



Geoinformatika

II – GIS jako zpracování dat

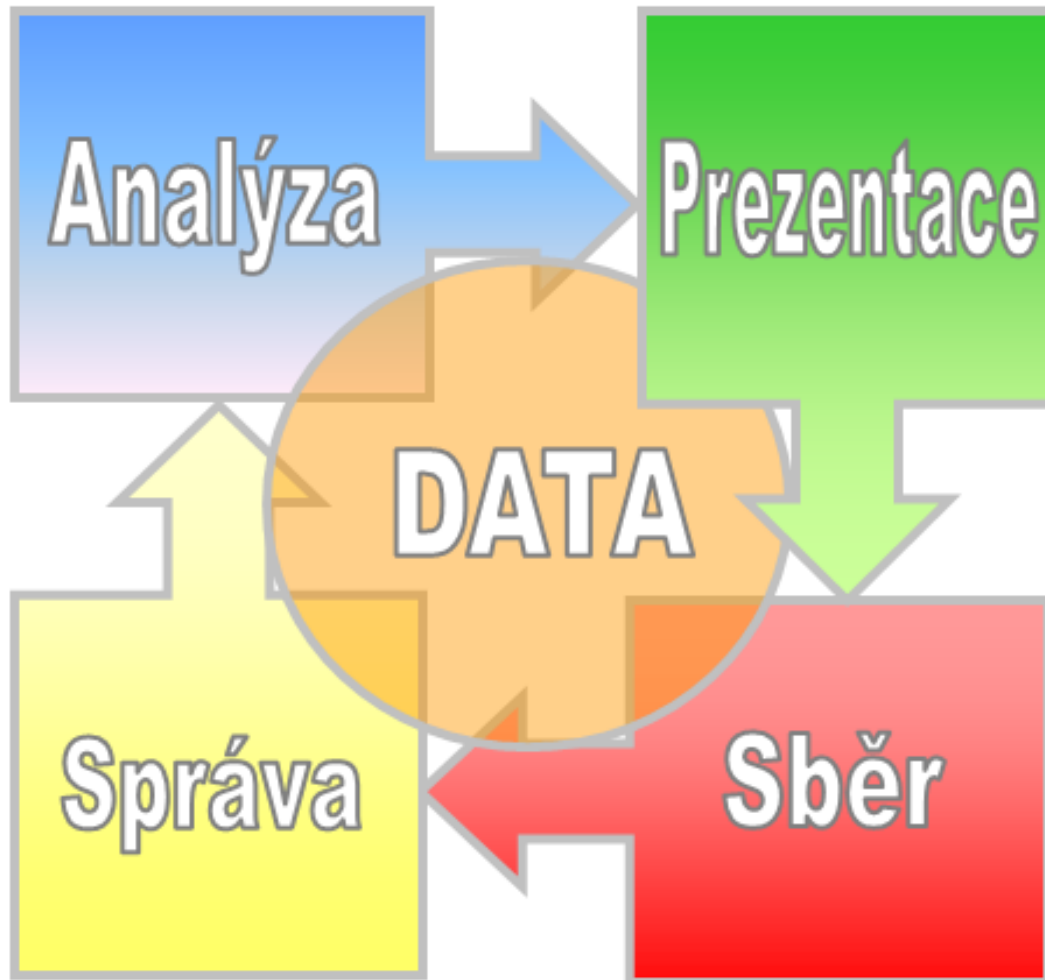
jaro 2017

Petr Kubíček

kubicek@geogr.muni.cz

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

GIS životní cyklus dat



Historie GIS

- V 50. letech 20. století začaly pokusy s automatizovaným mapováním za využití výpočetní techniky.
- V roce 1963 zavedl pojem GIS Kanadán **R. F. Tomlinson** a označil tak nové technologie pracující s daty a podávajícími informaci o terénu pomocí výpočetní techniky.

??Jak dostat mapu do počítače??

Historie GIS

- Pionýrské období (konec 60. let až 1975) - hlavně průkopnické práce, univerzity - důraz na digitální kartografii.
- 1975- začátek 80. let - ujednocení pokusů s institucemi na lokální úrovni - první **LIS**.
- 1982 - konec 80. let - **komeracionalizace problematiky** – běžně dostupné softwarové systémy pro GIS (ESRI, Intergraph, ...), první systémy založené na CAD (systémy před tím měly minimální grafické možnosti). Autocarto, **EuroCarto (1987 Brno)**.
- 1986 – P. Burrough - **učebnice GIS; Konečný a Rais – GIS v ČR**.

Historie GIS

- **NCGIA** – National Centre for Geographic Information and Analysis; GI science (USA). Změna pojetí GIS jako nástroje (research with GIS) na přijetí **GIS jako výzkumného směru** (research about GIS – spatial information theory).
- **90. léta** - počátky standardizace, uživatelské GIS, Desktop GIS, otevřené systémy (Open GIS), Internet.
- Výuka **GIS v Č(SS)R** – Brno, Ostrava, Olomouc...
- **AGILE**; Conference on Spatial Information Theory (**COSIT**).
- **Současnost** - vývoj objektově orientovaných systémů, masivní propojení s databázemi, vzdálený přístup přes Internet/Intranet, webové služby, sociální sítě, geoparticipace (VGI).
- **Mobilní GIS ...**

Komplexní GIS schéma

Transformace dat

- modelu
- polohy
- formátu

Sběr dat

- editace
- import

Uložení dat

- | | |
|-------------|------------------------------|
| Návrh | - struktura
- datové typy |
| Manipulace | - dotazování
- indexování |
| Dokumentace | - metadata |

