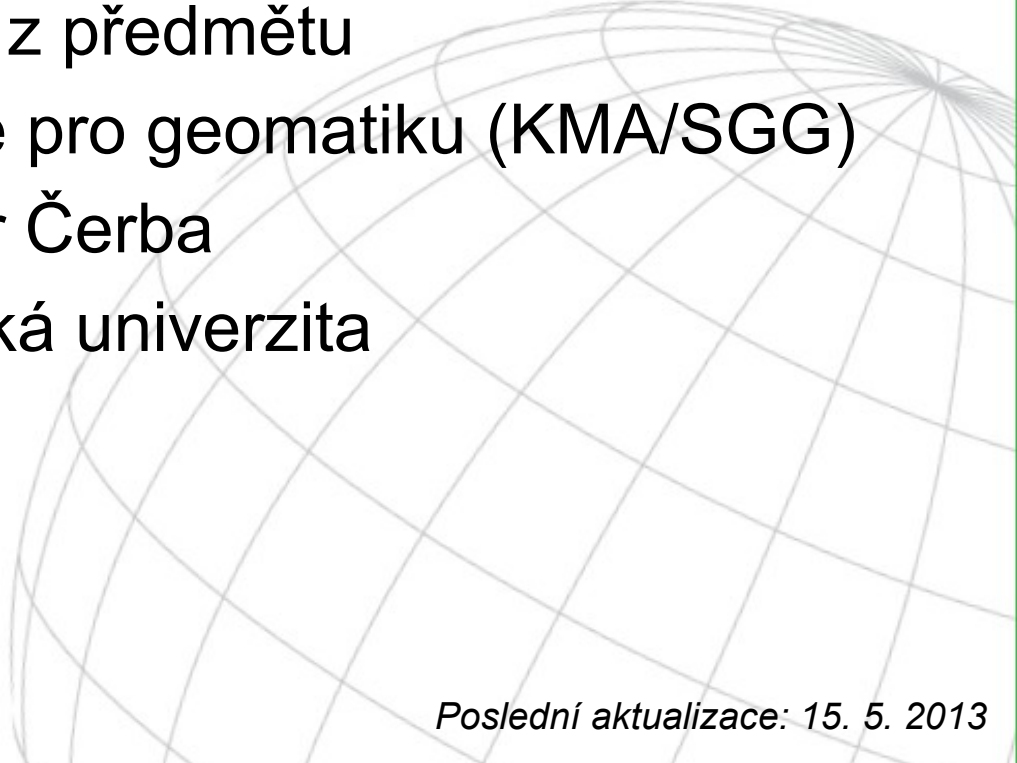


# Ontologie

Přednáška z předmětu  
Socioekonomická geografie pro geomatiku (KMA/SGG)  
Otakar Čerba  
Západočeská univerzita



# Proč ontologie v rámci geografie?

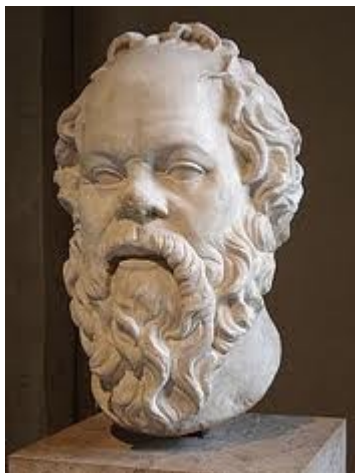
## Ontologie...

- ... zabraňují více-smyslnosti a více-významovosti
- ... podporují znovuvyužívání existujících struktur
- ... vytváří vazby a podporují komunikaci mezi geografii a jinými vědními disciplínami
- ... dokáží sjednotit a propojit odlišnou terminologii
- ... pomáhají vytvořit, propojit a sdílet různé datové modely
- ... existují a fungují bez ohledu na státní hranice, používané platformy poskytovatele dat i jejich uživatele
- ... podporují interoperabilitu prostorových dat a tvorbu SDI
- ... propojují svět prostorových dat a informací se světem Web 2.0 technologií a sémantickým webem

# Definice

- Antická filosofie: Ontologie = nauka o bytí
- T. Gruber: Ontologie = explicitní specifikace konceptualizace
- W. Borst: Ontologie = formální specifikace sdílené konceptualizace
- Wikipedia & INSPIRE: Ontology = representation of a set o concepts within the domain and the relationships between those concepts

90. léta  
20. století

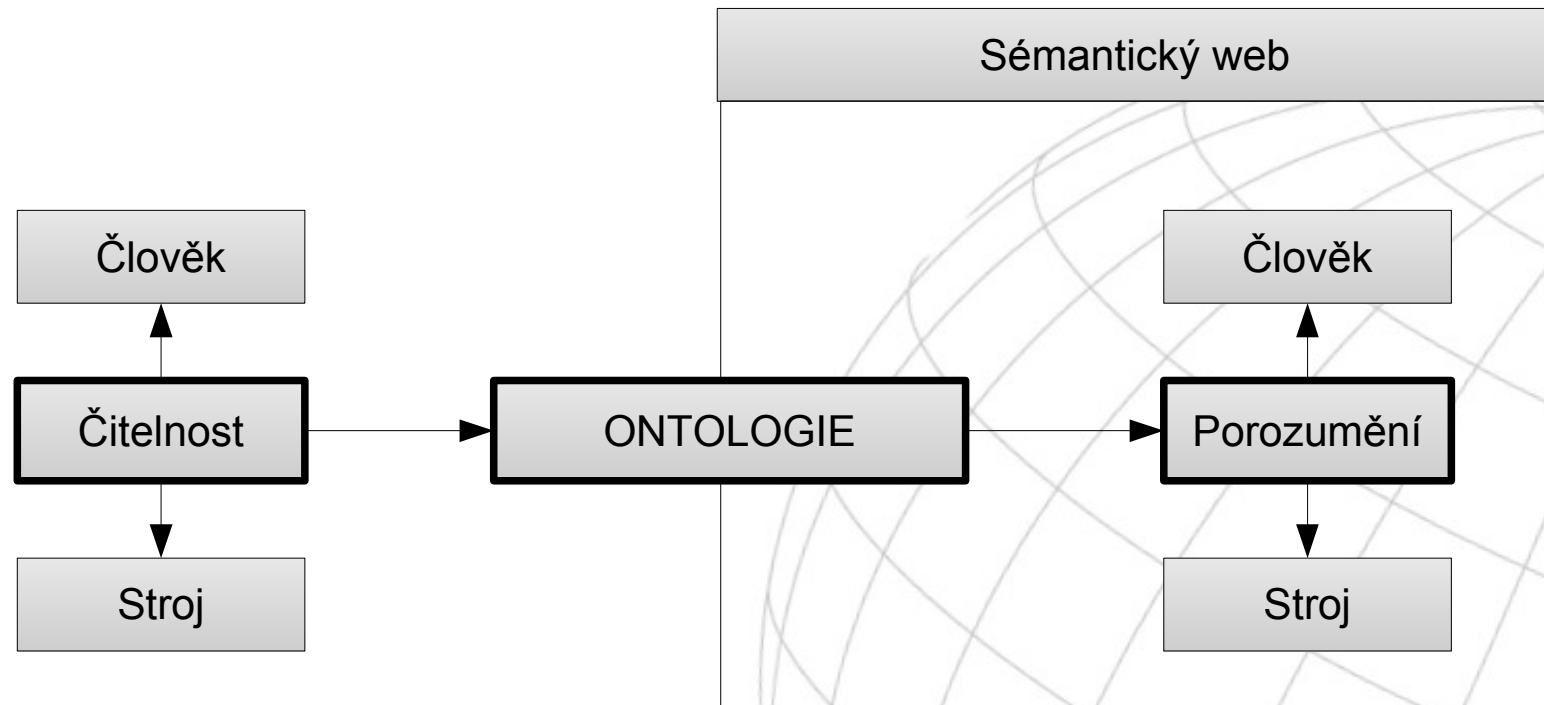


# Definice



# Vlastnosti ontologií

- Aplikačně nezávislé
- Formální & formalizovaná
- Dokáží pracovat nejen s daty, ale i s informacemi a se znalostmi



