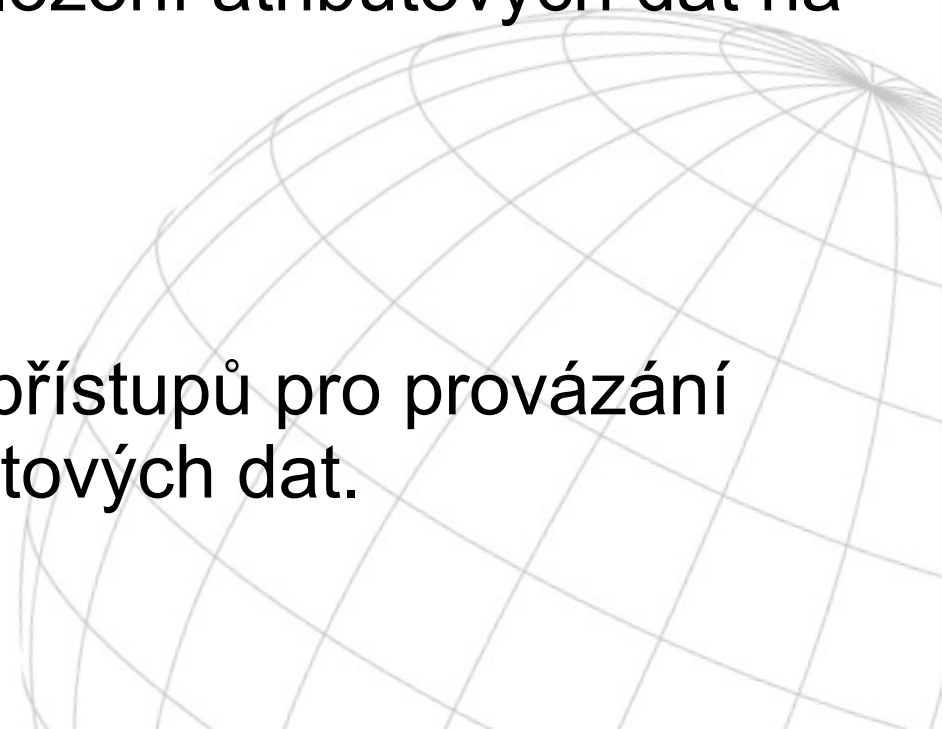
Pouze podkladová prezentace k přednáškám,
nejedná se o studijní materiál pro samostatné
studium.

Karel Jedlička



Atributy a jejich vztah k prostoru

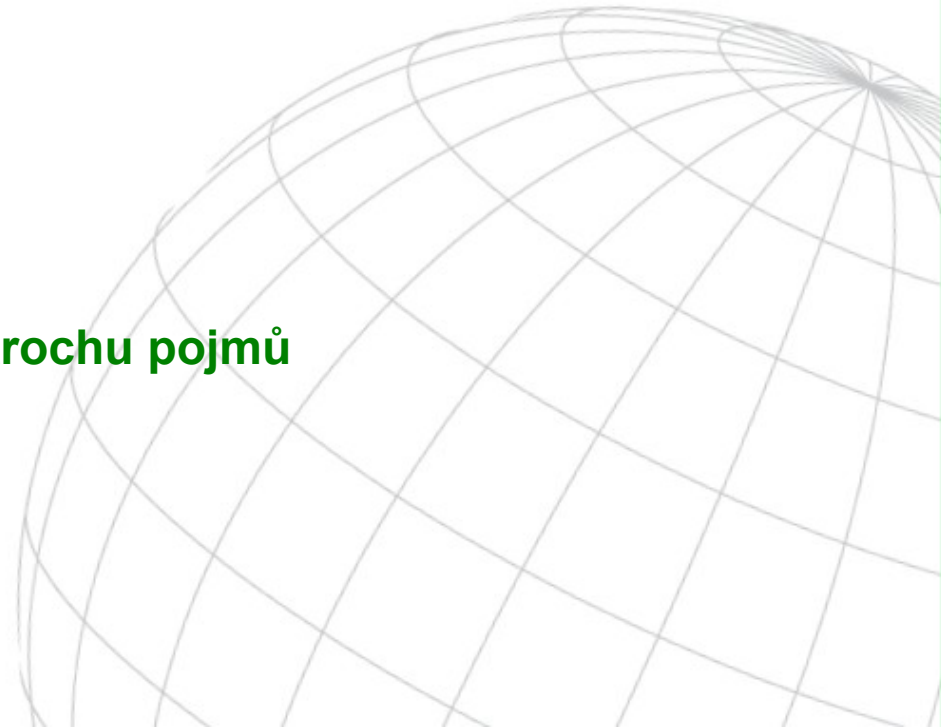
- Atributová data
 - Základní pojmy z klasických databází.
 - Možnosti přístupu k atributovým datům.
 - Datové modely pro uložení atributových dat na počítači.
- Rozdělení GIS
 - Generace GIS.
 - Srovnání vybraných přístupů pro provázání prostorových a atributových dat.



Základní pojmy z databází

- Atributová (popisná) data v digitální podobě
 - pro uchování popisných dat se v GIS používá separátního datového modelu, který vychází ze standardních modelů uchování dat v digitální podobě.

nejdříve trochu pojmů



Základní pojmy z databází

- **Databáze** je sdílená kolekce logicky uspořádaných dat (a popisu těchto dat – metadat), která je navržena tak, aby splňovala potřeby uživatele.
- **System řízení báze dat** (SŘBD, případně **Data Base Management System – DBMS**) je softwarový systém, který umožňuje definovat, vytvářet a udržovat databázi a který též poskytuje řízený přístup k této databázi.

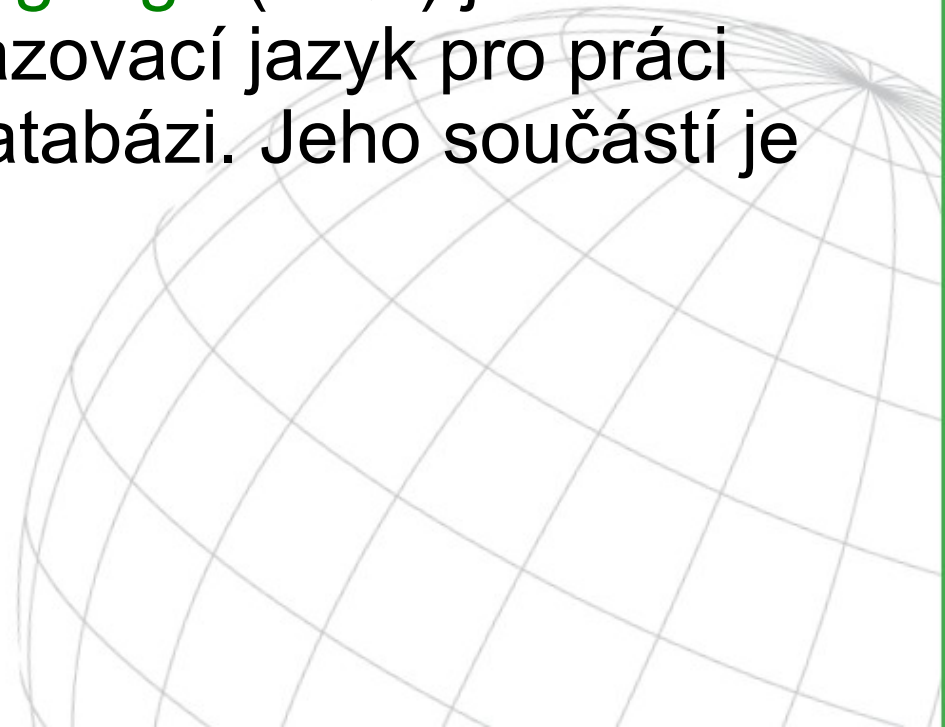
Základní pojmy z databází

- **Data Definition Language (DDL)** je jazyk pro definování datových struktur a způsobu jejich přístupu k nim.
- **Data Manipulation Language (DML)** je jazyk pro manipulaci s datovými strukturami.