Analýza dat

- průzkum
- modelování

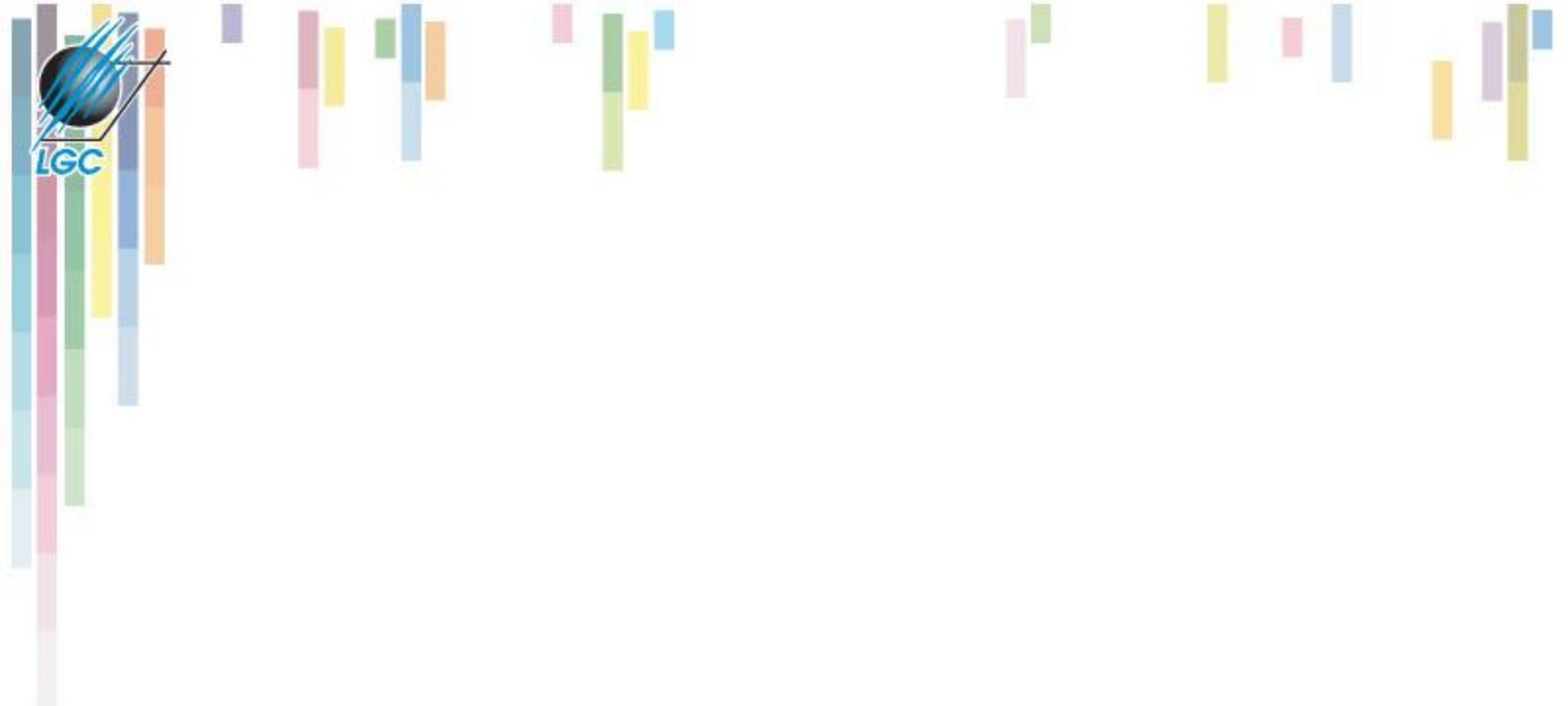
Prezentace dat

Vizuální

- kartografická
 - statické mapy
 - dynamické mapy
 - uživatelské rozhraní
- nekartografická
 - grafická
 - textová