# Klasifikace ontologií

Podle historických paradigmat

Terminologické (lexikální, taxonomické) → pokročilé tesaury

Informační → rozvinutí databázových schémat

Znalostní → logické teorie a jejich vazba na realitu, umělá inteligence

Podle míry formalizace

Neformální

Semi-formální → přirozený jazyk

Formální → další členění podle charakteru jazyka

Podle předmětu formalizace

Top-level (nejobecnější)

Doménové (konkrétní oblast)

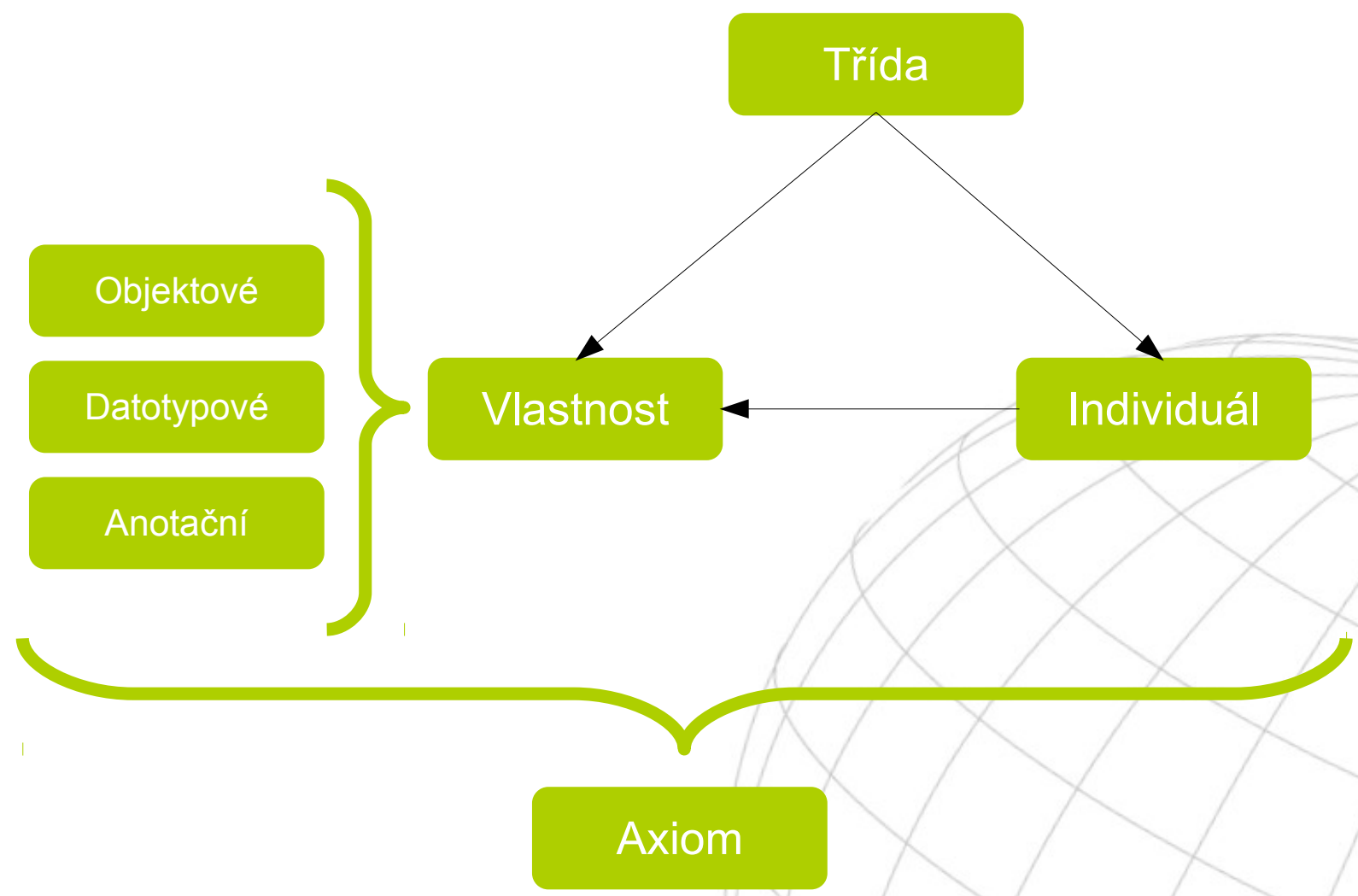
Generické (napříč doménami)

Úlohové, task (nejde o modelování reality, ale o využívání znalostních modelů → diagnostika, odvozování, plánování, hodnocení...)

Datové (zdroj dat nebo službu)

Aplikační (nejspecifičtější) – obsahují úlohovou i doménovou část

# Prvky ontologií



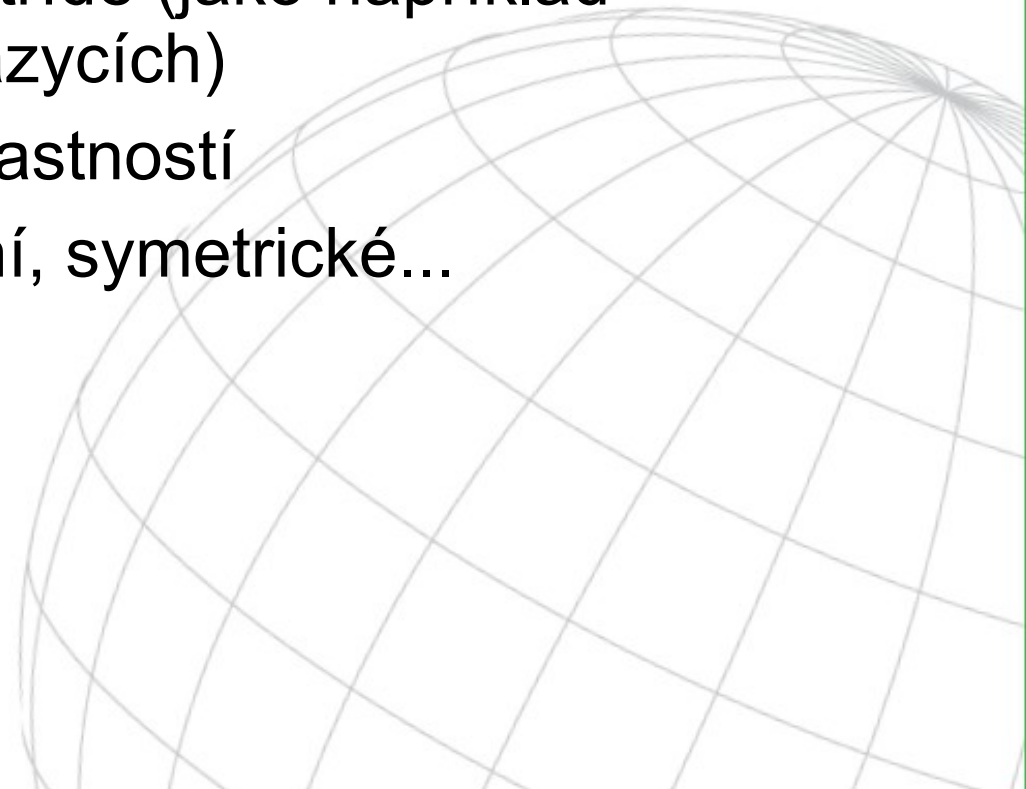
# Třídy

- Koncepty, typy, kategorie, rámce
- Základní stavební kámen – tvoří uspořádanou stromovou hierarchii (taxonomie)
- Podobné třídám z oblasti objektově orientovaných jazyků, například dědičností, ale také se liší (např. nezahrnují procedurální metody, možnosti vícenásobné dědičnosti)
- Dělení:
  - Definované (určené podmínky nutnosti i postačitelnosti)
  - Primitivní (pouze nutné podmínky nebo zcela bez definovaných podmínek)

