



Geoinformatika

II – GIS jako zpracování dat

jaro 2024

Petr Kubíček

kubicek@geogr.muni.cz

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

Co je geoinformatika?

- **Věda o zpracování geografické informace**
 - **Geografická informace** je soubor poznatků o nějakém jevu (události, předmětu, procesu...) jehož součástí je **vymezení** tohoto jevu **vůči** (zemskému) **povrchu**.
- **Technologický základ geoinformačních věd**
 - Kartografie, dálkový průzkum Země (DPZ, RS), geodézie (sruveying).

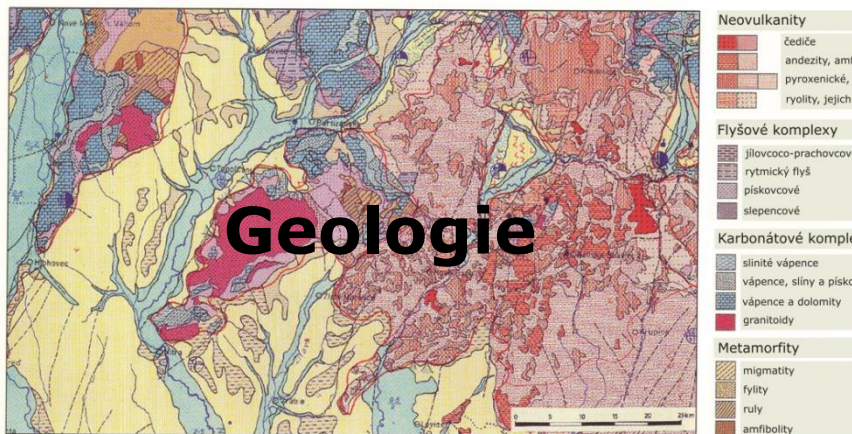
Proč GIS?



Rozdílné:

- Zdroje
- Umístění
- Měřítko
- Zobrazení
- Legenda

Obr. 4.2 Příklad přehledné inženýrskogeologické mapy Slovenska (Matula, 1969)



Území znehodnocené geodynamickými jevy

Hydrogeologické poměry

Co je to GIS?

- Co je to **informační systém**?
- Informační systém je soubor hardware a software na získávání, uchovávání, spojování a vyhodnocování informací.
- Informační systém se skládá ze zařízení na **zpracování dat**, systému **ukládání dat** (báze) a **vyhodnocovacích** programů.
- Co je **geografický**?? 😊

Co je to GIS?

- **Geografický** informační systém
- Je informační systém pracující oproti klasickým informačním systémům navíc i s **prostorovou složkou dat.**
- Také lze říci, že je výkonným nástrojem geověd, tedy že metody těchto věd umožňuje efektivně implementovat v počítačovém prostředí.
- **Předmět** výzkumu geoinformatiky.

- Varování před aktuální vlnou podvodných e-mailů
- Celodenní děkanské volno pro studenty PFF 14. května 2020

MOJE APLIKACE

- Pošta 0365
- Kalendář
- Garant
- Učitel
- Školitel

vyhledat v ISU

Život na MU

Výzkum: Děti v Česku i zahrani...

Psychologové z Masarykovy univerzity upozorňují, že jde o běžnou součást dospívání.