Základní pojmy z databází

- Příklad implementace DML a DDL pro relační databáze:
 - **Structured Query Language (SQL)** je standardizovaný dotazovací jazyk pro práci s daty uloženými v databázi. Jeho součástí je jak DDL, tak i DML.



Základní pojmy z databází

Ukázka syntaxe SQL:

SQL-DDL:

```
CREATE TABLE jméno_tabulky
    (jméno_sloupce datový_typ [NULL | NOT NULL] [,...]);
```

```
INSERT INTO jméno_tabulky VALUES ( , , NULL, NULL);
```

SQL-DML:

```
SELECT [DISTINCT|ALL] {*[sloupec_vyraz [AS nove_jmeno]][,...]}
    FROM tabulka [alias] [,...]
    [WHERE podminka]
    [GROUP BY seznam_sl] [HAVING podminka]
    [ORDER BY seznam_sl]
```

... více viz např. předmět KIV/DB1

Základní pojmy z databází

- Object Query Language (**OQL**)
 - Pro dotazování v objektově orientovaných a objektových databázích (viz. objektově orientovaný model dále).
 - Vyvinut Object Data Management Group (ODMG).
 - Pro jeho celkovou složitost zatím nebyl nikde plně implementován, nicméně mírně ovlivnil návrh novějších dotazovacích jazyků JDOQL a EJBQL.

Základní pojmy z databází

Common/Contextual Query Language (CQL):

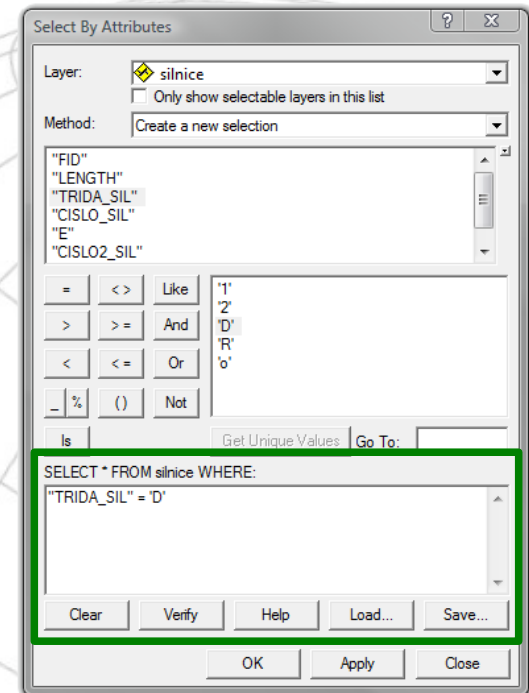
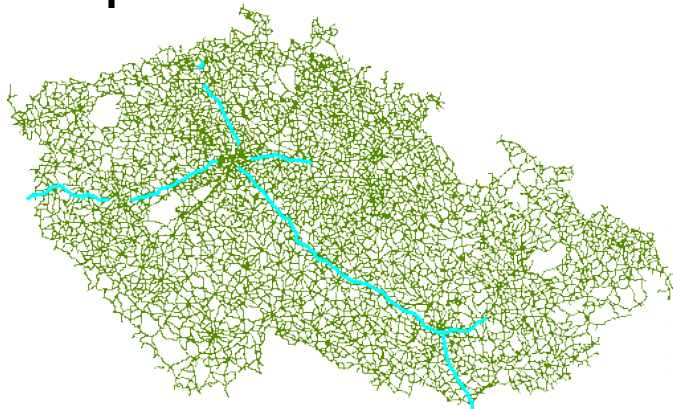
- Formální jazyk pro dotazování a prohledávání webu, bibliografických i jiných katalogů, atp.
- Intuitivnější než SQL.
- V GIS se zatím příliš nepoužívá.



Základní pojmy z databází

Query By Example (QBE):

- Způsob převodu uživatelského vstupu do formálního dotazovacího jazyka databáze (CQL/**SQL**).
- Jednodušší, ale neumožní zapsat vše.
- Často používaný v GIS GUI (Graphic User Interface)



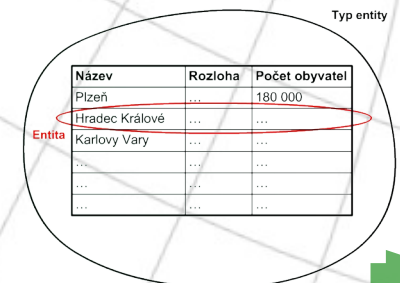
Základní pojmy z databází

- **Typ (třída) entity** je skupina entit se stejnými charakteristikami, například krajská města. *Lze si ji představit jako tabulku.*
- **Entita** je individuální element reprezentující reálný objekt, například Plzeň, Karlovy Vary, Pardubice... . *Představme si je jako řádky v tabulce.*

Typ entity

Název	Rozloha	Počet obyvatel
Plzeň	...	180 000
Hradec Králové
Karlovy Vary
...
...
...

Entita



Základní pojmy z databází

Typ entity

Entita

Název	Rozloha	Počet obyvatel
Plzeň	...	180 000
Hradec Králové
Karlovy Vary
...
...
...

Základní pojmy z databází

- Možná spojení mezi entitami – vztahy (**relace**):
 - **1:1** – Tomáš Marný má bydliště v Plzni.
 - **1:N** – Tomáš Marný má bydliště v Plzni ale protože vyhrál v loterii, koupil si ještě byt v Praze a chalupu na Slapech.
 - **M:N** – V Plzni bydlí více lidí než jen Tomáš Marný.
 - Mluvíme o kardinalitě/násobnosti relace/vztahu/vazby.

Základní pojmy z databází

- **Transakce** je posloupnost operací nad objekty báze dat, která realizuje **jednu ucelenou operaci** z pohledu uživatele. Pro názornost uveďme jako příklad transakce převod peněz z jednoho bankovního účtu na druhý.



Základní pojmy z databází

- Zadání: převed'te z účtu X 5000 Kč na účet Y.

Postup A:

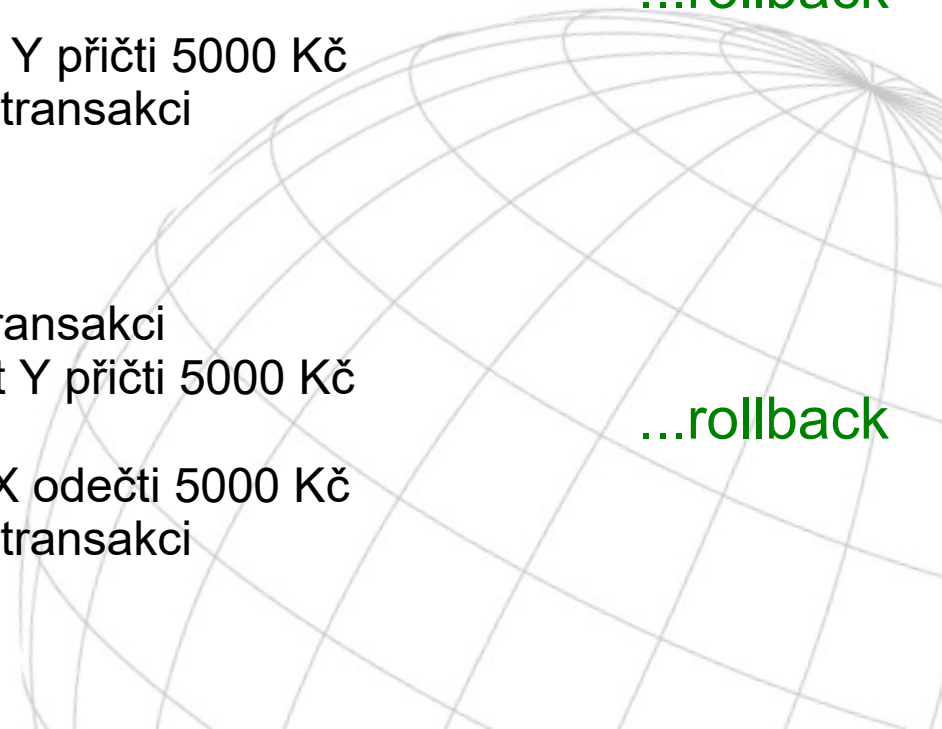
- (i) Zahaj transakci
- (ii) Z účtu X odečti 5000 Kč
- (iii) -----
- (iv) na účet Y přičti 5000 Kč
- (v) Ukonči transakci