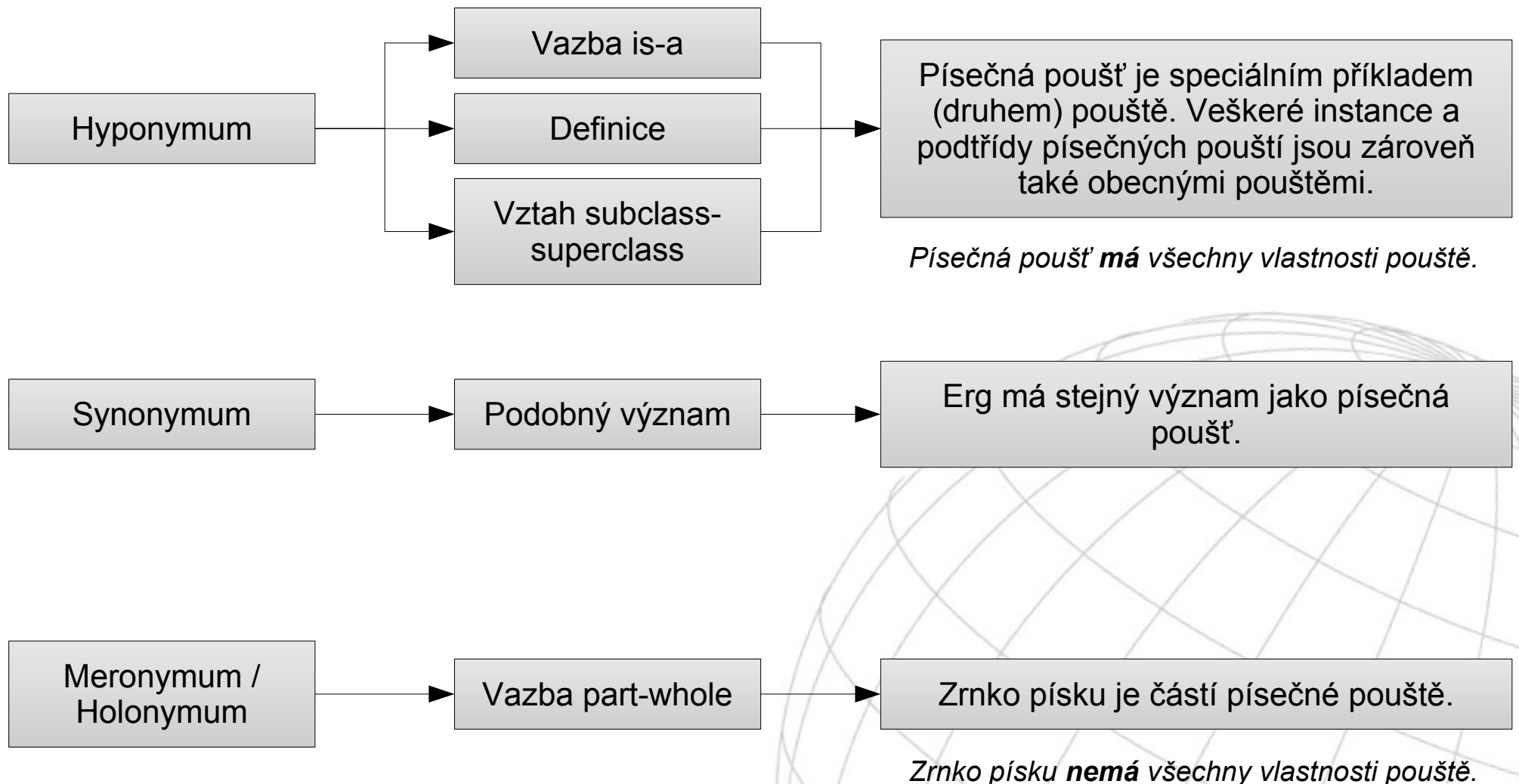


# Vlastnosti

- Relace, typ relace, funkce, sloty (většinou se vztahuje pouze na binární relace), role, „semantic link type“, atributy
- Definují vztahy mezi jednotlivými třídami a individui (nejde o atributy ve smyslu prostorových dat)
- Nejsou závislé na konkrétní třídě (jako například v objektově orientovaných jazycích)
- Možnost tvorby hierarchie vlastností
- Vlastnosti tranzitivní, inverzní, symetrické...

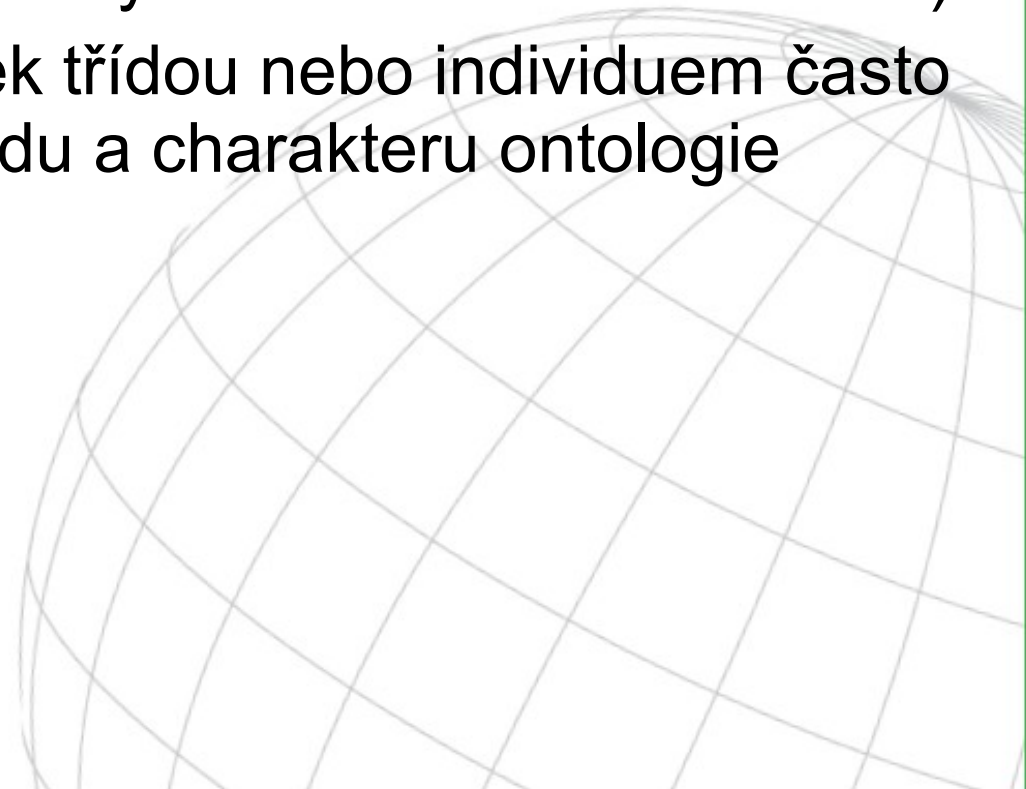


# Sémantické vztahy vlastností



# Individua

- Individuály, konstanty, objekty nebo instance
- Reprezentují reálné objekty, které již v rámci ontologie nebudou dále členěny
- Individua přibližuje ontologie jiným formám a úrovním popisu dat (například databázovým schémátům nebo UML)
- Volba, zda má být daný prvek třídou nebo individuem často závisí na subjektivním pohledu a charakteru ontologie



