

POPIS DAT PRO GEOINFORMAČNÍ TECHNOLOGIE

OTAKAR ČERBA

Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra matematiky, oddělení geomatiky



Obsah

POPIS DAT PRO GEOINFORMAČNÍ TECHNOLOGIE.....	1
Úvod.....	3
Prostorová data.....	5
Problematika definic.....	5
Vlastnosti prostorových dat.....	11
Klasifikace prostorových dat.....	12
Popis prostorových dat.....	14
Příklad 1: Proces porozumění.....	14
Příklad 2: Typy metadat a sémantického popisu.....	15
Příklad 3: Standardy pro popis dat.....	16
Shrnutí.....	17
Syntaxe a struktura dat.....	18
Jazyky pro popis dokumentu.....	18
Vybrané mezinárodní výměnné standardy pro prostorová data.....	18
Metadata.....	19
Sémantika.....	21
Termín „sémantika“.....	21
Sémantika a prostorová data.....	25
Zápis sémantických informací.....	27
Ontologie.....	28
Jak zapsat ontologie.....	31
Resource Description Framework a RDF Schema.....	33
Web Ontology Language (OWL).....	34
Ontologie a software.....	34
Ontologie – příklady a ukázky.....	35
Interoperabilita a standardizace.....	37
Spolupráce na poli GIT.....	37
The Open Source Geospatial Foundation (OSGeo).....	37
Přístupnost dat a její omezení.....	39
Licence dat a software.....	39
Open source.....	39
Licence.....	43
Harmonizace a mapování dat.....	46
Zdroje dat a jejich vyhledávání.....	47
Závěr.....	48
Bibliografie.....	49
Přílohy.....	52
Požadavky a zdroje.....	52
Přílohy v digitální podobě.....	53

Úvod

Prostorová data¹ představují základní stavební kámen veškerých geoinformačních technologií (GIT), aplikací a služeb. Jejich důležitost mohou přiblížit i následující body:

- 70-100% z veškerého množství dat má prostorový aspekt,
- při tvorbě geografického informačního systému cca 70% procent nákladů tvoří finanční prostředky vynaložené na data,
- základ veškerých kartografických a geovědních produktů tvoří geografická data apod.

V současnosti přestává být zásadním problémem většiny geovědních oborů množství dat, ale především jejich kvalita, včetně možnosti sdílení prostorových dat a kombinace datových sad z různých zdrojů. Přesto převažuje do jisté míry anachronický přístup, kdy se mnoho uživatelů nerozhoduje o datech na základě jejich kvality a vhodnosti pro dané řešení, ale na základě pouhé dostupnosti (nikoli přístupnosti). Je nutné si uvědomit, že charakter prostorových dat (často prezentovaný ve formě popisných informací jako jsou metadata nebo různé úrovně modelů) do jisté míry předurčuje způsob zpracování a vizualizace prostorových dat, včetně kartografických metod vizualizace. Můžeme hovořit o tzv. data-driven přístupu². Procesy zpracování a vizualizace prostorových dat by měl vznikat na základě shody vzešlé z požadavků autorů, zadavatelů a zákazníků, ze znalostí a informací týkajících se dané problematiky, z legislativních předpisů, technologických požadavků a také charakteru používaných prostorových dat.

Účelem tohoto dokumentu není seznámit čtenáře s kompletní problematikou prostorových dat. Tato oblast je natolik široká a dynamická, že byly zvoleny pouze některé typy prostorových dat (především data středních a velkých měřítek) a některé aspekty těchto dat související s jejich kvalitou. Především se jedná o význam a popis dat, jak uvádí titlek. Prostorová data v tomto textu jsou popisována především s ohledem na budování Spatial Data Infrastructure (SDI), což v současnosti představuje jednu z fundamentálních otázek geomatiky, geoinformatiky a ostatních geovědních disciplín. Cílem tohoto dokumentu je také základní přehled problematiky interoperability, standardizace, sémantiky, harmonizace a přístupnosti prostorových dat. Tento text se orientuje především na popisnou (atributovou) složku prostorových dat, proto nejsou uvedeny části věnující se geometrickým a topologickým aspektům geografických dat, technologickým otázkám, problematice syntaxe (která bude zmíněna pouze okrajově) ani například problematice přesnosti (ty jsou na Západočeské univerzitě v Plzni přednášeny v jiných předmětech a kurzech, které zajišťují pracovníci oddělení geomatiky).