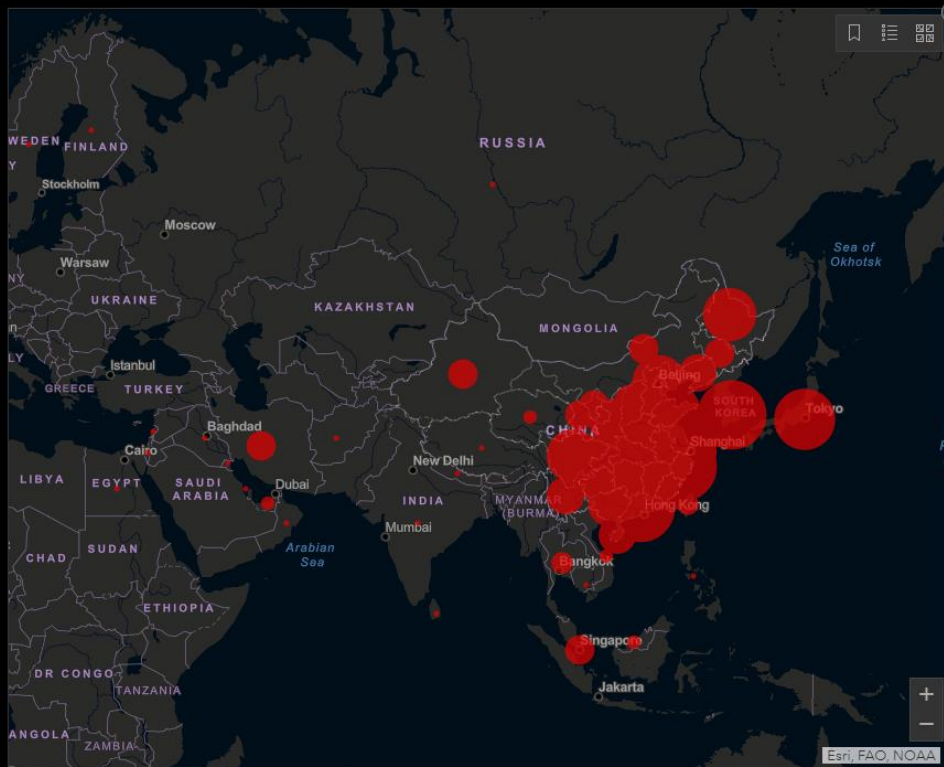


Coronavirus COVID-19 Global Cases by Johns Hopkins CSSE

Total Confirmed
80 150

Confirmed Cases by Country/Region

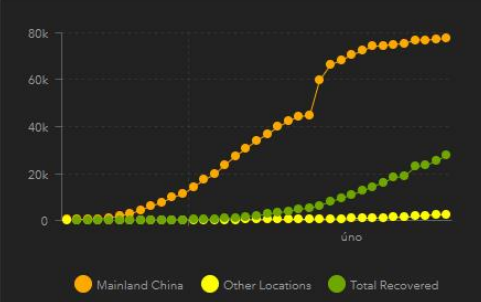
- 77 660 Mainland China
- 893 South Korea
- 691 Others
- 229 Italy
- 160 Japan
- 90 Singapore
- 81 Hong Kong
- 61 Iran
- 53 US
- 35 Thailand
- 31 Taiwan
- 22 Australia
- 22 Malaysia
- 16 Germany
- 16 Vietnam
- 13 United Arab Emirates
- 13 UK
- 12 France



Cumulative Confirmed Cases Existing Cases

Lancet Article: [Here](#), Mobile Version: [Here](#), Visualization: [JHU CSSE](#), Automation Support: [Esri Living Atlas team](#), Data sources: [WHO](#), [CDC](#), [ECDC](#), [NHC](#) and [DXY](#). Read more in this [blog](#). [Contact US](#). Downloadable database: GitHub: [Here](#). Feature layer: [Here](#). Point level: City level - US, Canada and Australia; Province level - China; Country level - other countries. Existing cases = total confirmed - total recovered - total deaths.

Total Deaths	Total Recovered
2 699	27 695
2 563 deaths Hubei Mainland China	18 889 recovered Hubei Mainland China
19 deaths Henan Mainland China	989 recovered Henan Mainland China
12 deaths Iran	807 recovered Guangdong Mainland China
12 deaths Heilongjiang Mainland China	794 recovered Zhejiang Mainland China
8 deaths South Korea	757 recovered Hunan Mainland China
7 deaths Italy	712 recovered Anhui Mainland China
7 deaths Guangdong Mainland China	682 recovered Jiangxi Mainland China
6 deaths Anhui Mainland China	458 recovered Jiangsu Mainland China
6 deaths Chongqing Mainland China	349 recovered Chongqing Mainland China



Actual Logarithmic Daily Increase

Definice GIS(?)

- **Pro GIS neexistuje jednotná definice, proto si jich uvedeme několik (mezinárodní):**
- *GIS je soubor prostředků pro sběr, ukládání, vyhledávání, transformaci, analyzování a zobrazování prostorových údajů z reálného světa z hlediska: 1. jejich polohy vzhledem k souřadnicovému systému; 2. jejich popisných – atributových vlastností; 3. jejich topologie. (**Burrough**, 1986)*
- *GIS je organizovaný soubor počítačového hardwaru, softwaru a geografických údajů navržený na efektivní získávání, ukládání, upravování, správu, analyzování a zobrazování všech forem geografických informací. (**ESRI**)*

Definice GIS

Národní:

- **GIS** je kolekce počítačového technického vybavení, programového vybavení geografických údajů a personálu, určená k účinnému sběru, ukládání, údržbě, manipulaci, analýze a zobrazování všech forem geograficky vztažené informace (Neumann, 1996).
- **Geoinformatika** je vědecká a technická disciplína, jejímž předmětem jsou prostorové údaje a která se soustřeďuje na jejich sběr, ukládání, manipulaci s nimi a jejich zprostředkování (Neumann, 1996).
- Další Rapant (2006)...

Podle různých kritérií:

- **Strukturální komponenty GIS**
- **Funkční komponenty GIS**
- **Koncepční přístupy k GIS**

Členění GIS - struktura

Strukturální komponenty GIS

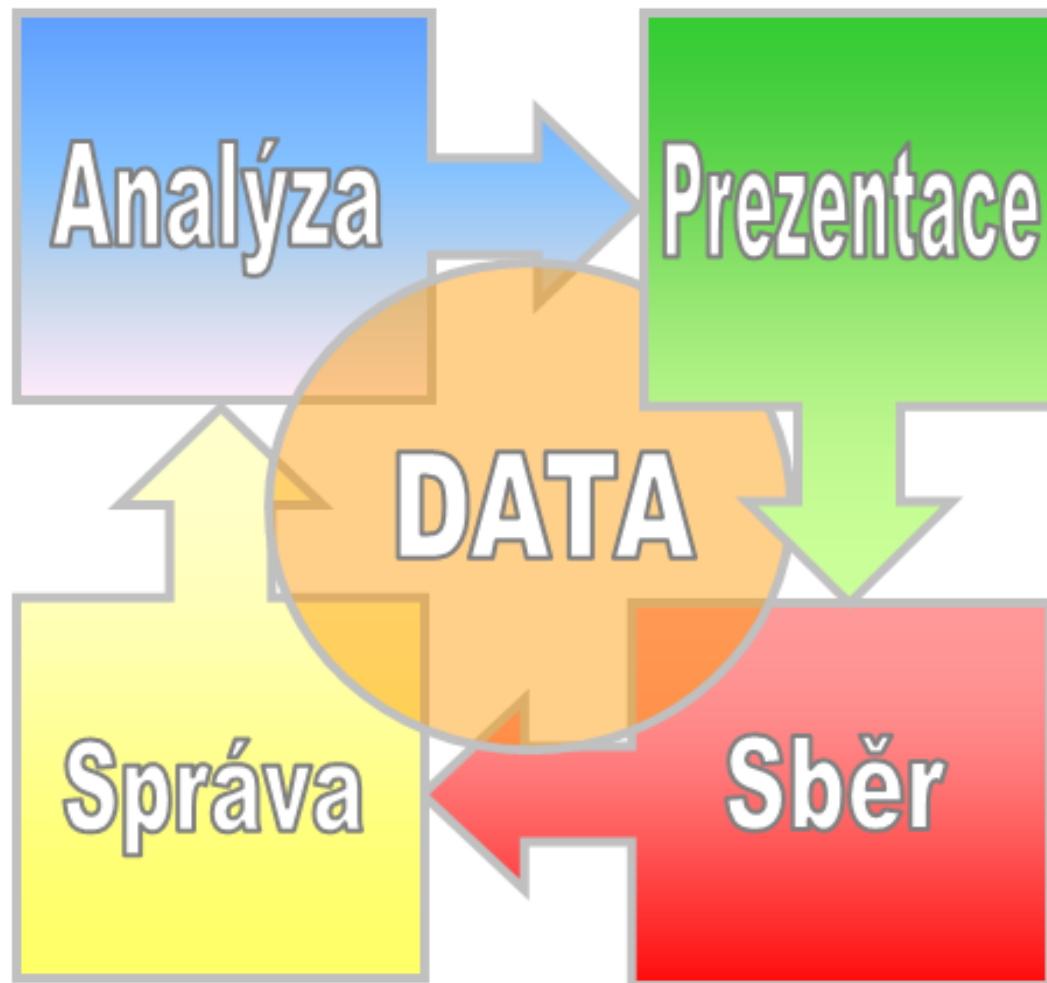
- Hardware
- Software
- Data
- **Lidé**
- Metody