# Příklady prvků ontologie

- Třída: Pohoří, Hora
- Individuál: Praha, Tok, Brdy
- Vztah mezi třídami: Hora → Praha, Tok; Pohoří → Brdy, Šumava
- Objektová vlastnost: Hora **leží v** Pohoří; Praha **leží v** Brdech
- Datotypová vlastnost: Praha **má nadmořskou výšku** 862m
- Anotační vlastnost: Praha **má www** '['http://cs.wikipedia.org/wiki/Praha\\_\(Brdy\)'](http://cs.wikipedia.org/wiki/Praha_(Brdy))
- Vazby mezi vlastnostmi: **ležet v** je inverzní k **obsahovat**

# Axiomy (pravidla)

- Logické formulace – umožňují vyjádřit například ekvivalenci tříd nebo vlastností, subsumpci (podřazení konkrétnějšího pojmu obecnějšímu termínu) tříd nebo vlastností, disjunktnost tříd, rozklad třídy na podtřídy apod.



# Jazyky pro zápis ontologií (1)

- CyC – jeden z nejstarších, založený na jazyku LISP, skládá se z tzv. mikroteorií, které představují vzájemně provázaná tvrzení
- eXtensible Ontology Language (XOL) – odlehčená verze jazyka OKBC, propojení ontologií a značkovacích jazyků
- Open Knowledge Base Connectivity (OKBC) – ekvivalent k Open Database Connectivity (ODBC) – univerzální rozhraní pro přístup ke znalostním systémům, které by umožňovalo komunikaci mezi takovými systémy

Důležitá je možnost sdílení, kombinace a znovupoužívání ontologií



## Jazyky pro zápis ontologií (2)

- Ontology Inference Layer (OIL) – kombinace deskripční logiky, XML, RDF a jazyka OKBC-Lite
- DARPA Agent Mark-up Language (DAML-ONT) – vytvoření sémantického jazyka na bázi RDF, který by nabízel větší možnosti než jazyka RDFS
- DAML+OIL – propojení formátů DAML-ONT a OIL, typické konstruktory (intersectionOf , unionOf, complementOf , oneOf, toClass, hasClass, hasValue, minCardinalityQ , maxCardinalityQ , CardinalityQ ) pro tvorbu logických výrazů a 11 typů axiomů (subClassOf , sameClassAs , disjointWith , subPropertyOf , samePropertyAs , inverseOf , transitiveProperty , uniqueProperty, unambiguousProperty , sameIndividualAs , differentIndividualFrom)

# Ukázka jazyka OIL

```
<slot-def>
```

```
  <slot-name = "has-component"/>
```

```
  <inverse> <slot-name = "is-component-of"/> </inverse>
```

```
  <properties> <transitive/> </properties>
```

```
</slot-def>
```

```
<class-def> <class-name= "nucleic-acid"/> </class-def>
```

```
<class-def>
```

```
  <class-name= "rna"/>
```

```
  <subclass-of> <class name = "nucleic-acid"/> </subclass-of>
```

```
  <slot-constraint>
```

```
    <slot-name = "has-backbone"/>
```

```
    <value-type> <class name= "ribophosphate" </value-type>
```

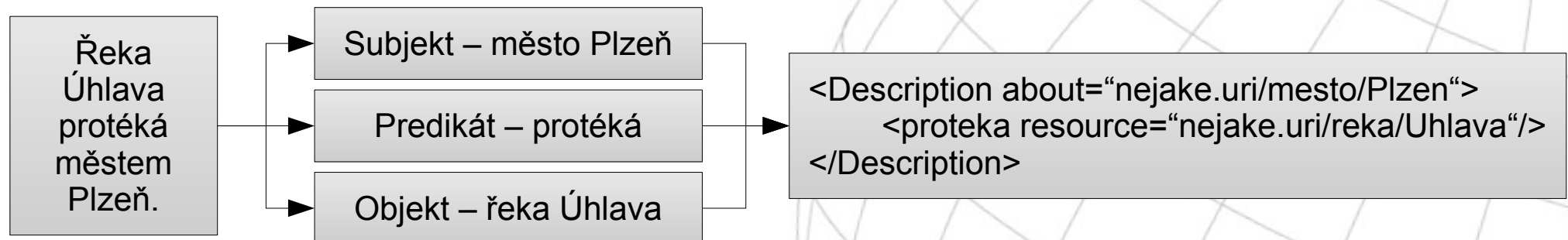
```
  </slot-constraint>
```

```
</class-def>
```



# Resource Description Framework

- W3C standard pro reprezentaci struktury metadat a popisu sémantiky → sémantický web
- Nahrazení stromové struktury grafovou strukturou
- Tři základní prvky (triple) – subjekt, predikát (vlastnost subjektu), objekt (hodnota vlastnosti)
- Subjekt a objekt – zdroj (resource) – specifikace pomocí URI (Uniform Resource Identifier) i primitivní datová hodnota



# RDF Schema (RDFS)

- Obohacení RDF o prvky z objektových systémů (třídy, relace, omezení apod.) – RDF → ontologie
- RDF vs. RDFS = XML vs. Schémové jazyky → konkrétní specifikace obecného jazyka
- Vlastnosti
  - Definuje specifický význam (způsob interpretace) určitým predikátům (subClassOf, type, range, domain) a zdrojům (Class)
  - nerozlišování mezi třídami a individuály
  - možnost používání predikátů jako subjektů a objektů – `<pravyPritok, subPropertyOf, Pritok>`

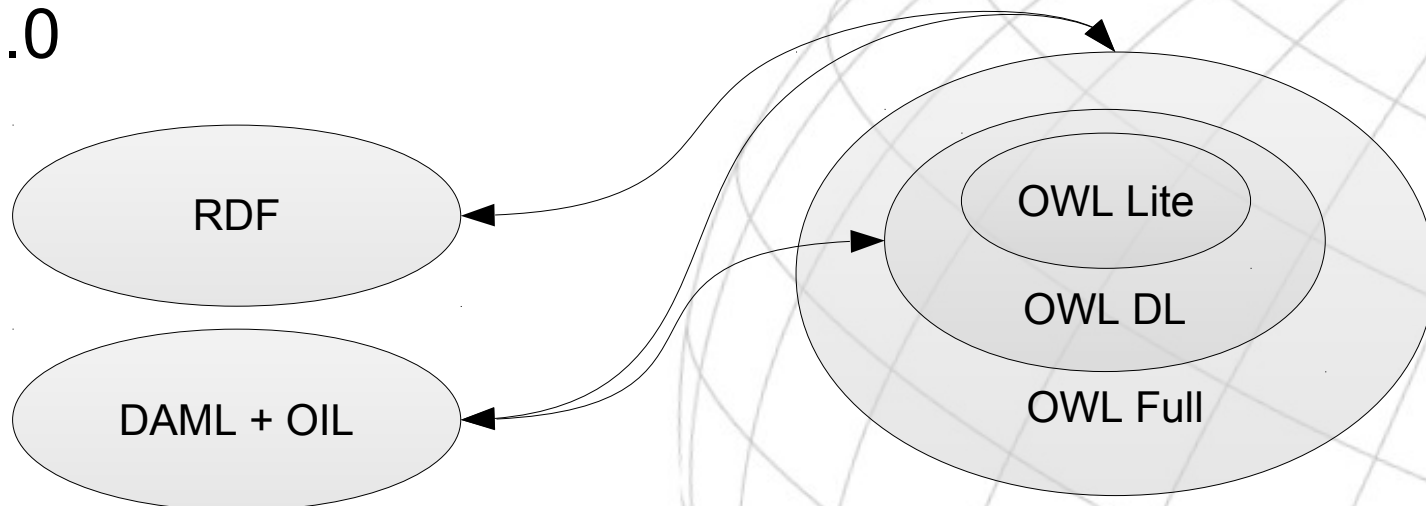
# Proč nestačí RDFS?