Dokument bude, vzhledem k dynamice vývoje GIT, průběžně doplňován a aktualizován. Proto se zřejmě nikdy nebude jednat o kompletní, celistvý text, ale jednotlivé, částečně nezávislé a především postupně doplňované kapitoly. Kde to nebude vyloženě nutné, nebude citován externí zdroj, ale bude na něj uveden pouze odkaz (jinými slovy tato publikace neslouží jako vyčerpávající zdroj informací, ale spíše jako jakýsi základní průvodce problematikou). Podobně se také bude postupovat při překladech z anglického jazyka nebo slovenského jazyka – některé citované pasáže mohou být uvedeny bez českého překladu, který by mohl původní text znehodnocovat.

Čtenář by se měl postupně seznámit s následující tematikou³:

- Definice prostorových dat, problematika názvosloví, klasifikace a popis prostorových dat

1 Problém terminologie bude zmíněn v následujících kapitolách.

2 Bohužel v současné době většina uživatelů inklinuje spíše k „software-driven“ přístupu.

3 Jedná se o zamýšlenou strukturu textu, která není definitivní a zatím také není zcela naplněna. Zároveň jde o částečný syllabus předmětu KMA/SGG.

- Syntaxe dat – jazyky pro popis dokumentu (schémata), vybrané mezinárodní výměnné standardy (GML, LandXML apod.)
- Metadata, včetně existujících standardů (INSPIRE, ISO) a metadatových systémů
- Sémantika dat (s důrazem na prostorová data), formát RDF a další
- Ontologie
- Interoperabilita a standardizace
- Harmonizace a mapování (sémantické, ontologické) dat
- Přístupnost dat a její limity
- Zdroje prostorových dat a jejich vyhledávání

Prostorová data

Problematika definic

Na začátku této kapitoly je uvedeno několik definic, které by měly ilustrovat problematiku a nejednoznačný pohled na data a informace s geografickými nebo prostorovými aspekty. Úvodní orientační seznam představuje vybrané definice pojmů označujících data obsahující prostorovou složku. Definice byly získány po zadání sémantického parametru *define*: ve vyhledávači Google (www.google.com).⁴

- Geodata
 - Geoinformation is an abbreviation of geographic information. (<http://en.wikipedia.org/wiki/Geodata>)
- Geo-data
 - Geo-data refers to the full spectrum of digital geographic data. Examples include: digital maps, raster image data, point vector, vector data, spatial/temporal data etc. (<http://www.geospatial.govt.nz/glossary/>)
- Geographic data
 - Geographic data is about much more than electronic pictures of maps. (http://en.wikipedia.org/wiki/Geographic_data)
 - The locations and descriptions of geographic features. The composite of spatial data and descriptive data. (<http://www3.richlandmaps.com/#eh>)
- Geographical data
 - Data that record the location and value characterising the phenomenon. (<http://learnline.cdu.edu.au/units/ses501/tools/glossary.html>)
- Spatial data
 - Any information about the location and shape of, and relationships among, geographic features. This includes remotely sensed data, map data, and digital data. In the context of digital data, the term may refer to a coverage, theme, or layer. (<http://www.umass.edu/tei/ogia/parcelguide/Glossary.html>)
 - Data that can be linked to locations in geographic space, usually via features on a map. (http://atlas.library.arizona.edu/glossaryq_s.htm; Clarke, K.C. Getting Started with Geographic Information Systems. 4th edition Upper Saddle River, N.J., 2003)
 - Any information with a spatial component. (http://www.colorado.edu/geography/courses/geog_2002_sm05/raleigh_articles/%2520for%2520class/terms.doc)