Nevizuální

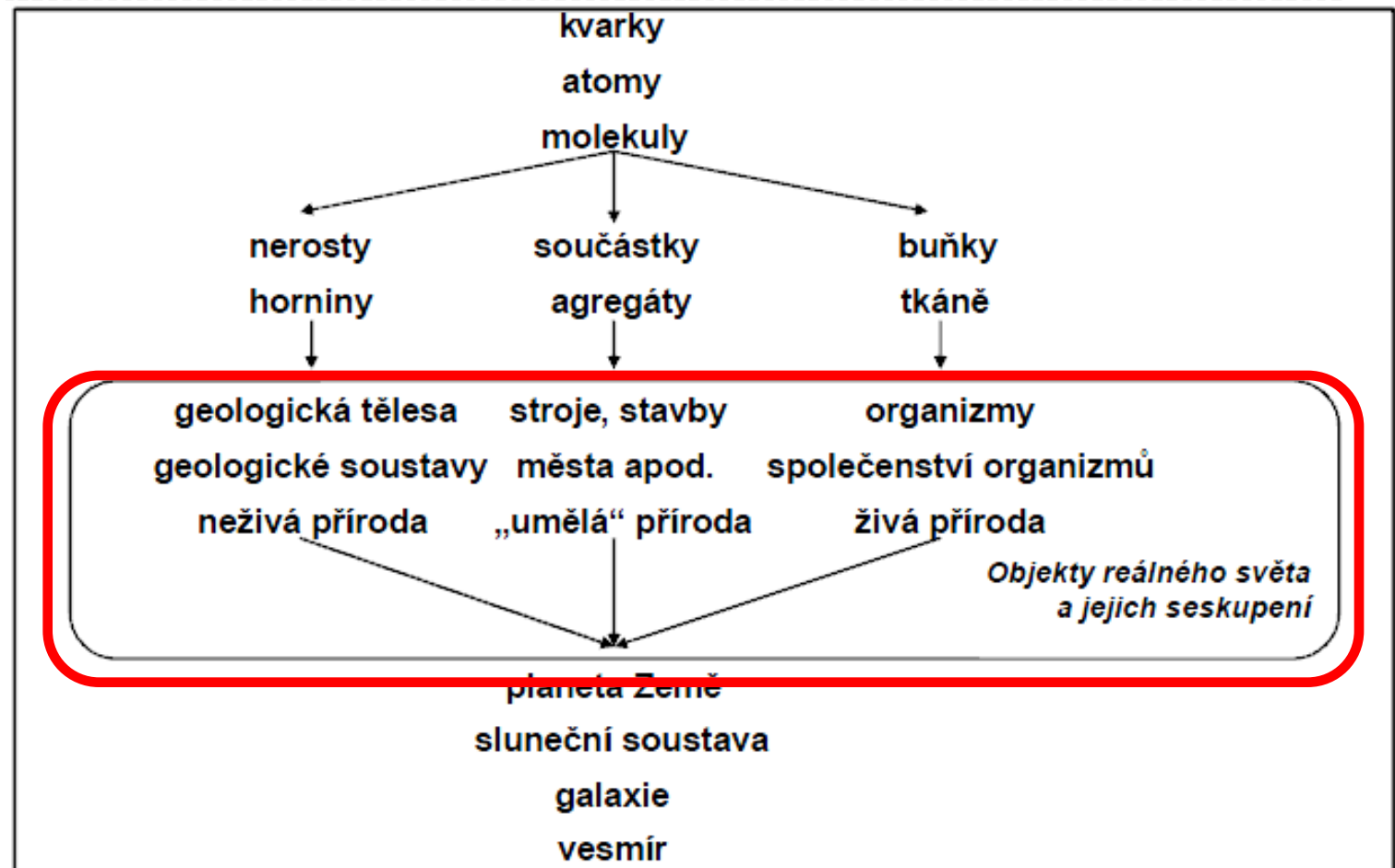
- export
- řídicí povely



TEORETICKÉ POJETÍ GIS

Pohledy na reálný svět

Tři základní pohledy - Objektový , Jevový, Procesní

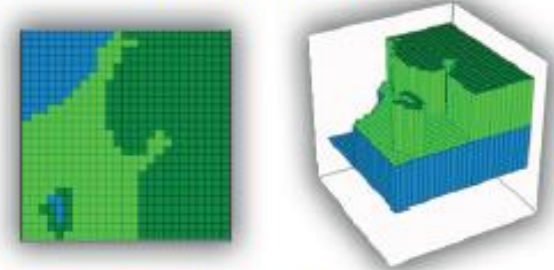




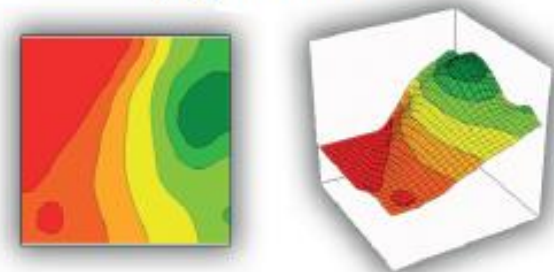
Jevový a procesní pohled

- **Dělení jevů** – kvalitativní x kvantitativní; statické x dynamické; kontinuální x diskrétní.
- **Procesní pohled** – vnáší do reálného světa **dynamiku** a ovlivňuje jak jevy, tak objekty.
- Pro modelování jevů a objektů využívá geoinformatika **odlišné datové modely**.

Diskrétní data



Spojité data



The slide features a decorative background with vertical bars in various colors (blue, green, yellow, orange, pink, purple) and a logo in the top left corner. The logo consists of a stylized globe with blue lines and the letters 'IGC' below it.

Modelování geografických objektů

- v geoinformatice se nezaobíráme reálnými objekty, ale z důvodu zjednodušené reprezentace - **modelem reality**.
- **Modelování** - abstrahování týkající se všech součástí geografické informace:
 - **geometrické,**
 - **topologické,**
 - **tematické**
 - **dynamické.**

Proces modelování

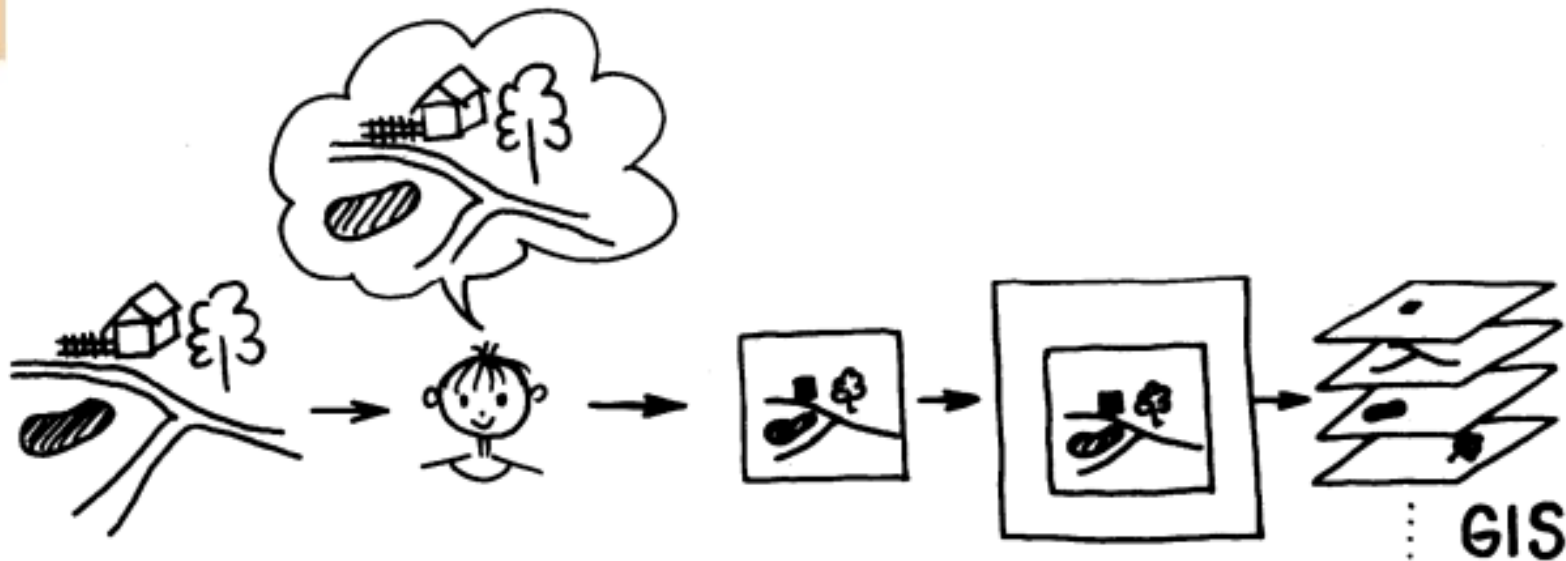
- **modelování** – je proces **abstrakce**, při kterém jsou **podstatné** elementy reálného světa **zdůrazněny** a **nepodstatné eliminovány** (s ohledem na cíl, který má toto modelování splnit):
 - **úmyslně** – zobrazují se jen ty elementy, které jsou předmětem zkoumání, ostatní se potlačují;
 - **neúmyslně** – v dané fázi poznání jsou nedostupné či nepoznatelné.
- Principem modelování je snaha o poznání vlastností studované části reality.