- RDFS (na rozdíl od ontologií) nepodporuje
  - lokalizaci (kontext) subjektů a objektů – příklad: Pokud se predikát jeSoucasti týká sídel i pohoří, RDFS nedokáže zajistit, aby součástí sídel byla pouze jiná sídla
  - kardinalitu
  - transitivitu nebo symetričnost vlastností

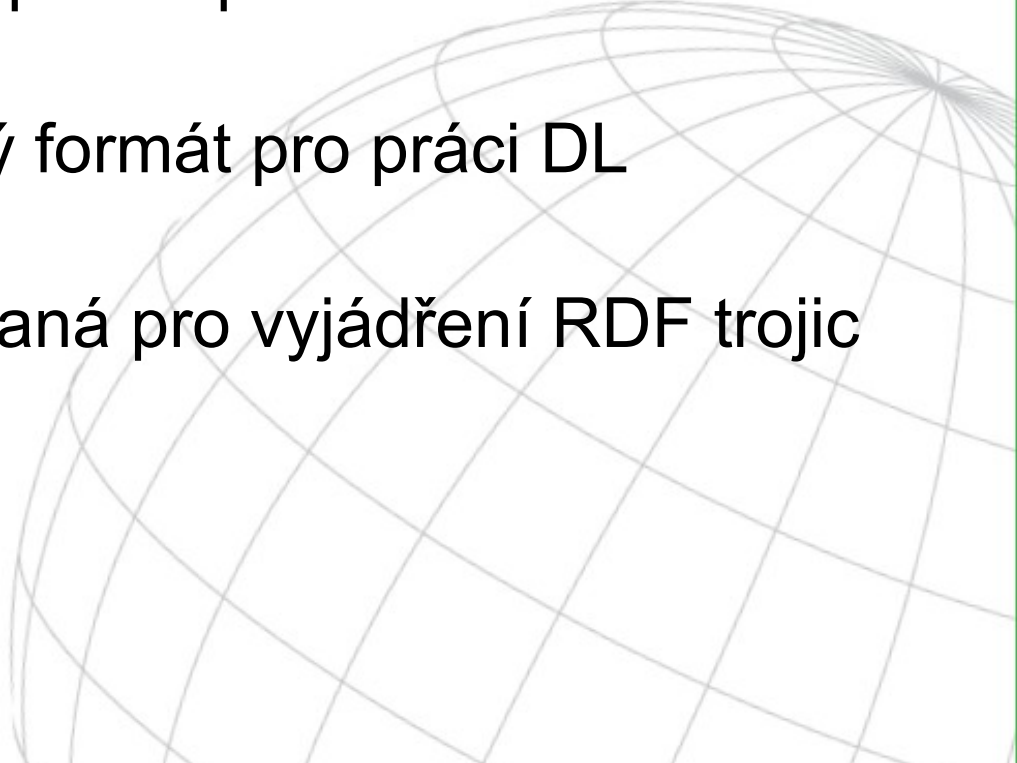


# Web Ontology Language (OWL)

- Standard W3C, který navazuje na formát DAML+OIL
- V současnosti existují tři verze OWL 1.0
  - OWL Full – sjednocení syntaxe OWL a RDF
  - OWL DL – omezení pouze na First Order Logic, tato verze je víceméně ekvivalentní DAML+OIL
  - OWL Lite – implementačně jednoduchá podmnožina OWL DL
- OWL 2.0

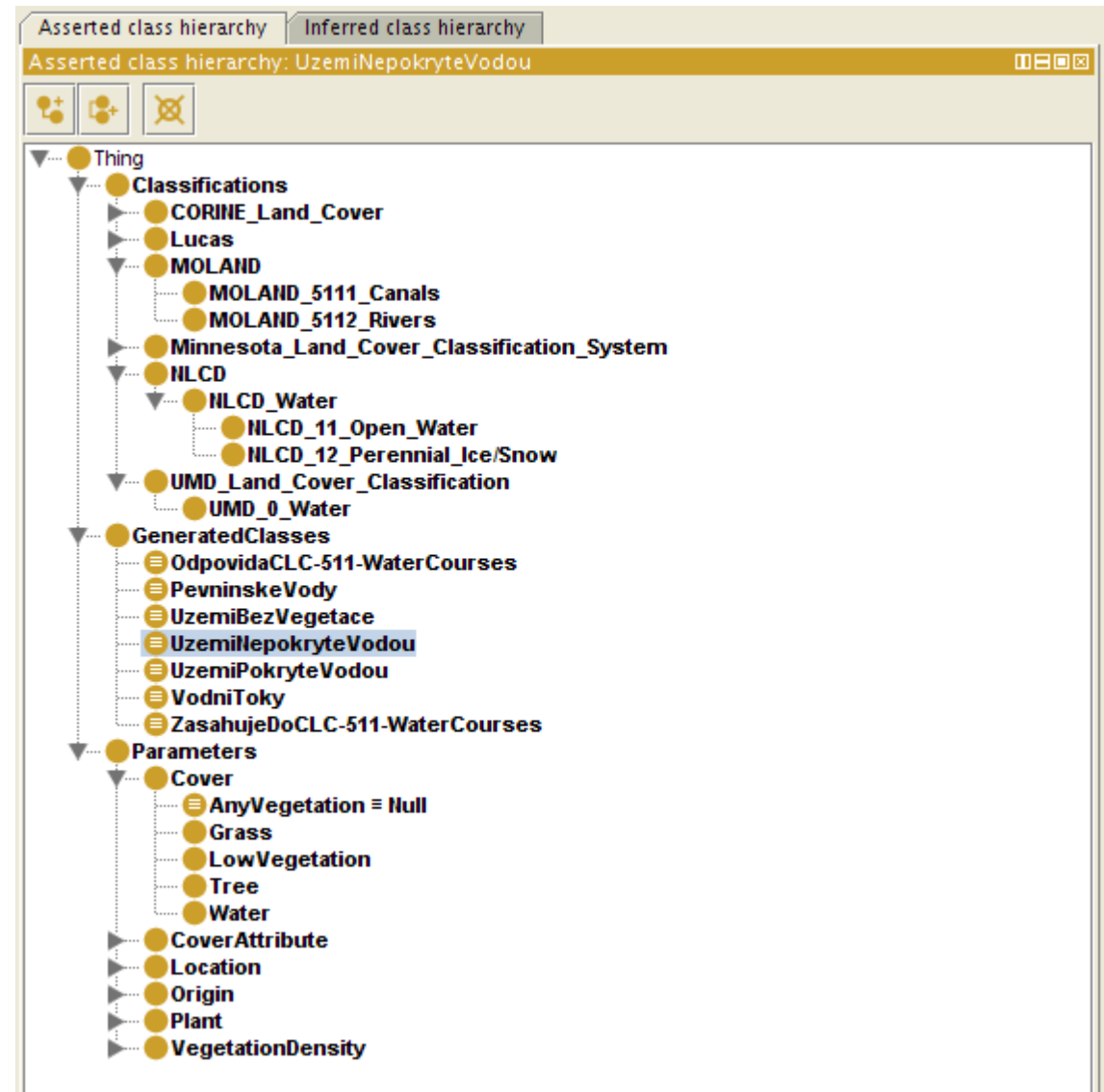


# OWL – syntaxe

- RDF/XML – základní způsob zápisu, slouží jako univerzální výměnný formát
  - OWL/XML – forma vhodná pro zpracování ontologií pomocí nástrojů pro XML.
  - Functional Syntax – ideální způsob pro ilustraci formální struktury ontologií
  - Manchester Syntax – vhodný formát pro práci DL ontologiemi
  - Turtle – forma zápisu používaná pro vyjádření RDF trojic
- 