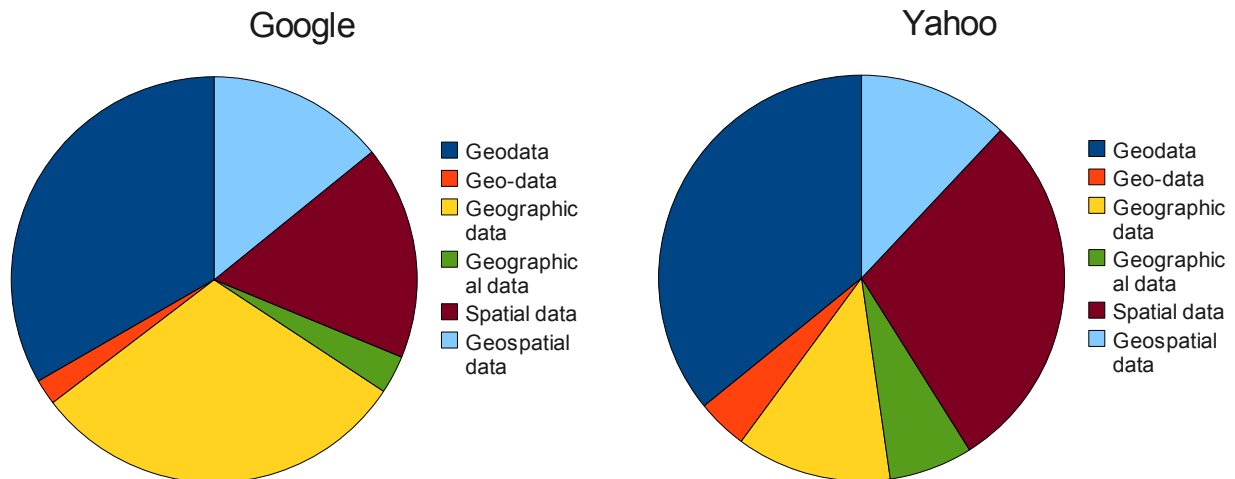
4 Problémem je fakt, že možnosti sémanticky označit definici využívá především Wikipedia, která zatím nemůže být považována za zdroj stoprocentně validních informací především v oblasti vědy a výzkumu. Na druhou stranu servery Google a Wikipedia patří mezi nejobsáhlejší a nejpoužívanější zdroje informací v současnosti (což částečně ospravedlňuje jejich využívání v tomto textu; druhým argumentem je využití sémantiky, alespoň v rámci současných možností).

V tomto případě („spatial data“) našel vyhledávač Google největší množství definice, celkem deset.

- Geospatial data
 - Information that identifies the geographic location and characteristics of natural or constructed features and boundaries on the earth. (<http://www.dir.state.tx.us/tgic/pubs/digtex2004/appendix3.htm>)
 - Any data with a geospatial reference. For example, an Ordnance Survey National Grid reference, postal address, place name or administrative area. (<http://www.geospatial.govt.nz/glossary/>)
 - Data and information that are referenced to a location on the Earth's surface through precise scientific coordinates such as maps, charts, air photos, satellite images, and land and water surveys. (<http://www.ic.gc.ca/eic/site/trm-crt.nsf/eng/rm00196.html>)
- Geoinformation
 - Geoinformation is an abbreviation of geographic information. (<http://en.wikipedia.org/wiki/Geoinformation>)
- Geographic information
 - Information in a digital map that records the physical position and shape of a map feature. (<http://www.angelfire.com/ak/ajitm/gisterms.htm#G>)
 - Information about objects or phenomena that are associated with a location relative to the surface of the Earth. A special case of spatial information. (<http://www.geospatial.govt.nz/glossary/>)
- Geographical information (vyhledávač nenalezl žádnou definici)
- Spatial information
 - Spatial information (also known as geographic information) is any information that can be geographically referenced, i.e. describing a location or any information that can be linked to a location. (http://www.anzlic.org.au/glossary_terms.html#S)
 - Data that has a geographical reference to a location on the earth's surface. This includes latitude and longitude co-ordinates, street address and lot number on plan. (<http://www.nrw.qld.gov.au/museum/glossary.html>)
 - Describes the transient and high frequency components of sound that allows our ears to determine where sounds are coming from (localization). It is also used to describe how accurately or distinctly speakers place recorded instruments in the stereo sound field. (http://www.musiciansfriend.com/document?doc_id=102222). Tato definice je uvedena spíše jako kuriozita ukazující, že podobné termíny jako prostorová informace se používají například v hudbě.
- Geospatial information
 - Information that is reference by a geographic location. (<http://www.cs.cornell.edu/wya/DigLib/MS1999/Glossary.html>)