GIS jako obraz reálného světa (Rapant)

- **Reálný svět** je pozorován pozorovatelem. Ten vytváří na základě svých vjemů vnitřní (**mentální model**) tohoto světa.
- **Mentální model** je velmi blízký reálnému světu, ale vždy je tu jisté zjednodušení.
- Pro převedení mentálního modelu do **GISu** je potřeba ji **zdigitalizovat**. Jednotlivé prostorové prvky (geoprvky) jsou nahrazeny základními geometrickými prvky a ty jsou dále tříděny podle tématu do jednotlivých vrstev.

GIS jako obraz reálného světa



Výsledný obraz reálného světa v GISu je:

- složen z bodů, linií a polygonů (geometrie)
- roztríděných do jednotlivých vrstev (tematika)
- dvourozměrný
- statický
- zjednodušený (chybí mnoho informací)
- obsahuje mnoho chyb a nepřesností vzniklých z převodem reality do podoby GIS.



Složky geografických dat

- **Neprostorová složka (tzv. Atributy)**
 - Čísla – kvantitativní hodnoty
 - Řetězce znaků – kvalitativní hodnoty
 - Datum – časové určení
 - Komplementární atributy – linky, videa, dokumenty ...
- **Časový aspekt** odráží změnu jevu v čase (od jednoho záznamu ke druhému)
- **Prostorová složka (tzv. Geometrie)**
 - tvar
 - poloha
 - topologie

Všechny měřitelné nebo popsitelné vlastnosti reálných entit spadají do jednoho z aspektů: **prostoru, tématu** nebo **času**.



Neprostorová složka - atributy

Popisující geografické objekty a jejich vlastnosti

- **Typy atributových dat**

- *poměr* - např. procenta
- *interval* - např. celá čísla z intervalu (0,10), desetinná čísla z intervalu (0.5-14.0)
- *pořadí* (ordinální) - řadová číslovka
- *výčet* (nominální) - např. pro typ silnice to může být (dálnice, rychlostní silnice, silnice 1.třídy, silnice 2.třídy, ostatní silnice)

Příklad:

- **Objekt** = lesní porost
- **Atribut** = dřevinná skladba, průměrná výška porostu, věková struktura, apod.
- Atributy jsou **neprostorové** (nereprezentují informaci o lokalizaci či o prostorových vztazích), mají vytvořenou **vazbu** na prostorové prvky atributové hodnoty, reprezentující **kvalitu** geografického objektu, nelze vždy měřit nebo udávat v jednotném měřítku.
- Př. Borovicový porost není nikdy 100% složen pouze z borovice. Při analýzách to nevadí, ale je nutno s touto skutečností počítat (míry kvality).

Časová složka

- **Čas – dynamický popis**
 - dynamika charakterizuje časovou variabilitu geografických objektů.
 - tyto změny se mohou týkat geometrie, topologie i tematického popisu.
- **Modelování** dynamických prostorových procesů v rozměrném prostoru vyžaduje složité modely a metody.

V praxi se ale používá zjednodušení:

- 1. Analýza časové série na jednom měřícím bodě - **časová změna.**
- 2. Prostorová změna atributové hodnoty mezi dvěma body v tom samém čase - **prostorová změna.**

Prostorová složka - tvar

- **Dvě chápání prostoru – vektorový a rastrový**
 - **Jev** -> jeho vymezení -> hranice -> **objekty**
 - Kartografický model reality
 - Bod, Linie, Plocha
 - Uspořádaný soubor souřadnic - **VEKTOR**.
 - **Prostor** -> jeho rozdělení -> jaký **jev** je přítomen v dílu
 - Vzorkování
 - Různé druhy členění podle charakteru buněk
 - Pravidelnost, komplexita
 - Matice hodnot – u pravidelného a jednoduchého členění - **RASTR**.

Datové modelování – vektorová reprezentace

Bod

- nemá délku, hloubku ani šířku - bezrozměrný (0D) prvek
- je jednotlivý pár souřadnic X, Y , reprezentující geografický prvek
- je příliš malý na to, aby byl zobrazen jako linie či plocha.

Základní geometrické objekty

Bod:

x
[x,y]

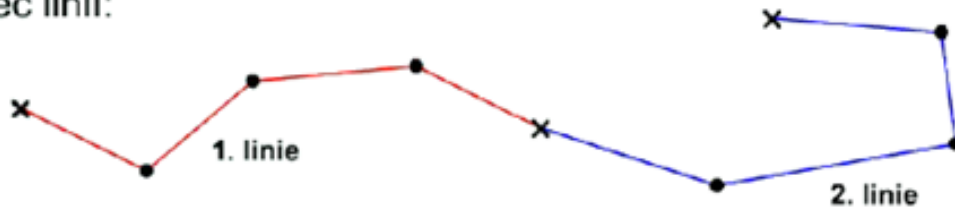
Typy geometrických objektů - linie



Linie:



Řetězec linií:

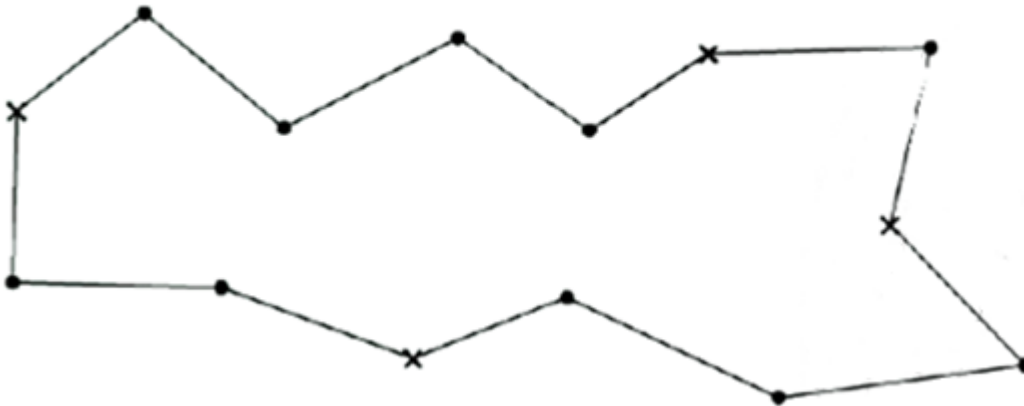


Linie

- má délku, ale nemá šířku ani hloubku - jednorozměrný (**1D**) geografický prvek;
- je sled orientovaných **úseček (hran)** definovaných souřadnicemi **vrcholů (vertex)** mezi dvěma **uzly (nodes)**;
- tvar reprezentovaného geografického prvku je příliš úzký na to, aby mohl být zobrazen jako plocha.

Typy geometrických objektů - plocha

Plocha - uzavřený řetězec linií:

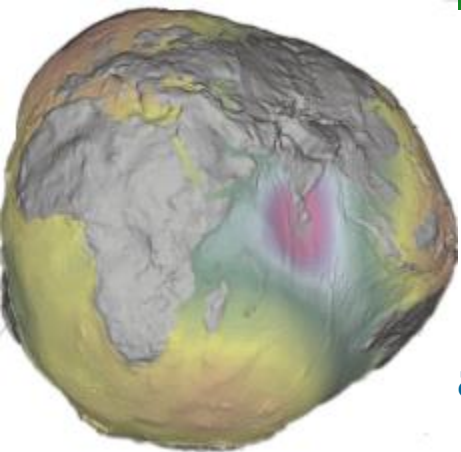
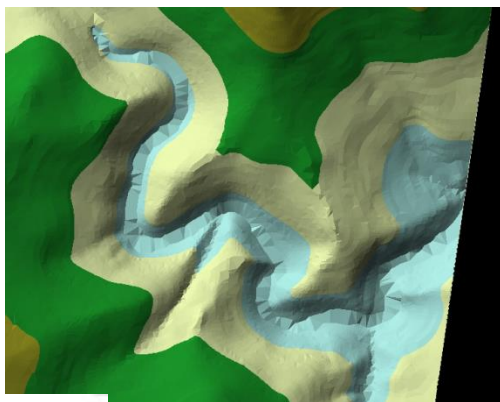


- **Polygon (plocha)**
 - mají délku a šířku, ale nemají hloubku - dvojrozměrný (**2D**) geografický prvek;
 - je uzavřený obrazec, jehož hranicí je uzavřená linie.
 - První a poslední vrchol jsou stejné.

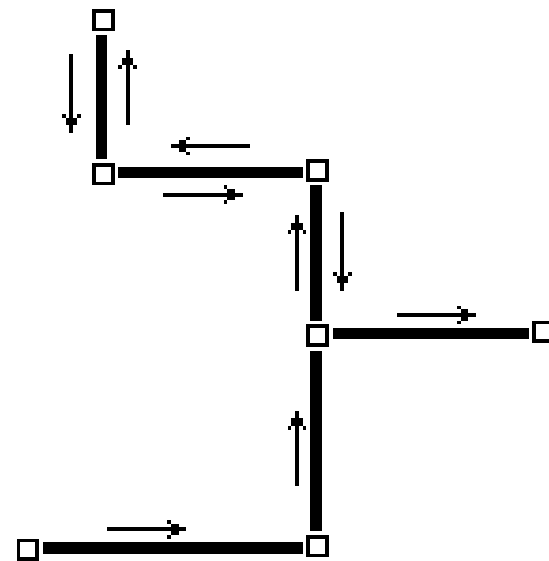


Sítě

- systém linií s topologickou strukturou;
- je řada vzájemně propojených linií, podél níž probíhá tok informací.



a



Povrchy

- je to souvislá entita, pro kterou v každém bodě existuje nějaká hodnota (2,5 D) x, y, z .