Členění GIS – funkční

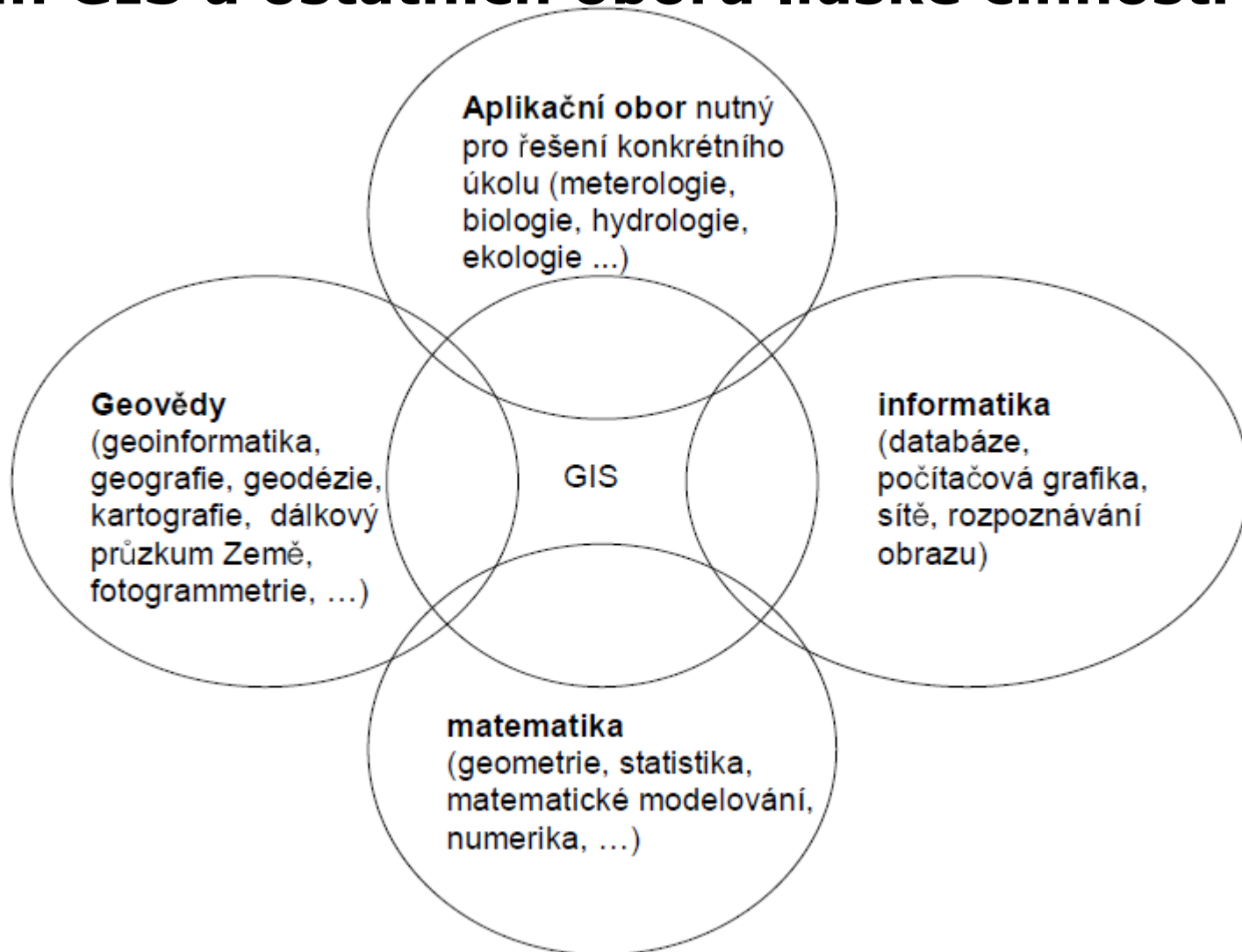
Funkční komponenty GIS

- **Vstup** dat.
- **Zpracování** a uchování dat.
- **Vykonávání analýz a syntéz** z využitím prostorových vztahů - jádro GIS, tedy to co nejvíce odlišuje GIS a jiné IS.
- **Prezentace** výsledků (výstupy grafické - mapy, negrafické - zprávy, souhrnné tabulky, statistická vyhodnocení, ...).
- **Interakce** s uživatelem (desktop GIS, Web GIS).