# Protégé

- Software pro zápis a zpracování ontologií
- Klasifikátory – FaCT++ , Pellet
- Velké množství zásuvných modulů (například pro vizualizaci ontologií)



# Reasoning



Když třídy A a B mají shodné hodnoty parametrů s výjimkou vegetace, kde A je pokryta pšenicí a B obilninami, pak A bude podtřídou B...

Jestliže jedna třída má stejné hodnoty parametrů jako jiná, pak se musí jednat o třídy se stejným významem...

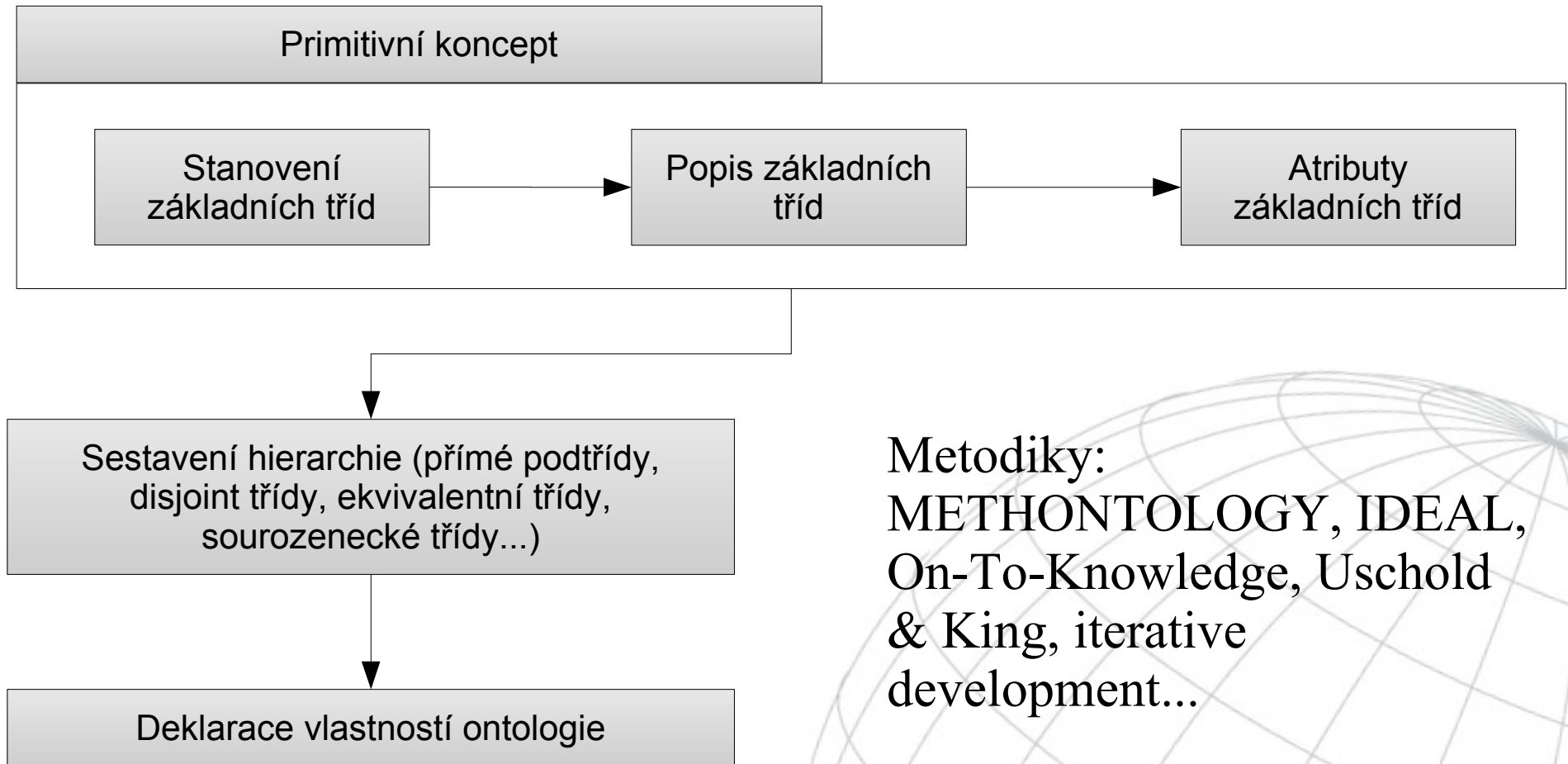


# Návrh ontologií – základní pravidla

- Neexistuje univerzální všeobjímající ontologie, protože taková by byla prakticky nekonečná.
- Žádnou doménu není možné komplexně formálně popsat (nemůže obsahovat všechny vlastnosti a omezení).
- Neexistuje jediný správný popis domény – záleží na
  - konkrétním účelu ontologie
  - úrovni detailu (podrobnosti) ontologie
  - požadavcích na kompatibilitu s dalšími ontologiemi a jinými systémy
  - subjektivních pohledech a osobních preferencích tvůrců ontologie i jejích správců a uživatelů