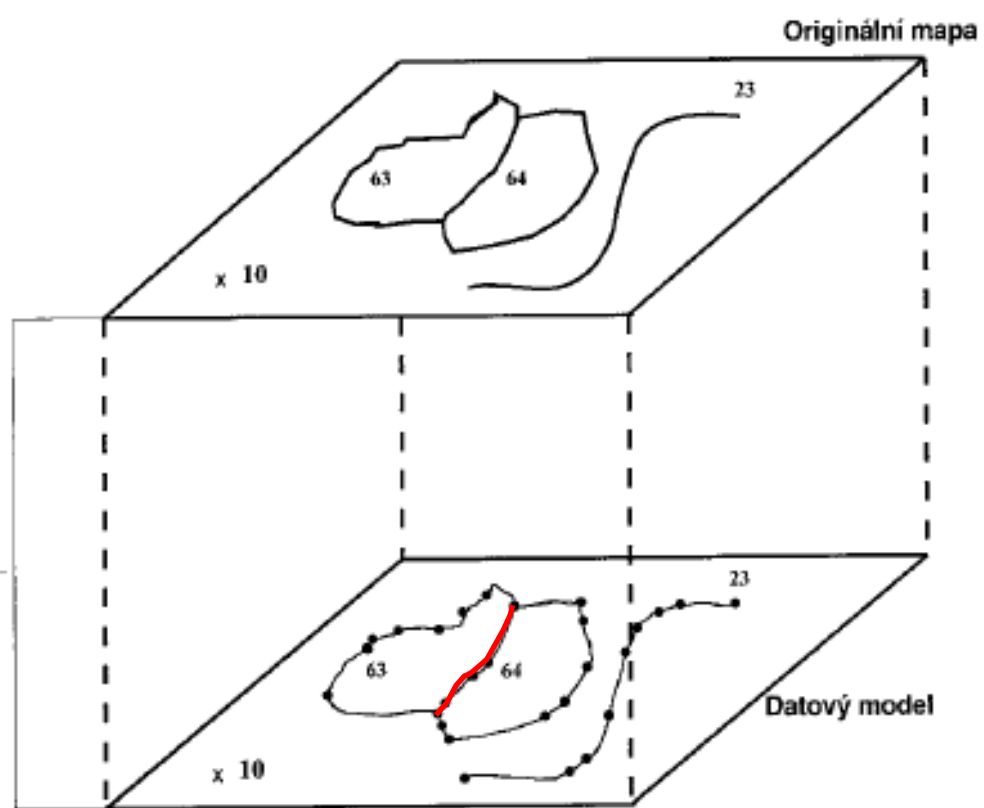
Objemy

- mají všechny rozměry (délku, šířku, hloubku) - trojrozměrné (3D) geografické prvky.



Špagetový datový model

- Nejjednodušší – CAD, CAM.
- Objekt na mapě se reprezentuje **jedním logickým záznamem** v souboru a je definovaný jako **řetězec x,y souřadnic**.
- Nevýhody - ačkoli jsou všechny objekty v prostoru definovány, struktura **neposkytuje informace o vztazích mezi objekty**.
- **Společná linie je pro každý polygon ukládána dvakrát**.
- Pro většinu **prostorových analýz je tento model nevhodný**, protože veškeré potřebné prostorové vztahy musí být spočítány před každou analýzou.



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	$X_1 Y_1, X_2 Y_2, \dots, X_1 Y_1$ Uzavřená smyčka
	64	$X_1 Y_1, X_2 Y_2, \dots, X_1 Y_1$



Topologický datový 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 (bez opakovaného zápisu).

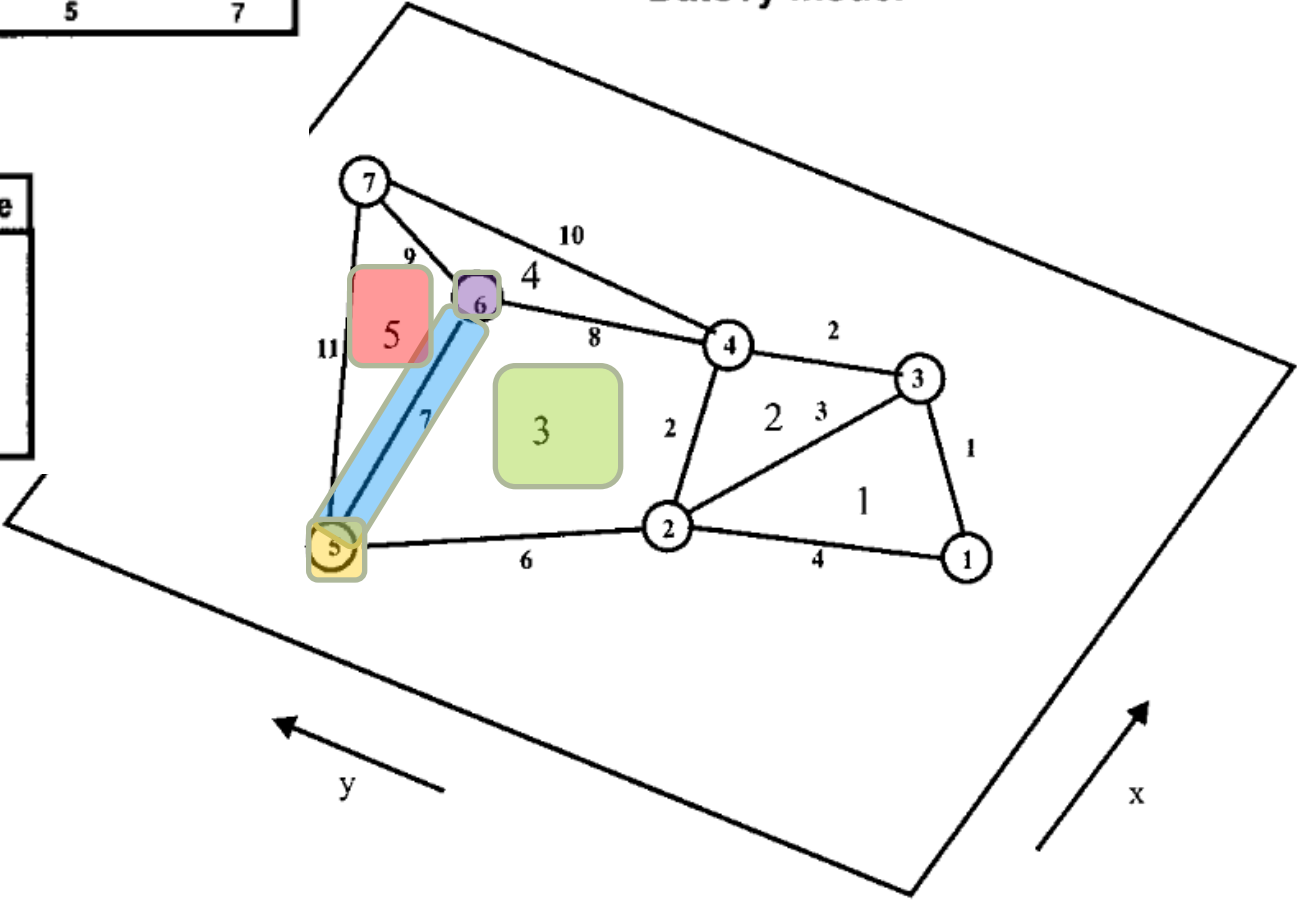
Soubor topologických vztahů

Hrana	Pravý Polygon	Levý Polygon	Počátek v bodě	Konec v bodě
1	1	0	3	1
2	2	0	4	3
3	2	1	3	2
4	1	0	1	2
5	3	2	4	2
7	3	5	5	6
9	4	5	7	6
10	0	4	7	4
11	5	0	5	7