GIS životní cyklus dat

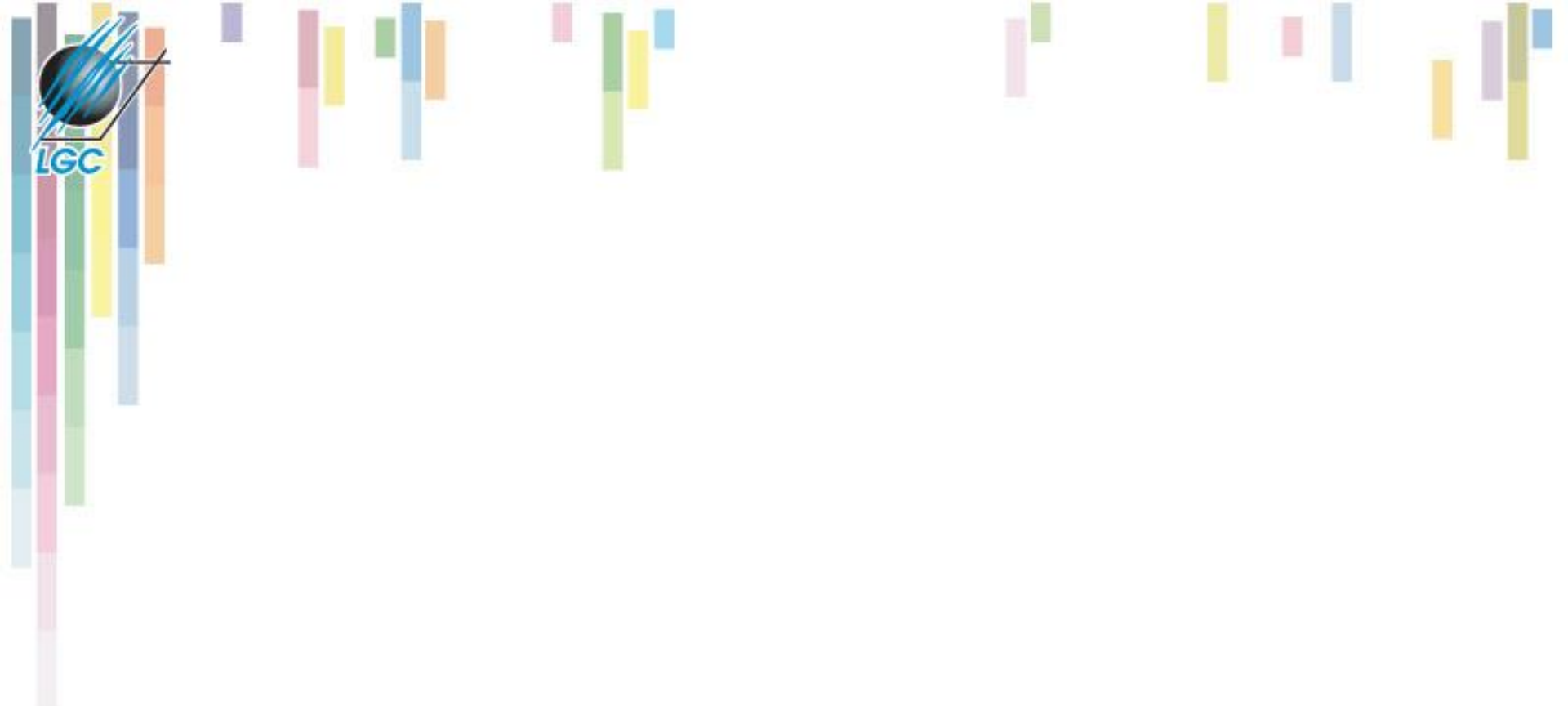


Vztah GIS a ostatních oborů lidské činnosti



Je GIS přínosný?

- **Proč vůbec používat GIS?**
 - 80 % dat lze prostorově lokalizovat.
- **Příklady využití GIS:**
 - mapové portály, služby,
 - obchod,
 - ochrana proti pohromám – krizové řízení,
 - správa inženýrských sítí (distribuční společnosti),
 - životní prostředí,
 - veřejná správa (ministerstva, kraje, města),
 - školství.



HISTORIE GIS – JAK TO VŠECHNO ZAČALO?

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 - komercializace 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 (1985).**

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** - vzdálený přístup přes Internet/Intranet/Cloud, webové služby, sociální sítě, geoparticipace (VGI), 3D GIS, strojové učení.
- **Mobilní GIS**, Imerzivní GIS, **GeoAI**...