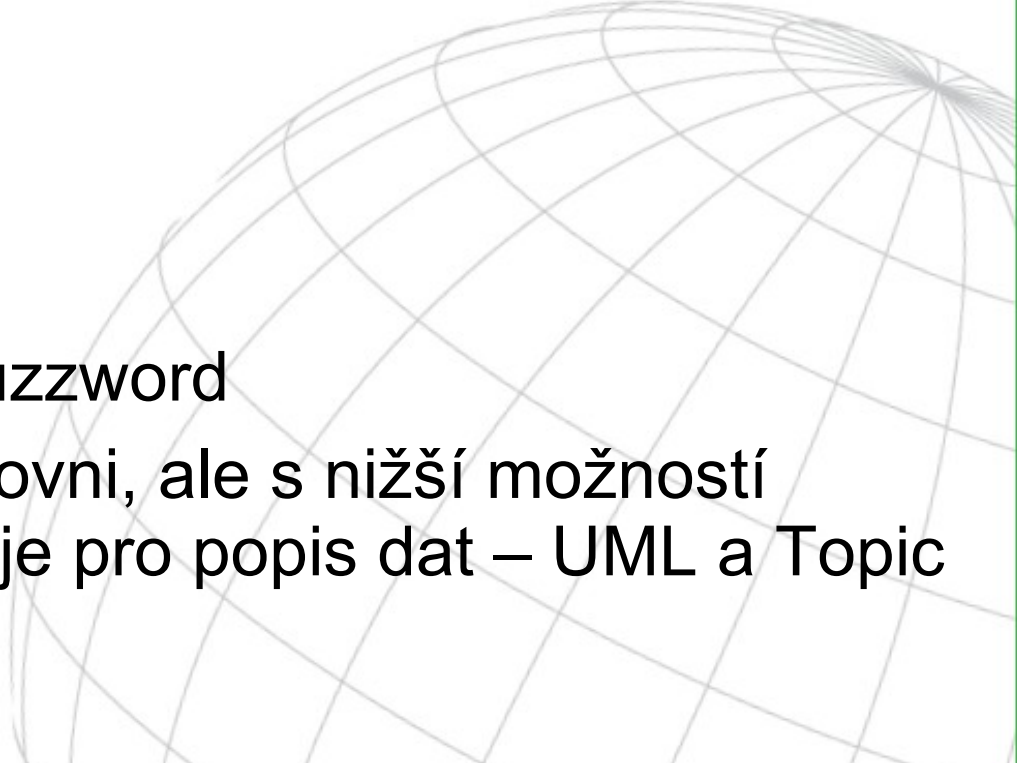


# Návrh ontologií – postup



Metodiky:  
 METHONTOLOGY, IDEAL,  
 On-To-Knowledge, Uschold  
 & King, iterative  
 development...

# Ontologie nejsou všemocné

- Představují pouze jeden pohled na doménu
  - Existují případy, kdy jejich nasazení je nevhodné nebo nedostatečně efektivní z pohledu
    - prostorových dat,
    - modelovaných procesů,
    - nasazených systémů,
    - časové náročnosti,
    - komplexity...
  - V současnosti ontologie = buzzword
  - Na podobné konceptuální úrovni, ale s nižší možností abstrakce, jsou i další nástroje pro popis dat – UML a Topic Maps
- 

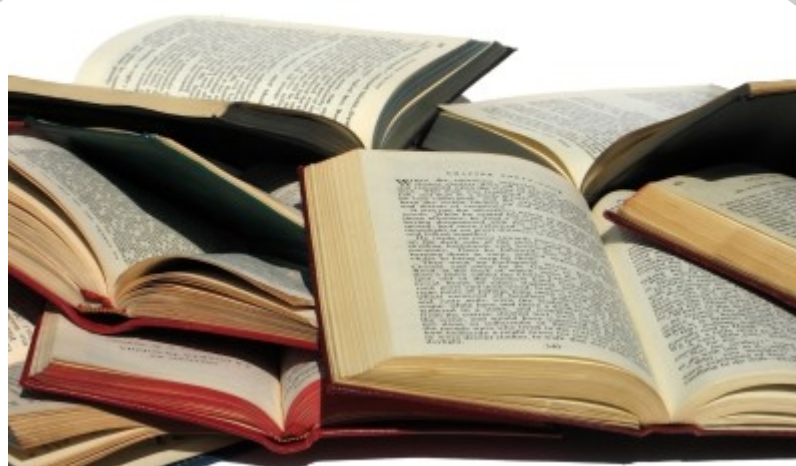
# Využití ontologií

Znalosti

Pro koho?

Počítače

Uživatelé



Čeho?

Domény

Co s nimi?

Porozumění

Znovu používání

Analýza

Sdílení

Popis

# Ontologie & GI

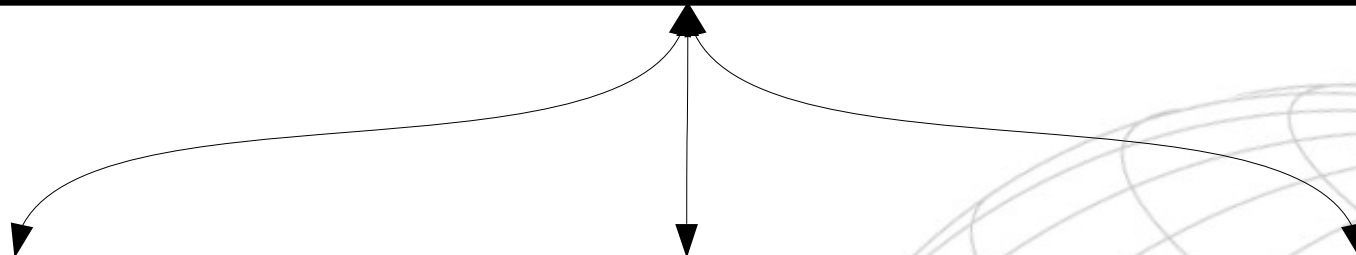
Konstrukce konceptů a vztahů mezi koncepty



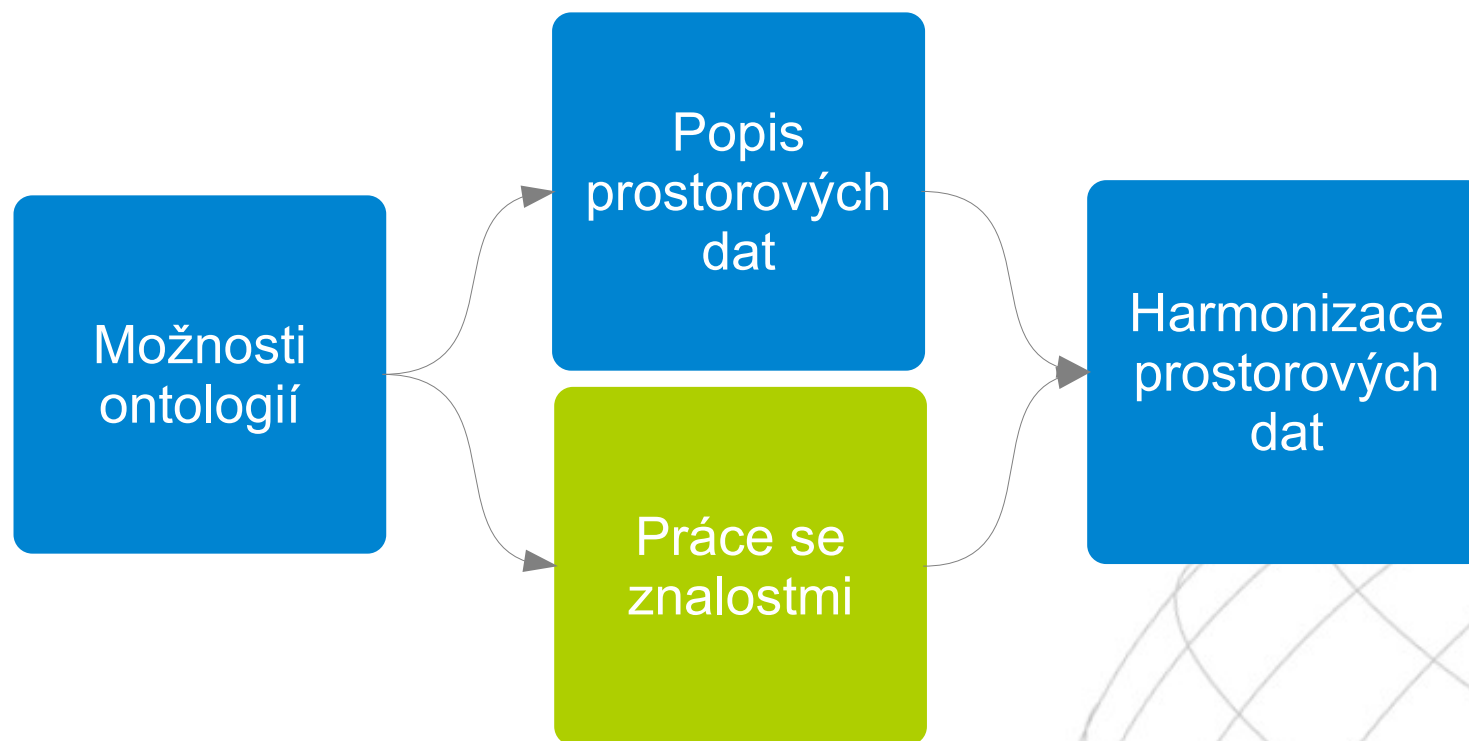
Datové modely  
& popis  
prostorových dat