...rollback

Postup B:

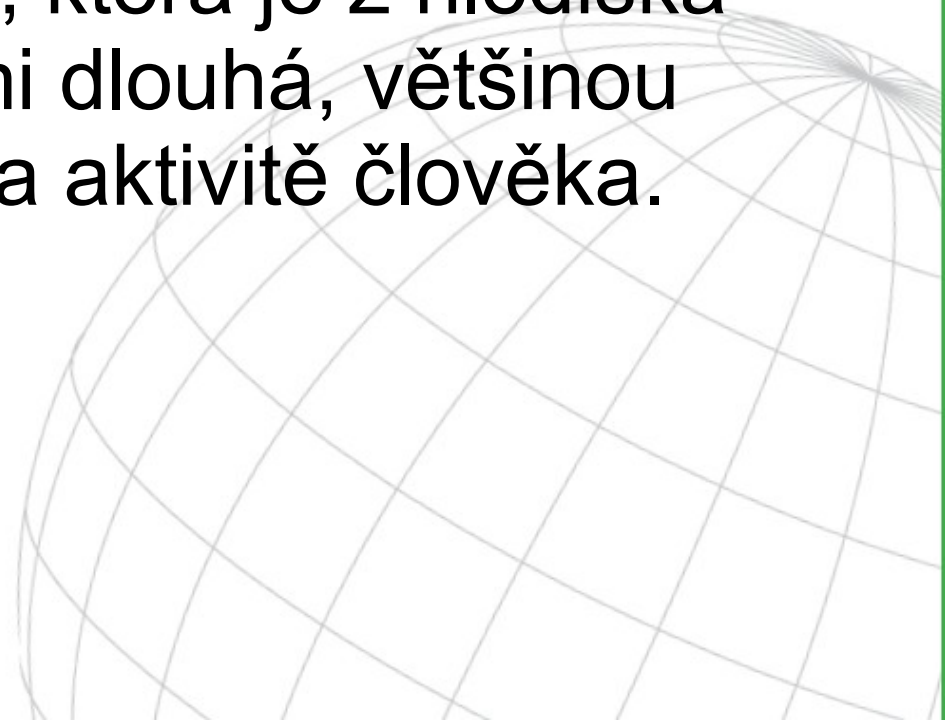
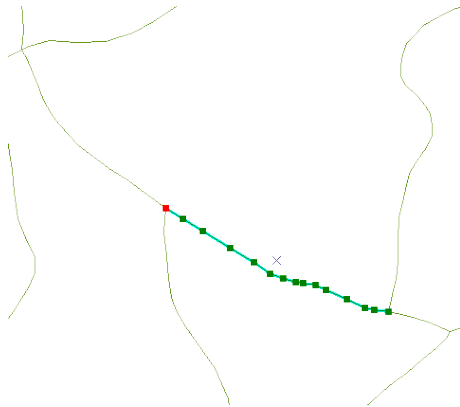
- (i) Zahaj transakci
- (ii) Na účet Y přičti 5000 Kč
- (iii) -----
- (iv) Z účtu X odečti 5000 Kč
- (v) Ukonči transakci

...rollback



Základní pojmy z databází

- **Dlouhá transakce** je již pojem používaný hlavně ve spojení s geografickými daty. Jedná se o transakci, která je z hlediska strojového času velmi dlouhá, většinou proto, že je závislá na aktivitě člověka.




Možnosti přístupu k atributům

- Možnosti přístupu k atributovým datům:
 - Přímý přístup (přímé zpracování souborů).
 - Přístup pomocí SŘBD.



Možnosti přístupu k atributům

- Přímý přístup
 - txt soubor
 - xls soubor
 - dbf soubor
 - ...



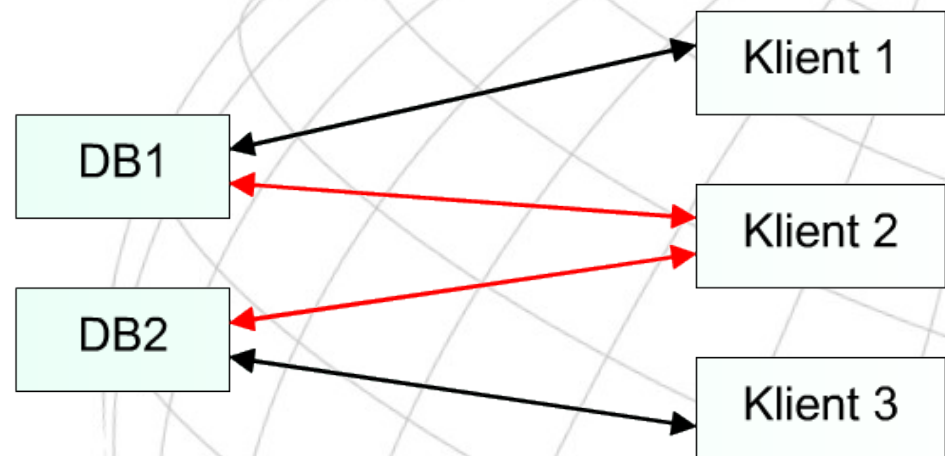
<i>Shape</i>	<i>Id</i>	<i>Mesta</i>	<i>Velikost</i>
Point	1	PLZEŇ	2
Point	2	Klatovy	1
Point	3	Sokolov	1
Point	4	K. Vary	1
Point	5	Cheb	1
Point	6	Fr. Lázně	0
Point	7	M. Lázně	0

Možnosti přístupu k atributům

- Přímý přístup

Ukázka konfliktu u přímého přístupu k datům

Uživatel 2 začíná pracovat na souborech, na kterých již někdo pracuje - **nastává konflikt**

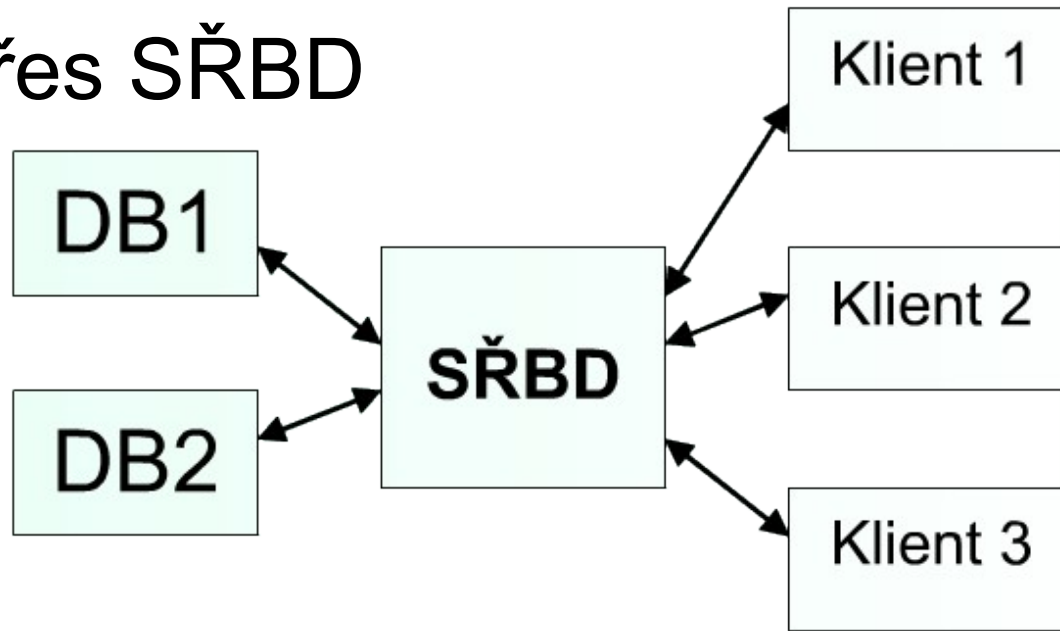


Možnosti přístupu k atributům

- Přímý přístup
- Nevýhody
 - Nebezpečí konfliktu.
 - Každá aplikace používající tabulku musí znát vnitřní strukturu uložení (na binární úrovni).
 - Nelze podporovat transakce.
 - Nelze používat vztahy mezi tabulkami.
 - Špatně se udržuje **integrita**(správnost) **dat**.