Opět využijeme majoritní vyhledávače webových stránek Google a Yahoo a pokusíme se najít četnost výskytu výše zmíněných pojmů (konkrétně termínů geodata, geo-data, geographic data, geographical data, spatial data, geospatial data). Následující graf ukazuje výsledky tohoto jednoduchého průzkumu, kde převažuje termín geodata následován výrazy spatial data a geographic

data.



Doposud jsme se zabývali pouze přívlastkem v označení dat s prostorovými aspekty. Kromě termínu „data“ se ovšem se stejným nebo podobným významem používají i jiné termíny – informace, údaje, features. Pro srovnání pojetí problematiky terminologie v České republice a ve světě byly použity definice publikované předními českými odborníky na názvosloví v oblasti GIT.⁵

- Údaj

Údaj je základní jednotkou informačního obrazu (popisu) objektu reálného světa, zaznamenanou zpravidla v neformalizované podobě. (Rapant, 2006)

Údaj je reprezentace skutečností, pojmů nebo instrukcí (návodů, pokynů) zpravidla v neformalizované podobě. (Šíma, 2003)

- Vědomost, poznatek, znalost

Data, jejichž informace se nemění v čase, nejsou závislá na uživateli a mohou tvořit mimořádně komprimovanou formu; tím umožňují lidem i strojům, aby je efektivně uchovávali a manipulovali s nimi. V prostoru dat jsou vědomosti navigačním systémem pro vyhledávání informací a odvozování jiných vědomostí. Existují různé kategorie vědomostí, obvykle se rozlišují deklarativní (odpovídající na otázku Co je to?) a procedurální (odpovídající na otázku Jak se to dělá? popř. Jak to probíhá?). Reprezentace vědomostí může být jazyková ve formě tvrzení, jinou formou jsou sémantické sítě. (Benda, 2005)

- Informace

1. Znalost týkající se objektů, jako jsou fakta, události, věci, procesy nebo myšlenky, včetně koncepcí, která má v jistém kontextu zvláštní význam.
2. Význam, který člověk přisuzuje datům.
3. Smysluplná interpretace dat a vztahů mezi nimi. (Šíma, 2003)

Význam či hodnota přisuzované datům; významy jsou datům přiřazovány na základě vědomostí (proces se nazývá interpretace). (Benda, 2005)

⁵ Zajímavý pohled na téma geografické objekty, prvky apod. poskytuje článek „Features, Objects, and other Things: Ontological Distinctions in the Geographic Domain“ autorů D.M. Marka, A. Skupina a B. Smithe (<http://ontology.buffalo.edu/smith/articles/COSIT01MSS.pdf>).

- Data

Data představují reprezentace skutečností, pojmů nebo instrukcí (návodů, pokynů) ve formalizované podobě vhodné pro komunikaci, interpretaci a zpracování lidmi nebo automatickými prostředky.⁶ (Rapant, 2006)