Slovníky &  
tezaury

Harmonizace &  
transformace



# Ontologie v GI



# Ontologie vs. datové struktury

Open vs.  
Closed World  
Assumption

Znalosti  
vs. data

Nižší vs.  
vysoká úroveň  
abstrakce

Značkovací  
jazyky

Slabší vs. silná  
podpora  
sémantiky

Návrh nových  
struktur vs. popis  
existujících  
struktur

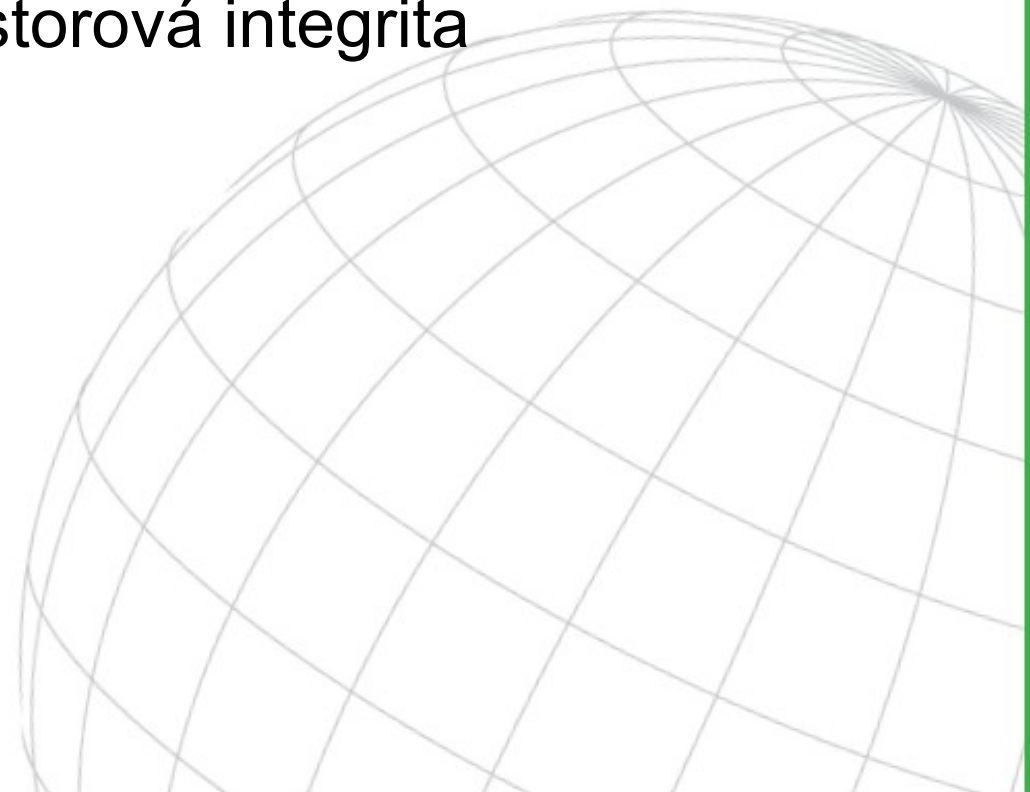
Aplikační  
závislost

Mezinárodní  
standardy

Polyhierarchie  
vs. stromové  
struktury

# Geo-ontologie – specifické vlastnosti

- Prostorové vazby
- Preference A-Box konceptu (místo T-Box)
- Sémantika na úrovni individuálů
- Obohacení mereologických vazeb o vazby topologické
- Logická + topologická a prostorová integrita



# Geo-ontologie – Koncepty

Koncepty – v případě geo-ontologií je nutné rozlišovat dva hlavní pohledy – **konstrukci ontologií z obecných geografických elementů** (například řeka nebo pohoří) nebo z takových prvků, které už představují **prostorová data** (jako modely geografických elementů). Ve druhém případě pak můžeme využívat všechny atributy prostorových dat, zatímco v prvním je možné pracovat pouze na úrovni konceptů (ve smyslu základních ontologických kategorií). **V praxi představují třídy geo-ontologie konceptuální pohled, zatímco individuály reprezentanty prostorových dat.**



## Geo-ontologie – Hierarchie

Běžná vazba **třída-podtřída**, která vytváří základní kostru konceptů, reprezentuje vztah typu is-a-part – je součástí. Tato vazba nemusí být realizována za použití objektové vlastnosti na rozdíl od podobného vztahu část-celek. Zatímco informace „Brdy jsou součástí Poberounské subprovincie“ bude vyjádřena vazbou třída „Brdy“ je podtřídou třídy „Poberounská subprovincie“. Pokud bychom ve stejné ontologii chtěli vyjádřit vztah mezi Padrt'skými rybníky (které leží na území Brd) a Brdy, pak je nutné použít objektovou vlastnosti (jinými slovy není možné říct, že Padrt'ské rybníky jsou speciální částí Brd). Otázky spojené se s korektní tvorbou hierarchického systému naplatí pouze pro geo-ontologie, ale v nich je tento problém akcentován ještě topologickými vazbami.

# Geo-ontologie – Terminologie

Hledání správných a jednoznačných termínů představuje další výrazný problém při tvorbě geo-ontologií. Například ve výše uvedeném příkladu je možné použít název „Brdy“ nebo „Brdská vrchovina“. Problematika terminologie na geografické doméně je také speciálně akcentována, neboť prostorová data nestojí většinou sama o sobě, ale jsou úzce propojena s atributovou složkou, která může zasahovat do nejrůznějších vědeckých oborů interagujících s geografii. Na druhou stranu terminologické problémy mohou být v současnosti řešeny pomocí nástrojů, které jsou podobně jako ontologie součástí sémantického webu.

# Geo-ontologie – Multilingualita

Ontologie by měly představovat univerzální komunikační nástroj, proto je nutné řešit i problematiku překladu označení jednotlivých konceptů a vlastností (multilingualita de facto představuje specifický případ výše uvedených potíží týkajících se terminologie). Z hlediska obsahu jsou řešením tezaury (například GEMET), z hlediska syntaxe je vhodné používat anotační vlastnosti s uvedením jazyka.



# Geo-ontologie – Souřadnice

V případě geo-ontologií vyvstává „nerudovská“ otázka „...kam s nimi...“, aby celou geo-ontologii zbytečně nezatěžovaly z hlediska velikosti souboru. Většina geo-ontologií, které používají souřadnice (zpravidla jako datotypovou vlastnost) pracuje na úrovni bodů, kde každý koncept (většinou individuál) obsahuje dvě, případně tři souřadnice. Mnohem větší problém nastává v případě linií nebo ploch. V současné době neexistuje zcela uspokojivé řešení tohoto závažného problému, který výrazně zpomaluje vyhodnocování a dotazování nad geo-ontologiemi. Je sice možné připojovat souřadnice externě pomocí URI (Uniform Resource Identifier). V tomto případě sice vlastní ontologický systém bude poměrně rychlý a nenáročný z hlediska prostoru, ale na druhou stranu jakékoli operace spojené s geometrií budou muset být prováděny nestandardním způsobem.

Děkuji za pozornost

**Linked in**

<http://cz.linkedin.com/in/otakarcerba/>