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

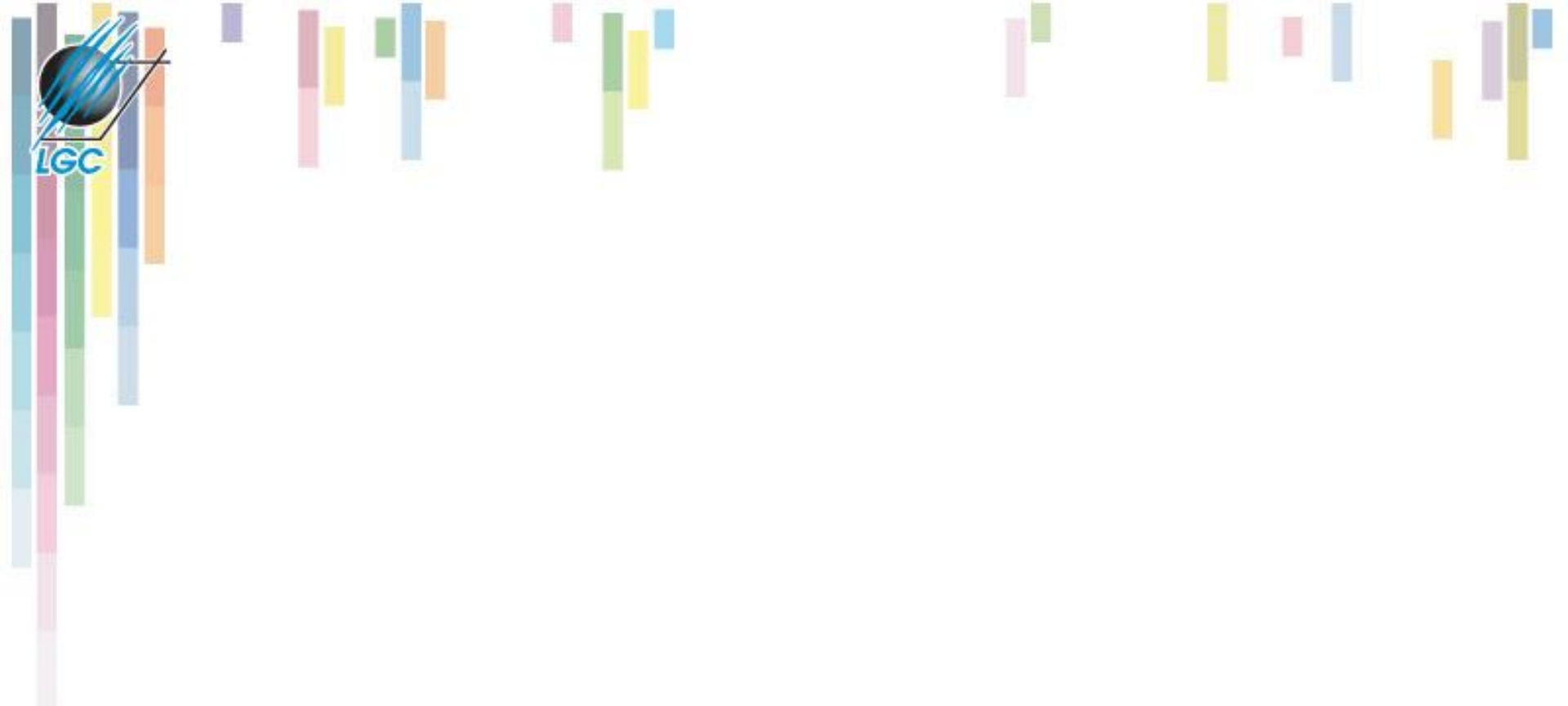
Prezentace dat

Vizuální

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

Nevizuální

- export
- řídicí povely



TEORETICKÉ POJETÍ GIS

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.



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.**



Složky geografických dat

- **Neprostorová složka (tzv. atributy) - tematická**
 - 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 (KDE)**, **tématu (CO)** nebo **času (KDY)**.



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.
- **Vyjádření kvality(nejistoty) atributu** - 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.
- Příklad: 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 měřicím bodě - **časová změna.**
 2. Prostorová změna atributové hodnoty mezi dvěma body v tom samém čase - **prostorová změna.**

Geoinformatika





Prostorová složka - tvar

- **Dvě chápání prostoru a jejich vyjádření v GIS**
 - **Prvek** -> 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]

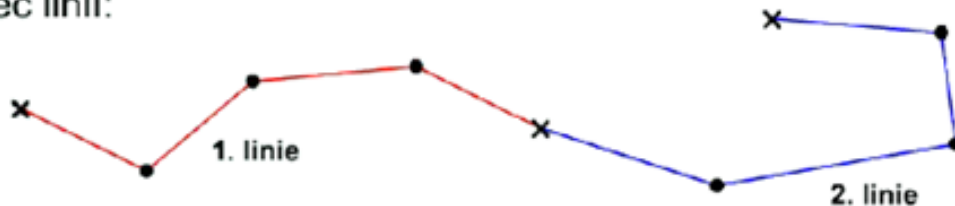
Typy geometrických objektů - linie



Linie:



Řetazec linií:

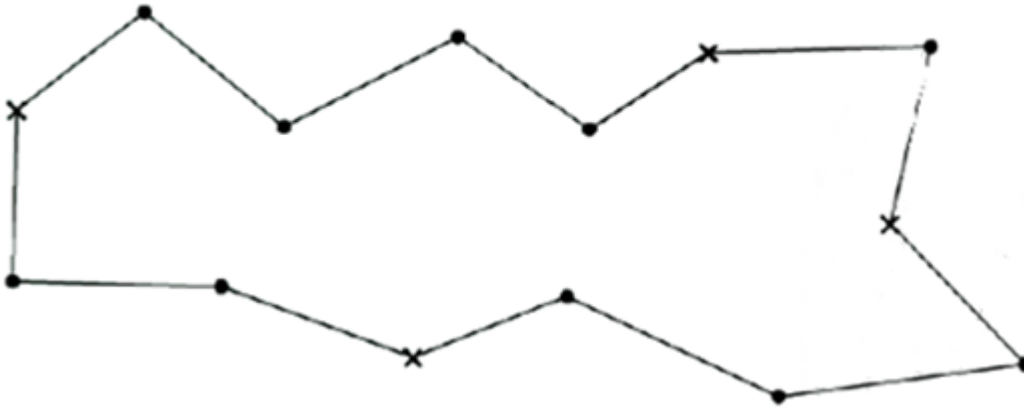


Linie

- 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:

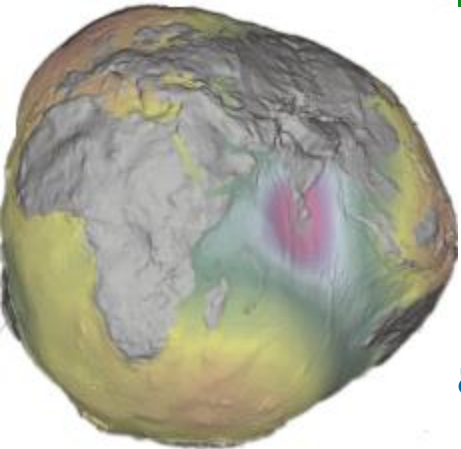
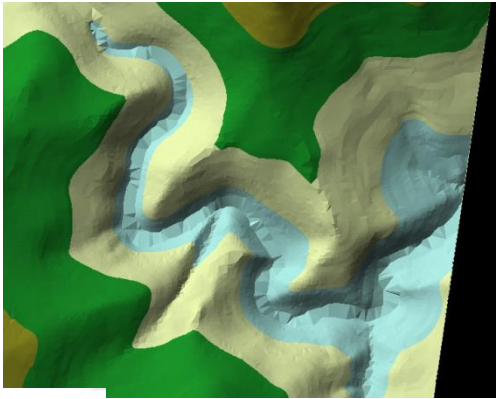


- **Polygon (plocha)**
 - mají délku a šířku, ale nemají hloubku - dvojrozměrný (**2D**) geografický prvek;
 - je uzavřený obrazec, jehož hranicí je uzavřená linie.
 - **První a poslední vrchol jsou stejné.**

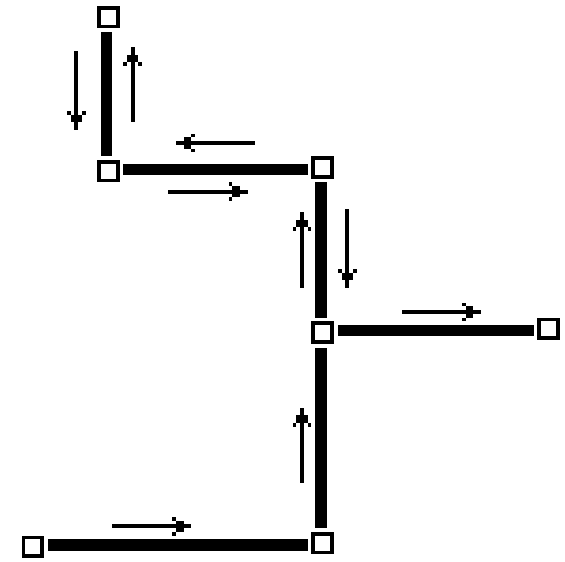


Sítě

- 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 .

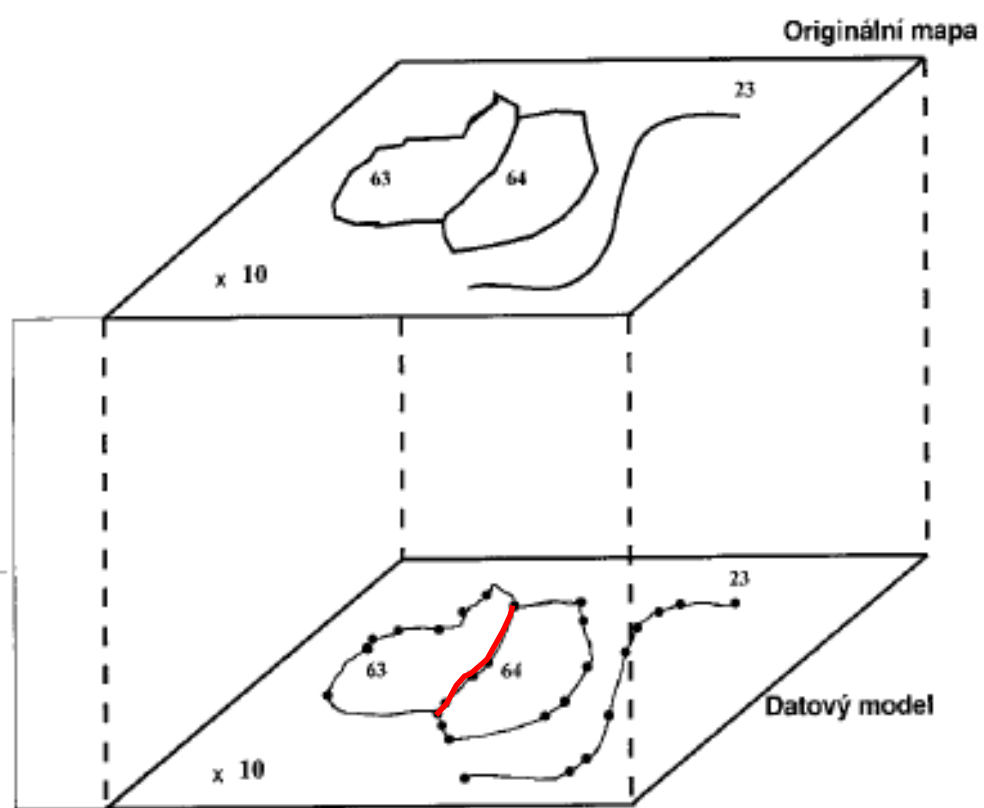
Objemy

- 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).