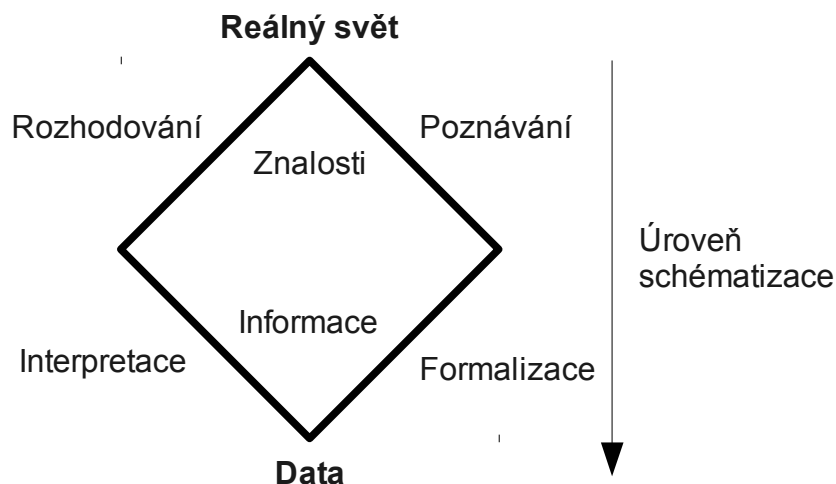
Data jsou opakovatelná reprezentace informace formalizovaným způsobem, vhodným pro komunikaci, interpretaci nebo zpracování. (Šíma, 2003)

- Jevy, vzhledy prvků (features)

Pojem „feature“, překládaný jako vzhled prvku, je podle (Šíma, 2003) definován jako

- Abstrakce jevů reálného světa.
- Modelový obraz geografické entity reálného světa, který je dále nedělitelný na jednotky stejné třídy nebo jako sada takových entit se společnou hodnotou atributu.
- Útvar ve vrstvě prostorových dat (jako například bod, linie nebo polygon), který reprezentuje geografický objekt.

Následující schéma ukazuje vazbu dat, informací, znalostí a objektů reálného světa. Zároveň obsahuje také procesy, které vedou ke vzniku dat (poznávání a formalizace) a zároveň k přenosu dat do objektů reálného světa (interpretace a rozhodovací procesy). V těchto procesech se nezastupitelná role informací a znalostí, které představují přechod od reálného světa k digitálním datům, ale zároveň jsou důležité z hlediska procesů poznávání a rozhodování, protože k těmto činnostem nestačí pouhá data. Na druhou stranu je potřeba velice pečlivě volit a vytvářet data, abychom mohli získat konkrétní informaci a užitečné znalosti.



Zpět k definicím základního pojmu tohoto dokumentu. Existuje velké množství dalších definic a jejich vzájemných vazeb⁷. Která varianta je správná nebo alespoň nejvhodnější? Budeme-li se chtít dohodnout na jednotné definice, bylo by nejvhodnější se obrátit na některou mezinárodní autoritu vyvíjející a spravující standardy a normy. Jako nejvhodnější se jeví organizace ISO (International Organization for Standardization, www.iso.org), která v rámci TC (Technical Committee) 211 definuje standardy z oblasti GIT. V názvu příslušného TC i v označení dalších norem se objevuje termín „geographic information“. Označení „geographic information“ nebo „geografická informace“ se vyskytuje například v normách CEN (European Committee for Standardization), v

⁶ ČSN 36 9001 oba termíny (data i údaj) považuje za synonyma. (Rapant, 2006)

⁷ Například vztah mezi geografickými a prostorovými daty je například popsán v publikaci (Jedlička, 2003).

názvech organizací a skupin jako například EUROGI (European Umbrella Organisation for Geographic Information; <http://www.eurogi.org/>)⁸ nebo UNGIWG (United Nations Geographic Information Working Group; <http://www.ungiwg.org/>). Tento termín je možné najít v názvu publikace (Neumann, 1996). Budeme-li respektovat terminologii ISO a další zdroje, pak se v českém jazyce měl objevovat výraz „geografická informace“. Norma ISO 19104 – Geographic information – Terminology uvádí definice následujících termínů:

- Geographic data
 - Data with implicit or explicit reference to a location relative to the Earth.
- Geographic feature
 - Representation of real world phenomenon associated with a location relative to the Earth.
- Geographic information
 - Information concerning phenomena implicitly or explicitly associated with a location relative to the Earth.⁹

S termínem „geographic information“ pracuje také projekt GINIE (Geographic Information Network In Europe; <http://www.ec-gis.org/ginie/>)¹⁰. V publikaci (Craglia et al., 2003) je uvedena definice tohoto pojmu – „the term geographic information (GI), or geo-spatial information, refers to any information record that has a reference to the Earth’s surface or near-surface. It does therefore include information on underground networks and geological features, as much as objects and features on the Earth’s surface such as parcels, buildings, rivers, transport networks, and urban and rural settlements. The term spatial is broader in the sense that it applies also to non- geographic spaces such as the cosmos, the space of the human body, and the internal design of buildings.“ V publikaci (Craglia et al., 2003) v kapitole věnované SDI v souvislosti s tímto často používaným a svým způsobem standardním termínem kniha začíná používat označení „spatial data“. Důvodem jsou zřejmě kromě ustáleného používání ve spojení SDI také národní zvyklosti, publikace se věnuje popisu situace v několika evropských státech.