Možnosti přístupu k atributům

- Přes SŘBD



- Odstraňuje nevýhody přímého přístupu.
- Standardní přístup k tabulkám (SQL, ODBC)
- Komplexnější řešení, složitější na instalaci, správu, údržbu, ...
- Zajišťuje **integritu**.

Možnosti přístupu k atributům

- Integrita (konzistence) databáze znamená, že **databáze vyhovuje zadaným pravidlům** – integritním omezením.
 - Integritní omezení se mohou týkat jednotlivých hodnot vkládaných do polí databáze (**například známka z předmětu musí být v rozsahu 1 až 4**), či může jít o podmínku na kombinaci hodnot v některých polích jednoho záznamu (**například datum narození nesmí být pozdější než datum úmrtí**).
 - Integritní omezení jsou součástí definice databáze, a za jejich splnění zodpovídá systém řízení báze dat.

Možnosti přístupu k atributům

- Integritní omezení
 - Entitní
 - **Specifikace unikátního identifikátoru záznamu** ~ primárního klíče v rámci tabulky.
 - **Primární klíč** je atribut, či minimální seznam atributů, které jednoznačně určují ntici (řádek) tabulky.
 - Doménová
 - Doménová integrita znamená, že **na úrovni sloupců** definujeme omezení na určitý datový typ, případně **omezení rozsahu hodnot** (viz známky ve škole).
 - Referenční
 - **Mezitableková** – definuje vztah dvou tabulek pomocí cizích klíčů (primární klíč jedné tabulky umístěn do druhé). Tomáš Marný „opustil databázi“, a tak se smažou i všechna jeho bydliště.

Datové modely pro atributy

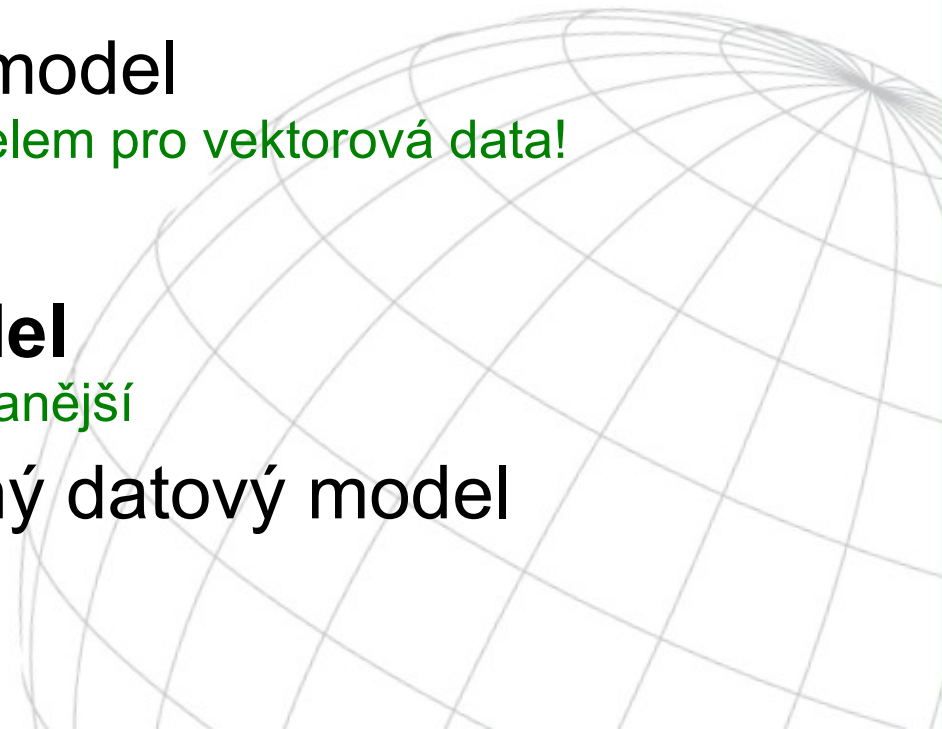
- Souborový (přímý přístup) x SŘBD
- Přímý přístup
 - Plochý model (jedna tabulka)



<i>Shape</i>	<i>Id</i>	<i>Mesta</i>	<i>Velikost</i>
Point	1	PLZEŇ	2
Point	2	Klatovy	1
Point	3	Sokolov	1
Point	4	K. Vary	1
Point	5	Cheb	1
Point	6	Fr. Lázně	0
Point	7	M. Lázně	0

Datové modely pro atributy

- ukládáme-li data přes SŘBD (DBMS), ukládáme data v některém z následujících datových modelů:
 - Hierarchický datový model
neplést s hierarchickým modelem pro vektorová data!
 - Síťový datový model
 - **Relační datový model**
pro atributová data nejpoužívanější
 - Objektově orientovaný datový model

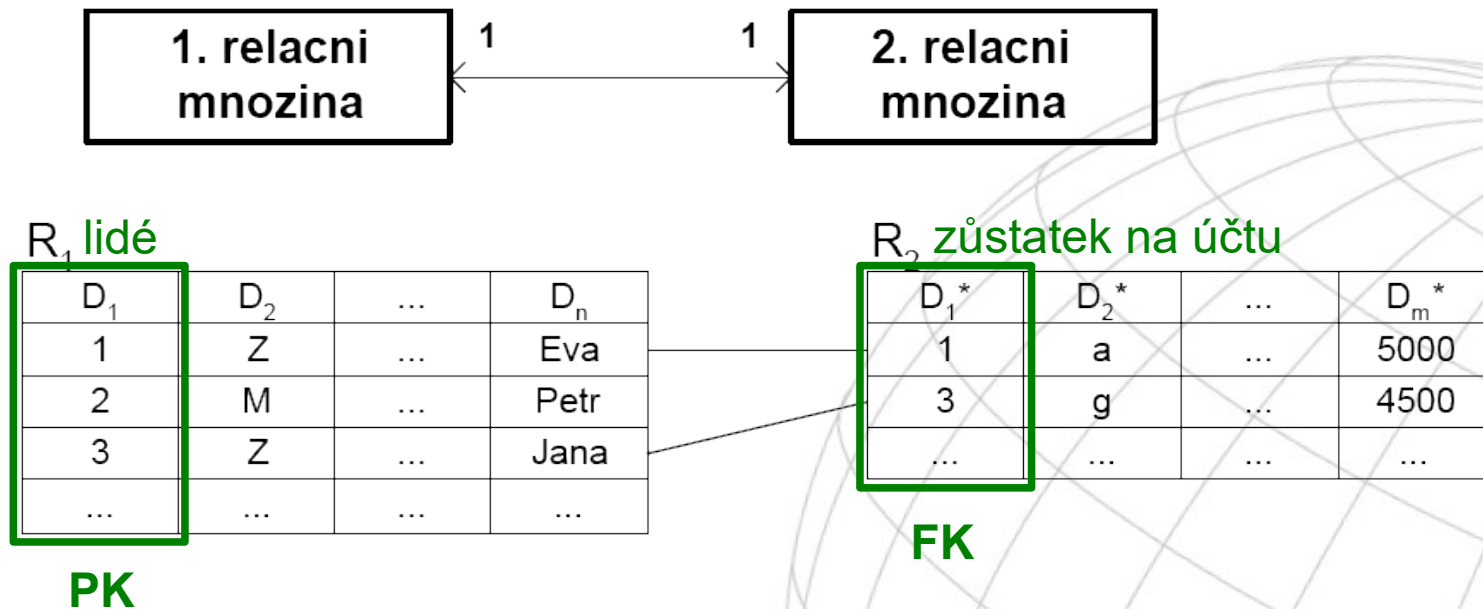


Datové modely pro atributy

- Relační datový model
 - organizuje data do tabulek,
 - v každé tabulce je jeden sloupec **primárním klíčem (PK)**,
 - PK umožňují vytváření vazeb (relací) mezi tabulkami (tím, že jsou do připojovaných tabulek vkládány jako cizí klíče),
 - pro odstranění nadbytečného (redundantního) ukládání jednoho záznamu se používá normalizace databáze (více v KIV/DB1).
 - pro popis struktury používá **ERA model**.