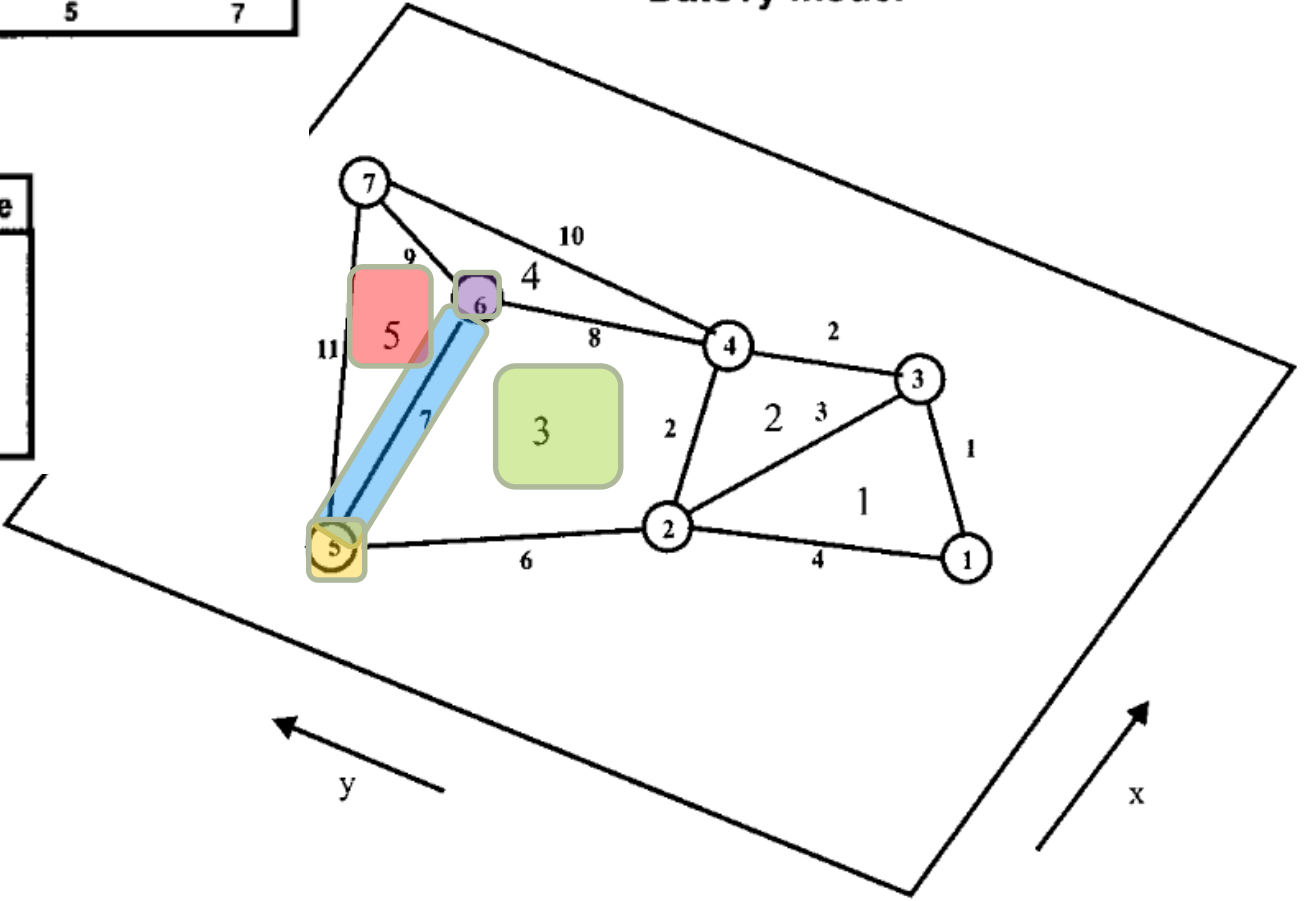
Soubor topologických vztahů

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

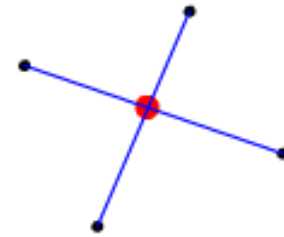
- **Topologie je matematický způsob, jak explicitně vyjádřit prostorové vztahy mezi jednotlivými geometrickými objekty.**
- **Proč vůbec topologie? Má jisté výhody, například:**
 - Umožní ukládat data efektivněji.
 - Mnoho analýz v GIS využívá pouze topologické a nikoli geometrické vztahy.
- **Důvod 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."



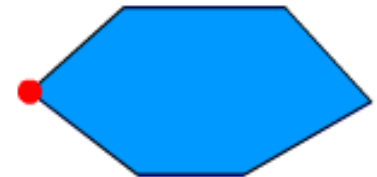
Tři základní topologické koncepty

- **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

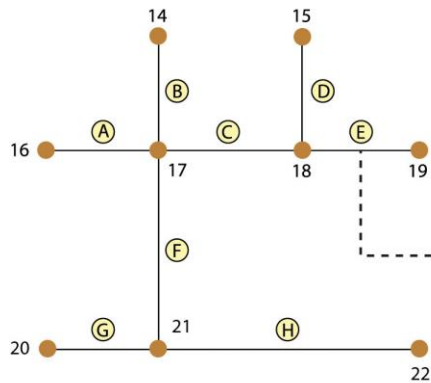


Princip okřídlené hrany:



Konektivita

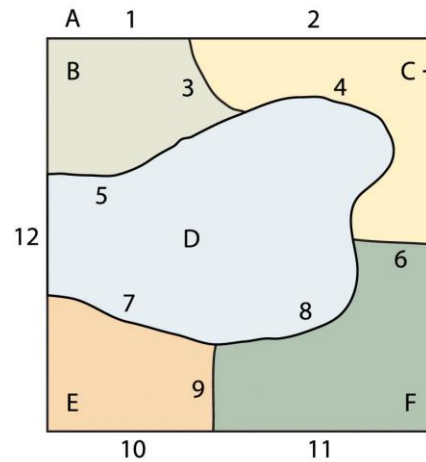
Arc-Node Topology



Arc-Node List

Arc	From-Node	To-Node
A	16	17
B	14	17
C	17	18
D	15	18
E	18	19
F	17	21
G	20	21
H	21	22