Jinou mezinárodně uznávanou organizací na poli GIT je Open Geospatial Consortium (OGC, www.opengeospatial.org). Toto konsorcium má sice přímo ve svém názvu přívlastek „geospatial“, ale aby se celá situace ještě znepráhlednila, ve svých standardech nejčastěji používá pro označení geografických dat termín „geographical features“.

V poslední době často zmiňovaná direktiva INSPIRE (Infrastructure for Spatial Information in the European Community; <http://inspire.jrc.ec.europa.eu/>) také pracuje s daty obsahujícími geografické nebo/a prostorové prvky. V normě INSPIRE se téměř výhradně setkáváme s výrazy „spatial information“ a „spatial data“. Předpona „geo-“ je používána víceméně pouze v souvislosti s geoportály. Stejný přívlastek (spatial) je používán také ve spojení s budováním jiných typů infrastruktury prostorových dat – např. GSDI (Global Spatial Data Infrastructure; <http://www.gsdi.org/>)¹¹.

Další komplikace přichází při pohledu na webové stránky organizace ICA (International Cartographic Association; <http://cartography.tuwien.ac.at/ica/>). Tato asociace, jejímž hlavním

8 Naopak organizace CAGI (Česká asociace pro geoinformace; www.cagi.cz), český člen EUROGI, používá ve svém názvu termín „geoinformace“.

9 Výrazy s přívlastkem „spatial“ se používají pouze pro atributy, referenční systémy, objekty apod.

10 Stránky projektu GINIE (<http://www.ec-gis.org/ginie/>) poskytují obrovské množství informací ohledně geografických dat.

11 Na stránkách GSDI se objevuje další dosud nezmiňovaný termín „(geo)spatial data“.

posláním je kartografická vizualizace prostorových dat, používá termín „geospatial data“, viz Commission on Geospatial Data Standards (<http://ncl.sbs.ohio-state.edu/ica/>). Zajímavé je, že vlastní stránky této komise obsahují poněkud pozměněný název – ICA Spatial Data Standards Commission. Přívlastek „spatial“ je používán i v dalších materiálech této komise.

Na závěr této kapitoly věnované zmatkům v definicích se pokusíme znovu nahlédnout do tří českých publikací, které se problematikou terminologie zabývaly v posledních létech.

Publikace Geoinformační terminologie pro geodety a kartografy (Šíma, 2003) a Terminologický slovník zeměměřičství a katastru nemovitostí (Pražák et al., 2005) shodně staví na roveň termíny geografická data, geodata, geoprostorová data, stejně tak i termíny geografická informace, geoinformace, geoprostorová informace. Kromě těchto definic se v obou textech vyskytuje také termín prostorová data:

- Geografická data, geodata, geoprostorová data
 - 1: data s implicitním nebo explicitním vztahem k místu na Zemi.
 - 2: počítačově zpracovatelná forma informace týkající se jevů přímo nebo nepřímo přidružených k místu na Zemi.
 - 3: data identifikující geografickou polohu a charakteristiky přírodních a antropogenních jevů a hranic mezi nimi.
- Geografická informace, geoinformace, geoprostorová informace
 - Informace týkající se jevů implicitně nebo explicitně přidružených k místu vztaženému k Zemi; znalost získaná jako výsledek syntézy, analýzy nebo integrace geografických dat.
- Prostorová data
 - Data o poloze, tvaru a vztazích mezi jevy reálného světa, vyjádřená zpravidla ve formě souřadnic a topologie.