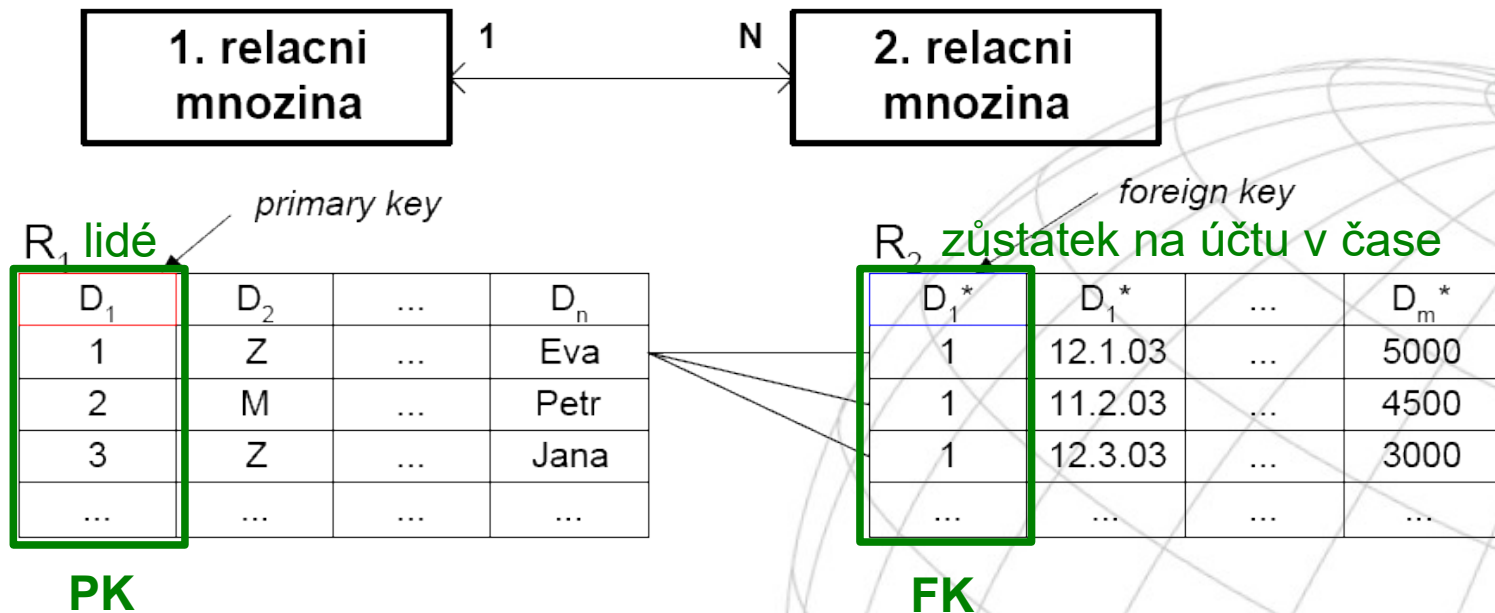
Datové modely pro atributy

- ERA model je založen na vazbách a atributech
1 : 1



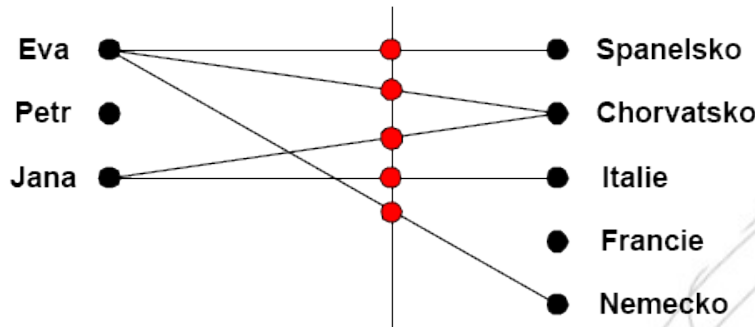
Datové modely pro atributy

- ERA model je založen na vazbách a atributech
1 : N



Datové modely pro atributy

- ERA model je založen na vazbách a atributech
M : N



R_1 lidé

D_1	...	D_n
1	...	Eva
2	...	Petr
3	...	Jana
...

PK

1	z1
1	z2
1	z5
3	z2
3	z3

FK1 FK2

R_2 navštívené země

D_1^*	D_1^*	...	D_m^*
z1	26.7.03	...	Spanelsko
z2	22.5.03	...	Chorvatsko
z3	20.12.03	...	Italie
z4	5.6.03	...	Francie
z5	10.9.03	...	Nemecko

PK

Datové modely pro atributy

- Relační datový model

+

- 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 (Tabulkové procesory – např. Excel, statistický software, ...),
- **díky výkonu, standardizaci a rozšíření se dnes jedná o nevyužívanější model v GIS.**

Datové modely pro atributy

- Relační datový model
 - - není vhodný pro příliš složité datové konstrukce, tam pak nastupují objektové technologie, resp. kombinace relačního a objektového přístupu – viz dále.

- ... další zdroje: <http://gis.zcu.cz/studium/dp/2003/> (Vokounová Lucie)



Datové modely pro atributy

- Objektově orientovaný model
 - data spravována jako **objekty**, což více přibližuje model reálnému světu,
 - vychází z objektově orientovaného programování,
 - pro každý objekt, jsou popisovány nejen jeho **vlastnosti** (atributy), ale i způsob jeho **chování** (metody),
 - objekty spolu komunikují pomocí **zpráv**.

Objekt1
-Vlastnost1
-Vlastnost2
-Vlastnost3
+Metoda1()
+Metoda1()
«signal»-Zprava1()

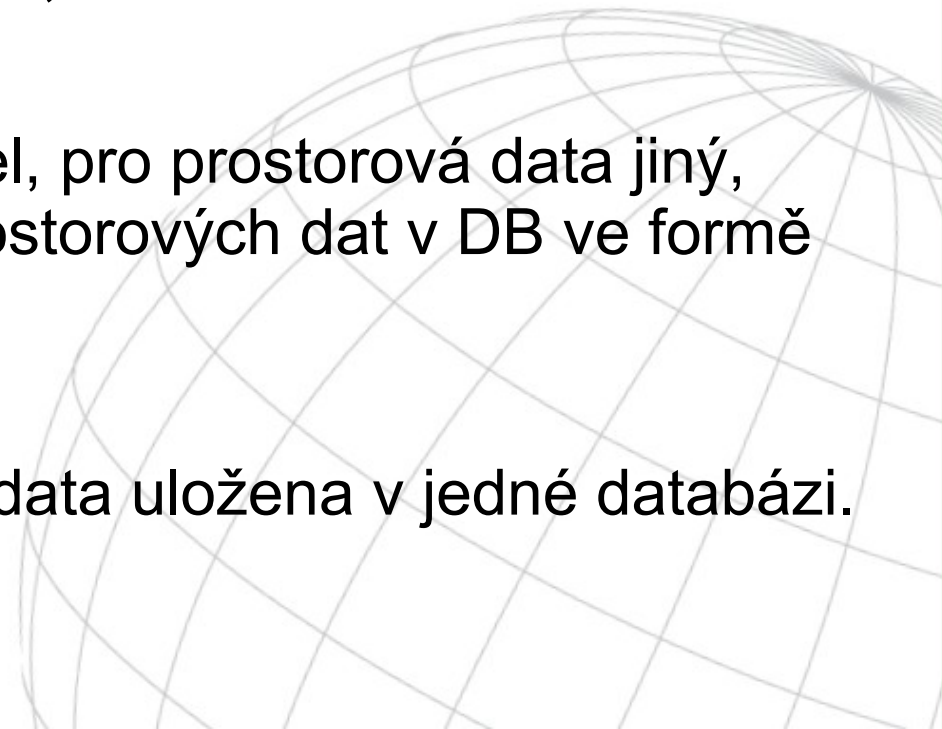
GeografickyObjekt
-GeometrieJakoVlastnost1
-Vlastnost2
-Vlastnost3
+Metoda1()
+Metoda1()
«signal»-Zprava1()

Datové modely pro atributy

- Po vyhodnocení všech výhod a nevýhod objektového a relačního modelu nakonec vznikl **objektově-relační model**, který zachovává výhody relačního modelu a přidává výhody objektového modelu. Dnes je objektově-relační model možno použít u většiny velkých databází (Oracle, Informix, SDE Geodatabase, ...).
- Viz. 3. generace GIS dále.