Soubor souřadnic bodů

Uzel	X souřadnice	Y souřadnice
1	23	8
2	17	17
3	29	15
4	26	21
5	8	26
6	22	30
7	24	36

Datový model





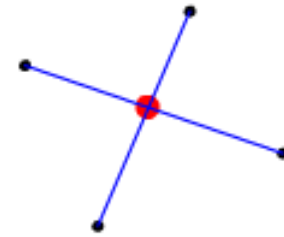
Vektorová reprezentace - topologie

- **Topologie je matematický způsob, jak explicitně vyjádřit prostorové vztahy mezi jednotlivými geometrickými objekty.**
- **Proč vůbec topologie? Má jisté výhody, například:**
 - Umožní ukládat data efektivněji.
 - Mnoho analýz v GIS využívá pouze topologické a nikoli geometrické vztahy.
- **Důvod pro využívání topologie (ESRI 1995):**
- "Topology is useful in GIS because many **spatial modeling operations don't require coordinates, only topological information**. For example, to find an **optimal path between two points** requires a **list of the arcs** that connect to each other and the **cost to traverse each arc in each direction**. Coordinates are only needed for drawing the path after it is calculated."

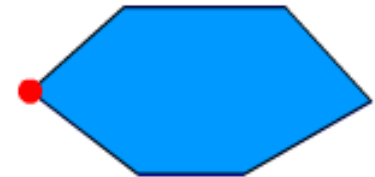
Tři základní topologické koncepty

- **Konektivita** – dvě linie se na sebe napojují v uzlech.
- **Definice plochy** – linie, které uzavírají nějakou plochu, definují polygon.
- **Sousednost** - linie mají směr a nesou informaci o objektech nalevo a napravo od nich.

Konektivita



Definice plochy

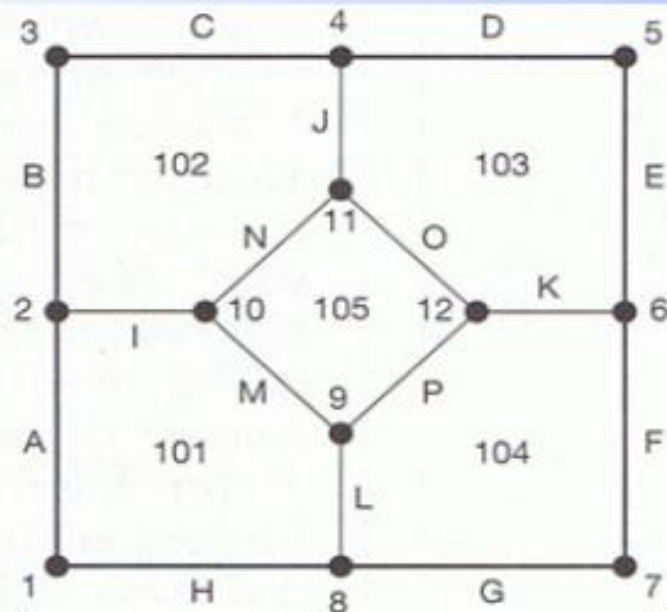


Princip okřídlené hrany:





Kahoot 😊



Node file		
ID	X	Y
1	0	0
2	0	10
3	0	20
⋮	⋮	⋮
12	14	10

Polygon structure file	
ID	Chain/Segment list
101	A, I, M, L, H
102	B, C, J, N, I
103	D, E, K, O, J
104	F, G, L, P, K
105	M, N, O, P

Chain/Segment file					
ID	Start-node	End-node	Left-poly	Right-poly	Length
A	1	2	Outside	101	10
B	2	3	Outside	102	10
C	3	4	Outside	102	10
⋮	⋮	⋮	⋮	⋮	⋮
P	12	9	104	105	4

Polygon attribute file		
ID	VAR 1 (Name)	VAR 2 (Area)
101	Cars	96
102	Cars	96
103	Staff	96
104	Buses	96
105	Info kiosk	16

ArcGIS Shapefile

- Jeden soubor obvykle reprezentuje jeden typ mapového prvku, např. silnice, jezera, obce
- Shapefile specifikuje i další pomocné soubory.
- „**Jméno.přípona**“ prefix zůstává stejný, přípona se mění:
- **Povinné**
 - **.shp** – samotný hlavní soubor s geodaty (geometrie).
 - **.shx** – indexový soubor (posun vůči počátku souboru, délka záznamu).
 - **.dbf** – soubor s atributy resp. popisné data.
- **Nepovinné**
 - **.prj** – zdrojový souřadnicový systém.
 - WKT-string
GEOGCS["GCS_WGS_1984",DATUM["D_WGS_1984",SPHEROID["WGS_1984",6378137,298.257223563]],PRIMEM["Greenwich",0],UNIT ["Degree",0.017453292519943295]]
 - **.cpg** – specifikuje kódování v dbf souboru.
 - UTF-8.

Struktura *.SHP

- Geometrický prvek v záznamu - **shape**
- Samotnou geometrii shape ukládá jako **sekvenci bodů** (např. GPS souřadnic).
- Nedefinuje topologickou strukturu.
- Jeden záznam shape – jeden řádek v attributech.