Dokument s názvem Geoinformatika a geoinformační technologie (Rapant, 2006) nabízí definice termínu prostorová data¹² a dále v textu zmiňuje pojmy geodata a geografická data:

- Prostorová data
 - Prostorová data jsou jakákoliv data, která obsahují formální polohovou referenci, např. odkaz na buňku gridu. Jedná se např. o data DPZ nebo mapy.
 - Prostorová data (angl. spatial data) jsou data, která se vztahují k určitým místům v prostoru a pro která jsou na potřebné úrovni rozlišení známé polohy těchto míst.
- Geodata
 - Se stejným významem jako prostorová data se také používá termín geodata, což ovšem není považováno za zcela správný přístup, protože geodata jsou vztažena pouze na zemské těleso a jeho přilehlé okolí.
- Geografická data
 - Geografická data jsou data vztahující se k předmětu geografie.
 - Geografická data jsou data týkající se fenoménů přímo nebo nepřímo svázaných s místy vztahujícími se k povrchu Země.

12 Samotná publikace (Rapant, 2006) upřednostňuje ve vlastním textu druhou definici termínu „prostorová data“.

Shrneme-li předchozí definice a další informace, můžeme dospět k následujícím závěrům a doporučením.

- Je důležité si uvědomovat významové rozdíly mezi termíny data, informace, údaje a znalosti.
- Pokud hovoříme o datech vztažených k zemskému tělesu, přívlastky prostorový, geoprostorový, geografický a geo- můžeme považovat za synonyma.
- V případě dat popisujících kosmická tělesa je na místě výraz prostorová data.
- Různé přívlastky dat s geografickým aspektem by měly být používány s ohledem na různé kontexty (standarty, oficiální názvy apod.).
- Používané pojmy je vhodné na začátku textu vysvětlit a odkázat se na příslušné zdroje (legislativní či technologické standarty a normy, případně v praxi používaná, ale dosud nepsaná pravidla).
- Ačkoli (možná právě proto) se problematice věnovalo tolik stránek, ukazuje se, že pro vzájemnou interoperabilitu nejsou důležité definice, ale především popis dat a jejich významu, kterému se budou věnovat další kapitoly.

Tato kapitola sice nadala jednoznačnou odpověď na otázku ohledně správné a univerzálně platné definice, ale přesto snad přispěla k hledání cesty z labyrintu různých termínů.

V tomto dokumentu bude převážně používán termín „prostorová data“ a „prostorová informace“ (ve správných kontextech, nikoli jako synonyma). Jedním z důvodů je zaměření tohoto textu na interoperabilitu a tedy souvislost s budování Spatial Data Infrastructure (SDI), se kterou se právě výše uvedený pojí. Případné použití ostatních termínů bude vázáno na standarty, zvyklosti, odlišný kontext nebo konkrétní význam.

Vlastnosti prostorových dat

Prostorová data mají velké množství nejrůznějších vlastností, přičemž se nejedná pouze o technické parametry, na které se posuzování vhodnosti, přístupnosti nebo kvality dat často omezuje. Všechny vlastnosti prostorových dat by měly být zohledněny při volbě datových vrstev pro konkrétní projekt. Otázkou zůstává dominantních a marginálních vlastností, která by měla záviset na charakteru projektu, konkrétních výstupech, požadavcích uživatele apod. Pokud však nebudou všechny parametry prostorových dat zohledněny nebo jejich posouzení bude neobjektivní, mohlo by dojít (a v praxi často dochází) k neúmyslné nebo i úmyslné manipulaci s uživatelem.

Seznam vlastností prostorových dat popisovaný v tomto dokumentu vychází z publikací (Annoni et al., 2008), (Cada & Mildorf, 2005), (Cerba & al., 2009), (Directive 2007/2/EC) a (ISO 19115:2003). Vlastnosti jsou rozděleny do čtyř skupin – obecné vlastnosti, vlastnosti prostorové složky dat, vlastnosti časové (temporální) složky dat a vlastnosti atributové (popisné) složky dat. Velké množství těchto vlastností dat je součástí popisu dat – metadat nebo nejrůznějších typů datových modelů (schémata, UML modely, databázová schémata, ontologie apod.).