Rozdělení GIS podle generací

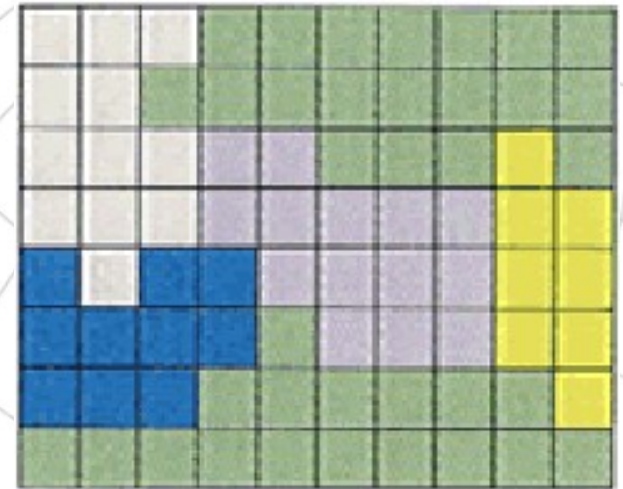
- jedná se o rozdělení na základě **vztahu atributových a prostorových dat.**
 1. generace:
 - nevelká podpora atributů,
 2. generace:
 - pro atributy jeden model, pro prostorová data jiný, případně ukládání prostorových dat v DB ve formě „BLOB“,
 3. generace:
 - atributová i prostorová data uložena v jedné databázi.



Rozdělení GIS podle generací

1. generace

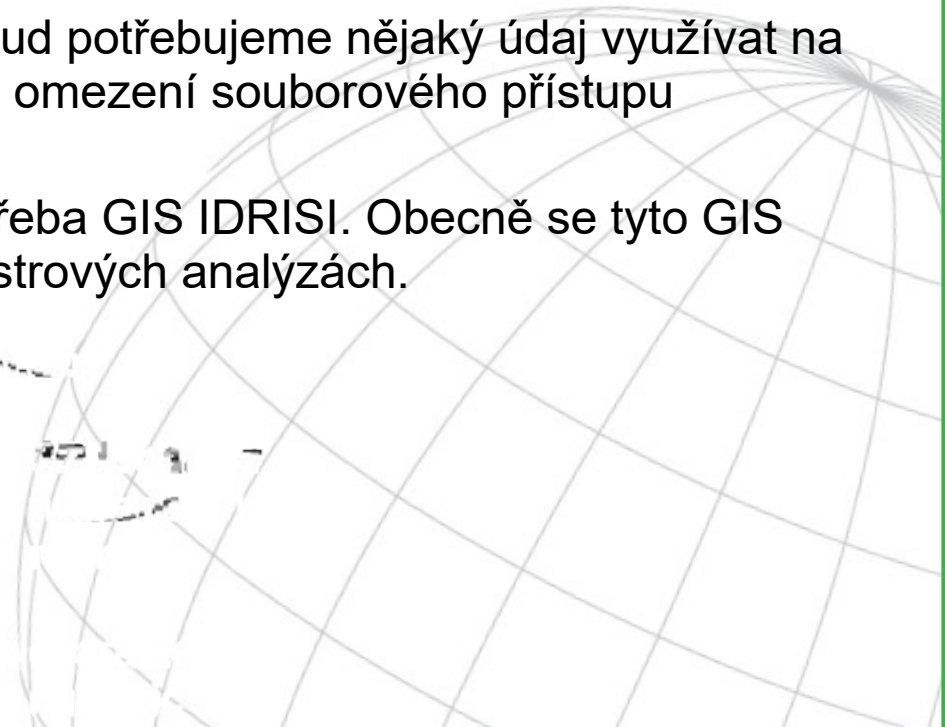
- Systém **bez atributových souborů**:
 - Hodně se používá u čistě rastrových systémů - není oddělen prostorový a popisný údaj.
 - Jediným atributem je **hodnota buňky**.
 - Např. GRASS ve svých počátcích.



Rozdělení GIS podle generací

1. generace

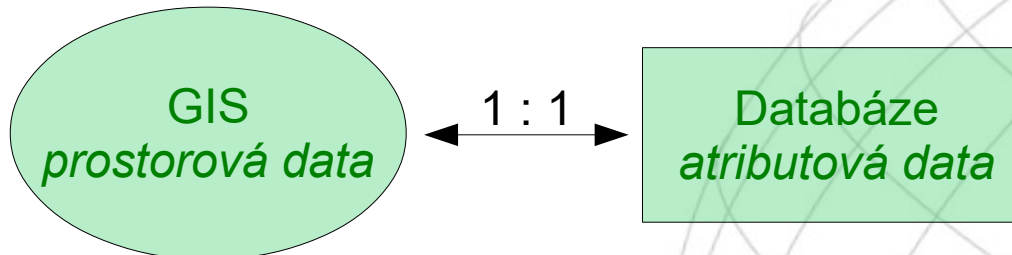
- „Flat“ soubory – stručný popis:
 - V podstatě o tabulky bez možnosti relací, využívají souborově orientovaný systém.
 - Ke každému objektu je možné přiřadit pouze 1 tabulku (spojení prostorových a atributových dat je provedeno přes id_objektu, které funguje jako klíč).
 - Tento způsob je nepraktický, pokud potřebujeme nějaký údaj využívat na více místech. Problém pramení z omezení souborového přístupu k datům.
 - Příkladem takového systému je třeba GIS IDRISI. Obecně se tyto GIS využívají dodnes převážně při rastrových analýzách.



Rozdělení GIS podle generací

2. generace

- **Duální/hybridní způsob (model)** propojení prostorových a atributových dat
 - Prostorová složka je zpracovávána jednom systému, atributy v DBMS někde jinde.
 - Problémy s **integritou** (konzistencí) dat.
 - Neřeší problematiku dlouhých transakcí.



- Příklad **dřívější** verze ARC/INFO – jeden systém (ARC) zpracovává grafiku a druhý (INFO) se stará atributy.