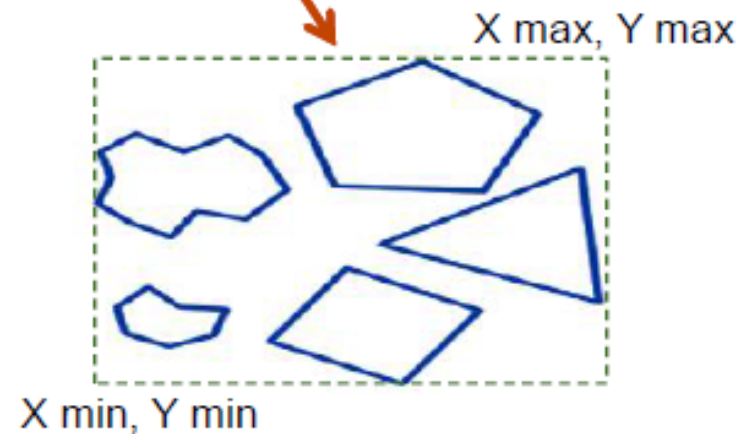
Point

File header [100b]	
Record header [8b]	Record content
Record header [8b]	Record content
Record header [8b]	Record content
Record header [8b]	Record content
...	...
...	...
Record header [8b]	Record content

Hlavička souboru *.SHP

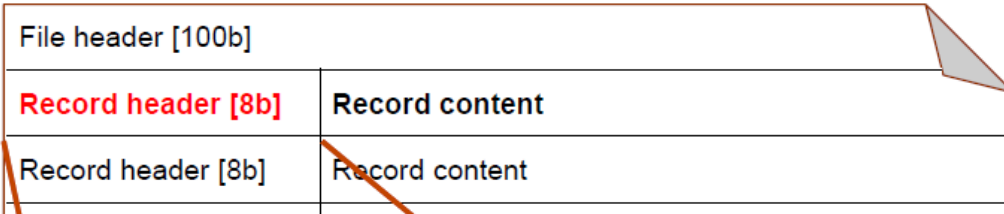
0	4	24	28	32	36	68	84
int	int	int	int	int	4 x double	2x double	2 x double
File code vždy 94440	nepouž.	délka souboru	verze	Typ shape	MBR X min Y min X max Y max	Z min Z max	M min M max

0	NullShape	15	PolygonZ
1	Point	18	MultiPointZ
3	PolyLine	21	PointM
5	Polygon	23	PolyLineM
8	MultiPoint	25	PolygonM
11	PointZ	28	MultiPointM
13	PolyLineZ	31	MultiPatch





Hlavička/obsah záznamu

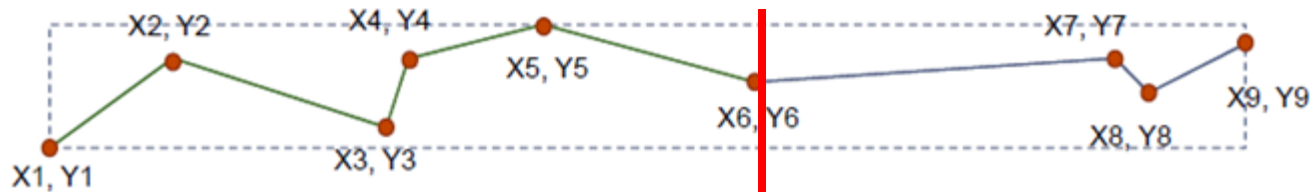


0	4
int	int
číslo záznamu	délka záznamu



0	4	36	40	44	X	EOF
int	4 x double	int	int	Int[]	Point[] (2 x double na 1 bod)	
Shape type	MBR	Počet částí	Počet bodů	Indexy na části	Body	$X = 44 + 4 * \text{počet částí}$

3	Xmin, Ymin, Xmax, Ymax	2	9	[0,5]	[X1,Y1], [X2,Y2], ..., [X6,Y6],..., [X9,Y9]
---	------------------------	---	---	-------	---



Atributová data *.dbf

- Standartní DBF soubor (tabulka)
- Ke každému záznamu existuje právě jeden řádek v tabulce ve stejném pořadí jako ve zdrojovém shapefile.
- Stejný prefix jako zdrojový shapefile.
- Kódování uloženo v **.cfg** souboru.
- Velké množství dat, redundance.

	SHAPE_ID	LINK_ID	ST_NAME	FEAT_ID	ST_LANGCD	NUM_STNM	ST_NM_F
▶	0	565809744	CINGROVA	1248477527	CZE	2	
	1	565809752		0		0	
	2	565809753	CINGROVA	1248477527	CZE	2	
	3	565809754		0		0	
	4	565809755	CINGROVA	1248477527	CZE	2	
	5	565809756	CINGROVA	1248477527	CZE	2	
	6	565809757	56	1410701498	CZE	2	
	7	565809758	56	1410701498	CZE	2	

Shapefile - shrnutí

+ **výhody**

- Neukládá topologii dat
- Snadná editace bodů
- Rychlá vizualizace geodat
- Jednoduše pochopitelná struktura
- Podpora v GIS softwarech
- Snadná projekce do jiných souřadnicových systémů

- **výhody**

- Neukládá topologii dat
- Redundance dat (např. body sousedících polygonů)
- Manipulace s detailní shapefile (až 100MB soubor) je pomalá.
- Špatná podpora Unicode.

The slide features a decorative header with a grid of vertical bars in various colors (blue, green, yellow, orange, pink, purple) and a logo in the top-left corner consisting of a blue sphere with white lines and the letters 'IGC' below it.

Výhody

- lze pracovat s jednotlivými objekty jako se samostatnými celky;
- menší náročnost na paměť;
- dobrá reprezentace jevové struktury dat;
- vysoká geometrická přesnost
- kvalitní grafika, přesné kreslení, znázornění blízké mapám;
- jednoduché vyhledávání, úpravy a generalizace objektů a jejich atributů.

Vektorová data

Nevýhody

- výpočtová náročnost (problémy při náročných analytických operacích);
- komplikovanost datové struktury;
- složitější odpovědi na polohové dotazy;
- obtížná tvorba překryvů vektorových vrstev (overlay) ;
- problémy při modelování a simulaci jevů.