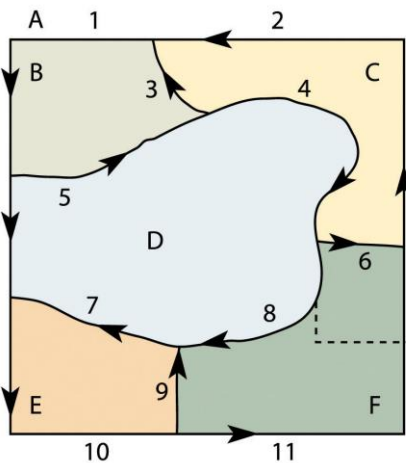
Definice plochy



Polygon-Arc Topology

Polygon	Arc List
B	1,3,5
C	2,3,4,6
D	4,5,7,8,12
E	7,9,10
F	6,8,11

Sousednost










Left-Right Topology

Arc	Left Polygon	Right Polygon
1	A	B
2	C	A
3	B	C
4	C	D
5	B	D
6	C	F
7	E	D
8	F	D
9	E	F
10	E	A

ArcGIS Shapefile

ESRI formát pro netopologický zápis GIS vektorového formátu.

- 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 – (Well Known Text)
 - **.cpg** – specifikuje kódování v dbf souboru.
 - UTF-8.

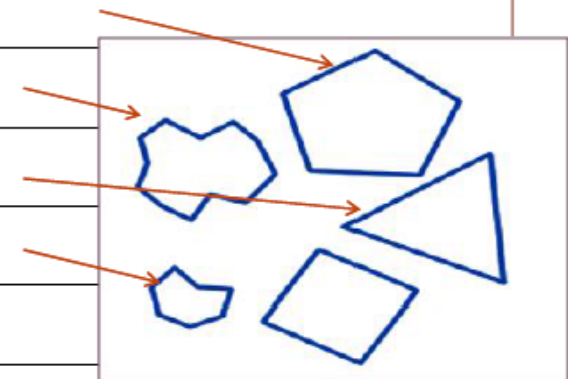
Name	Type		
 Lines.shp	Shapefile		
 Points.shp	Shapefile		 Lines.dbf
 Polygons.shp	Shapefile		 Lines.shp
		 Lines.shx	

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.

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



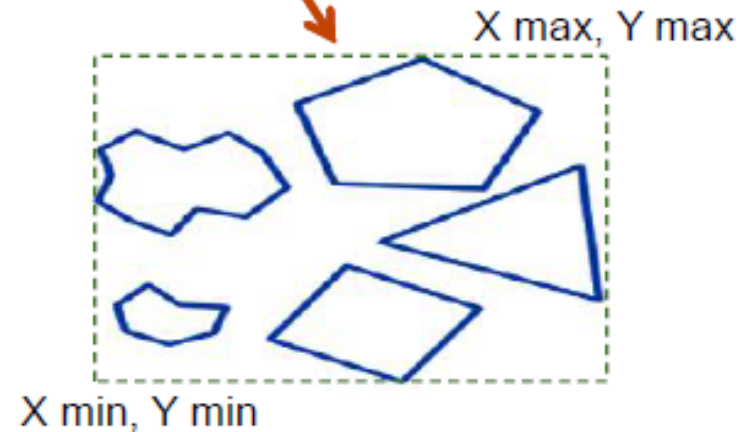
Polygon



Hlavička souboru *.SHP

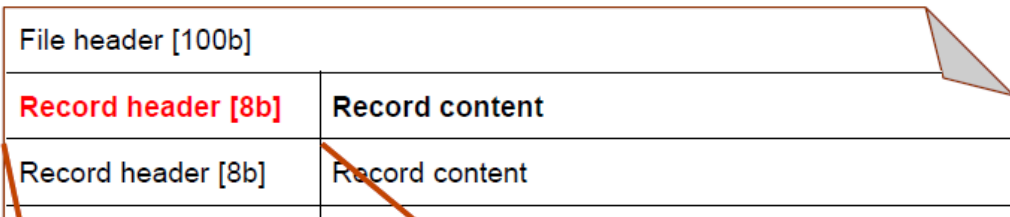
0	4	24	28	32	36	68	84
int	int	int	int	int	4 x double	2x double	2 x double
File code vždy 94440	nepouž.	délka souboru	verze	Typ shape	MBR X min Y min X max Y max	Z min Z max	M min M max

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

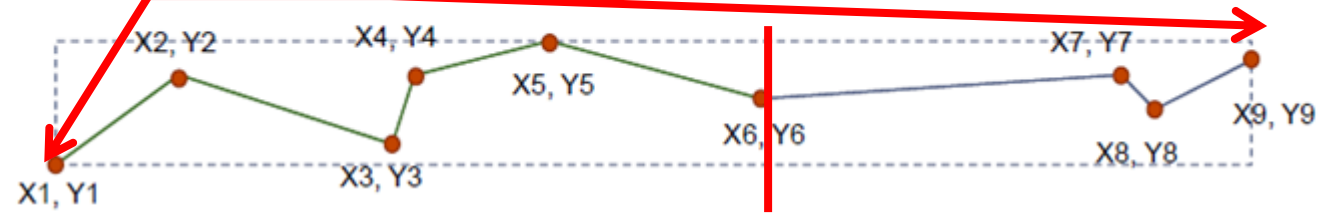


0	4
int	int
číslo záznamu	délka záznamu



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, max 2GB) je pomalá.
- Špatná podpora Unicode (kódování češtiny).



Vektorová data

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.

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.