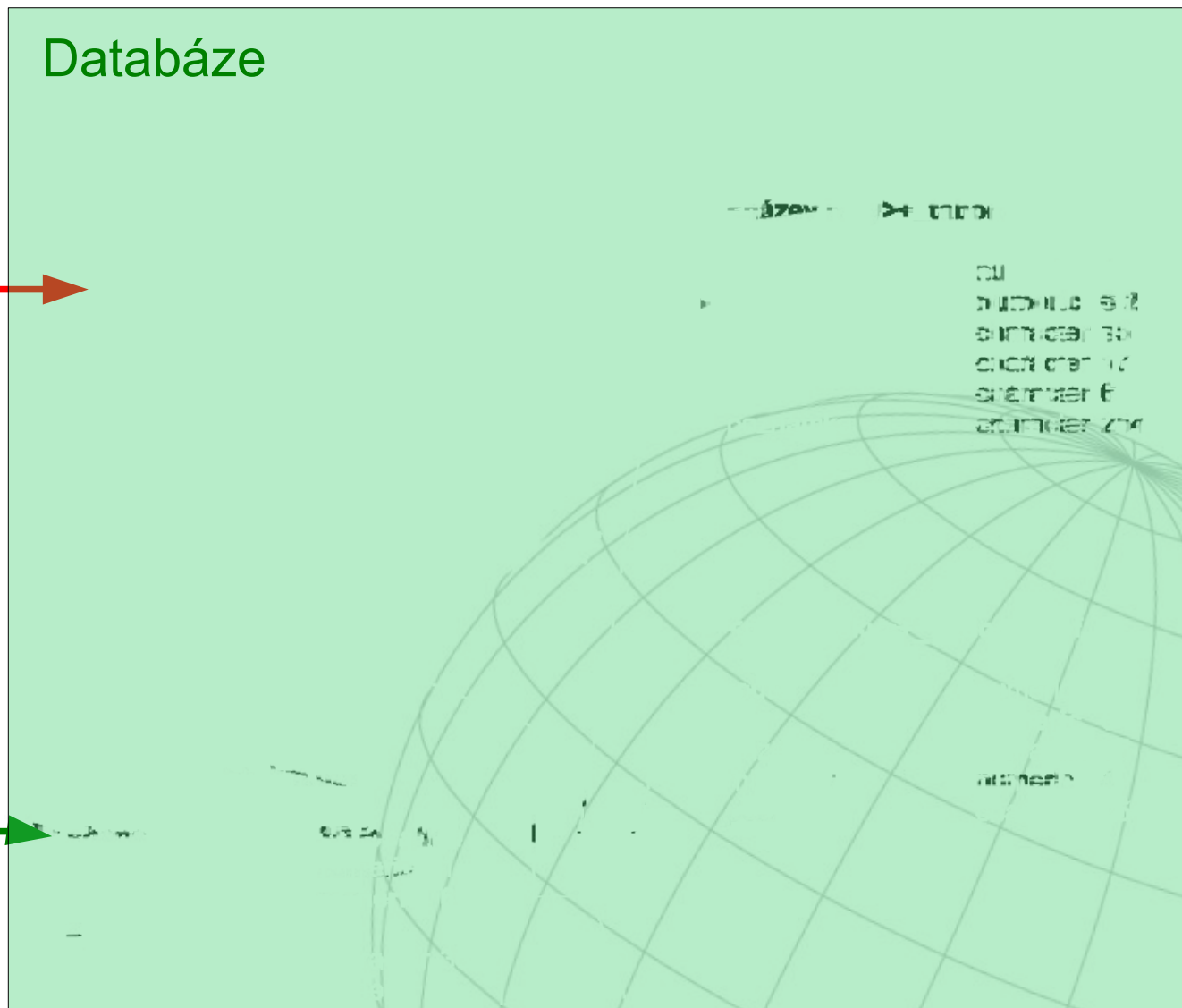
Rozdělení GIS podle generací

2. generace

Duální/hybridní model

Přímý přístup k atributům

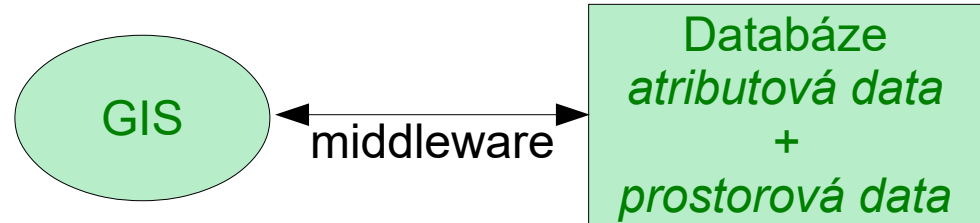
Přístup přes GIS



Rozdělení GIS podle generací

2. generace

• Integrovaný model



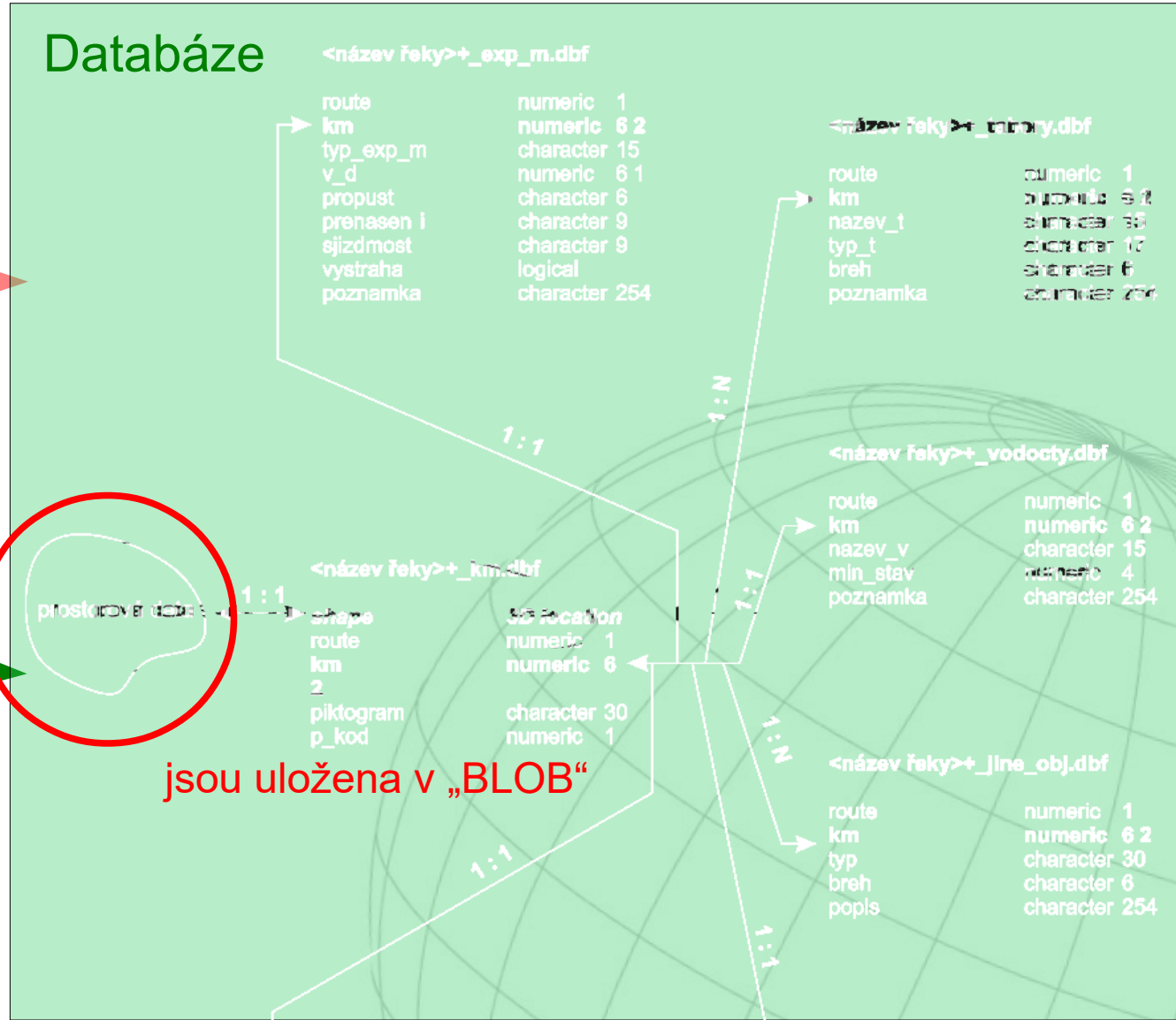
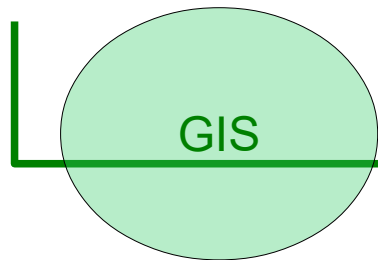
- Vše uloženo v jedné databázi (jak prostorová, tak popisná složka), ale využíván je pouze standardní relační model.
- O management všech geodat se stará tzv. **middleware**, což je produkt tvořící komunikační vrstvu mezi databází a GIS, nikoli tedy databáze samotná.
- Díky uložení prostorové části v databázi je možné používat bezešvá prostorová data a odpadá dělení prostorou na mapové listy.
- Model je relativně pomalý (standardní relační databáze neumí efektivně ukládat prostorová data).
- Pokud data nejsou uložena podle standardu pro ukládání prostorových dat v databázi jsou prostorová data jsou uložena jako binární posloupnosti (Binary Large Objects - BLOBs), což znemožňuje data zpracovávat již v DBMS či pomocí jiného SW.
- Model neodstraňuje problémy s dlouhými transakcemi a integritou dat.
- Příkladem je ArcGIS a Personal Geodatabase.

Rozdělení GIS podle generací

2. generace Integrovaný model

Přímý přístup
k atributům

Přístup přes GIS

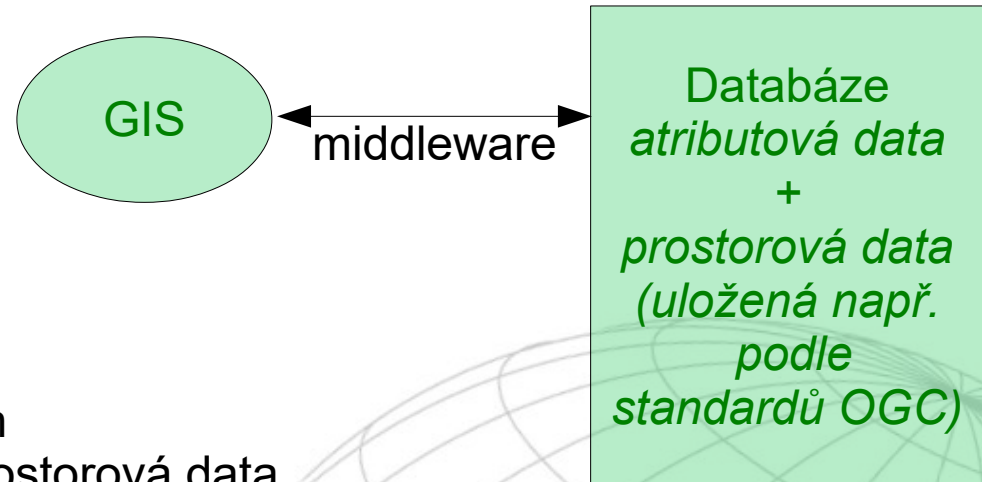


Rozdělení GIS podle generací

3. generace

• Objektově-relační model

- kombinace relačního a objektového modelu.
- Je možné vytvářet objekty, ale podporuje i klasická data.
- Využívá speciálních indexačních mechanismů, které podporují prostorová data.
- Umožňuje snadno integrovat stávající relační DB do nového prostředí.
- **O prostorová data se stará přímo DBMS** – integrita dat je bezproblémová.
- Příkladem je Oracle Spatial, ArcGIS a ArcSDE Geodatabase – za předpokladu dodržení OGC standardu pro ukládání prostorových dat.



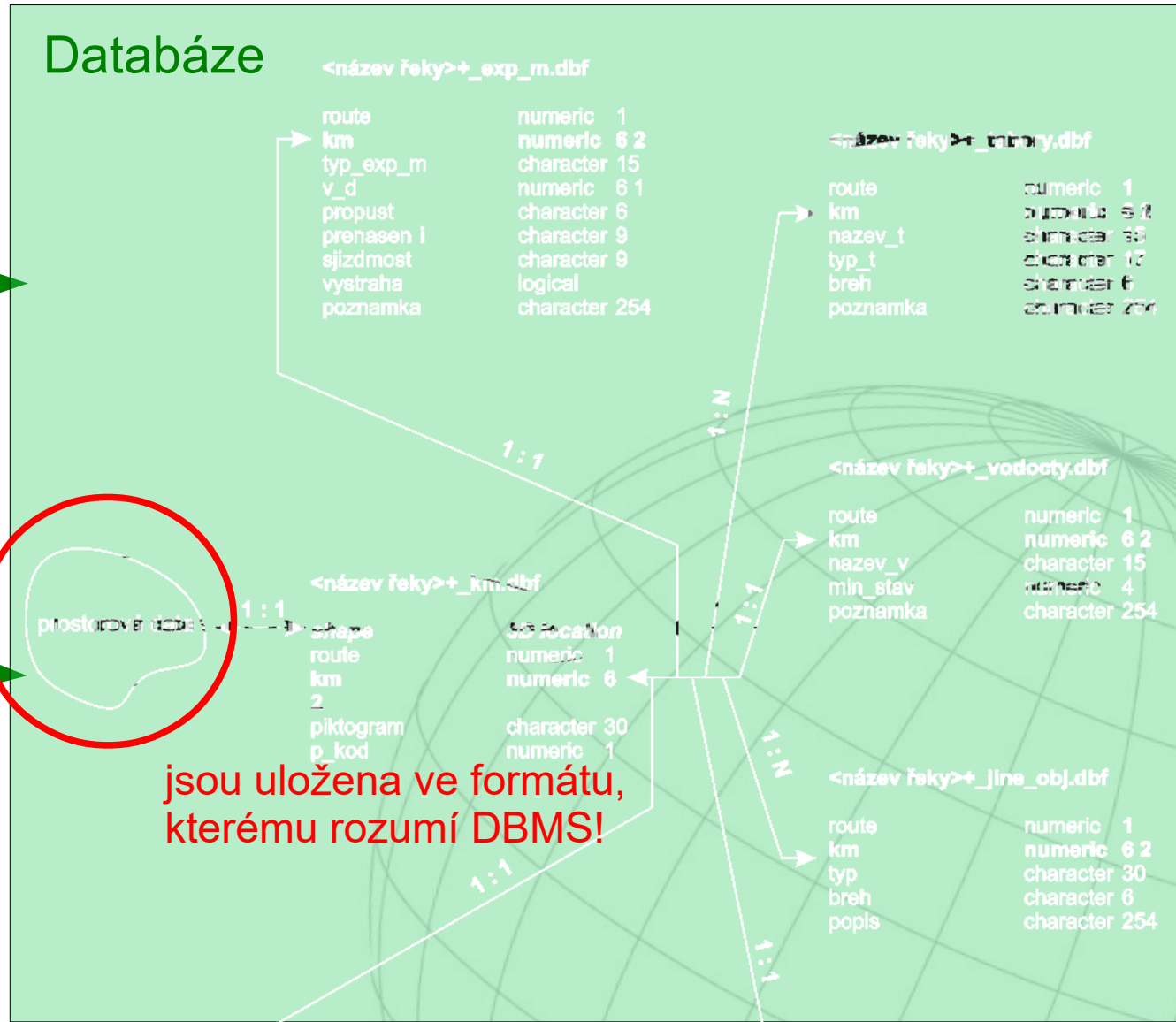
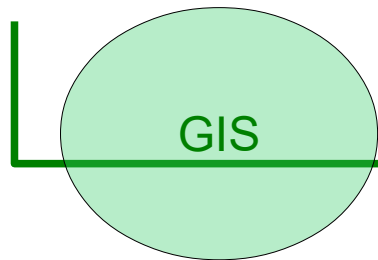
Rozdělení GIS podle generací

3. generace

Objektově-relační model

Přímý přístup k atributům

Přístup přes GIS



Rozdělení GIS podle generací

1. generace (nevelká podpora atributů)

- bez atributových souborů
- „flat“ soubory

2. generace

- duální/hybridní přístup (pro atributy jeden model, pro prostorová data jiný)
- integrovaný model (prostorová data v DB ve formě „BLOB“)

3. generace (atributová i prostorová data uložena v jedné databázi)

- objektový model
- objektově-relační přístup s podporou prostorových dat.

Používané modely

1. generace (nevelká podpora atributů)

- bez atributových souborů
- „flat“ soubory

2. generace

- **duální/hybridní model** (pro atributy jeden model, pro prostorová data jiný)
- **integrovaný model** (prostorová data v DB ve formě „BLOB“)

3. generace (atributová i prostorová data uložena v jedné databázi)

- objektový model
- **objektově-relační model** s podporou prostorových dat.