1. Obecné vlastnosti

- | | |
|--|----------------------------------|
| 1. Podpora interoperability a přístupnosti | 4. Soulad s platnou legislativou |
| 1. Multilingualita | 5. Vazby na jiné datové sady |
| 2. Kulturní adaptabilita | 6. Datové modely |
| 3. Metadata | 2. Původ dat |

1. Tvůrce dat
2. Charakter dat (primární /přímo měřená/ nebo sekundární /data získaná zpracováním/ data)
3. Metody použité pro získání dat
4. Správa dat
5. Frekvence aktualizací
6. Aktualizované informace
3. Distribuce dat
 1. Licence a autorská práva
2. Vlastnosti prostorové složky
 1. Použité jednotky
 2. Přesnost
 3. Podrobnost
 4. Konsistence
 5. Spolehlivost
3. Vlastnosti časové (temporální) složky
 1. Použité jednotky
 2. Přesnost
 3. Podrobnost
4. Vlastnosti atributové (popisné) složky
 1. Použité jednotky
 2. Přesnost
 3. Podrobnost
 4. Konsistence
 5. Spolehlivost
 6. Téma
2. Ceny
3. Poskytovatel a distributor dat
4. Prezentace dat
 1. Vizualizační model
 2. Vícenásobná reprezentace (včetně kontextové reprezentace)
5. Technické parametry ukládání a distribuce dat
 1. Medium
 2. Formát dat
6. Prostorový rozsah
7. Geometrie (prostorová reprezentace)
8. Topologie
9. Dimenze
10. Geodetické datum
4. Konsistence
5. Spolehlivost
6. Časový rozsah
7. Terminologie
8. Klasifikační systém
9. Správa identifikátorů
10. Registry a číselníky
11. Katalogy prvků (Feature catalogues)

Popis prostorových dat

Největším současným problémem při využívání (nejen) prostorových dat¹³ je jejich heterogenita (různorodost). Popis dat, včetně prostorových dat, v současnosti představuje téměř nutnou podmínku pro jejich plnohodnotné a efektivní využívání. Základním důsledkem popisu a dokumentace dat není odstranění všech heterogenit, ale eliminace jejich následků, kterými jsou především snížená přístupnost a použitelnost takových dat. Což bývá spojené s omezenými možnostmi sdílení, kombinace a výměny jednotlivých datových sad mezi jednotlivými uživateli, institucemi i systémy. Vzhledem k tomu, že zřejmě v žádném oboru nebo odvětví nikdy nebude fungovat jediný univerzální datový model nebo formát (jinými slovy data budou stále mít různé vlastnosti a charakteristiky), je právě eliminace následků různorodosti dat velice důležitou činností vedoucí k vyšší míře interoperability (schopnosti vzájemné spolupráce systémů¹⁴), přičemž interoperabilita a její podpora tvoří jeden ze základních pilířů budování SDI (Spatial Data Infrastructure). Popis dat hraje důležitou roli v procesu pochopení a porozumění také z pohledu člověka nejen z hlediska automatizovaného zpracování prostorových dat.

Bohužel existence heterogenit je symptomatická nejen pro prostorová data, ale také pro popis dat. Následující odstavce nabízí několik různých pohledů na problematiku popisu prostorových dat. Na závěr této kapitoly se pokusíme shrnout úrovně popisu dat a jejich hierarchii.

Příklad 1: Proces porozumění

Proces porozumění prostorovým datům a tedy i jejich popis můžeme rozdělit na tři základní fáze (úrovně, části)¹⁵ (modifikováno podle Šlapák, 2003), přičemž jsou všechny tři složky považovány za stejně důležité a nezbytné pro kvalitní práci s prostorovými daty:

1. Syntaktická relevance – subjekt, který data přijímá, musí je být schopen detekovat a rozumět jim.
2. Sémantická relevance – Subjekt musí vědět, co data znamenají.
3. Pragmatická relevance – Data musí mít pro daný subjekt nějaký význam nebo smysl (subjekt by je měl dokázat využít).

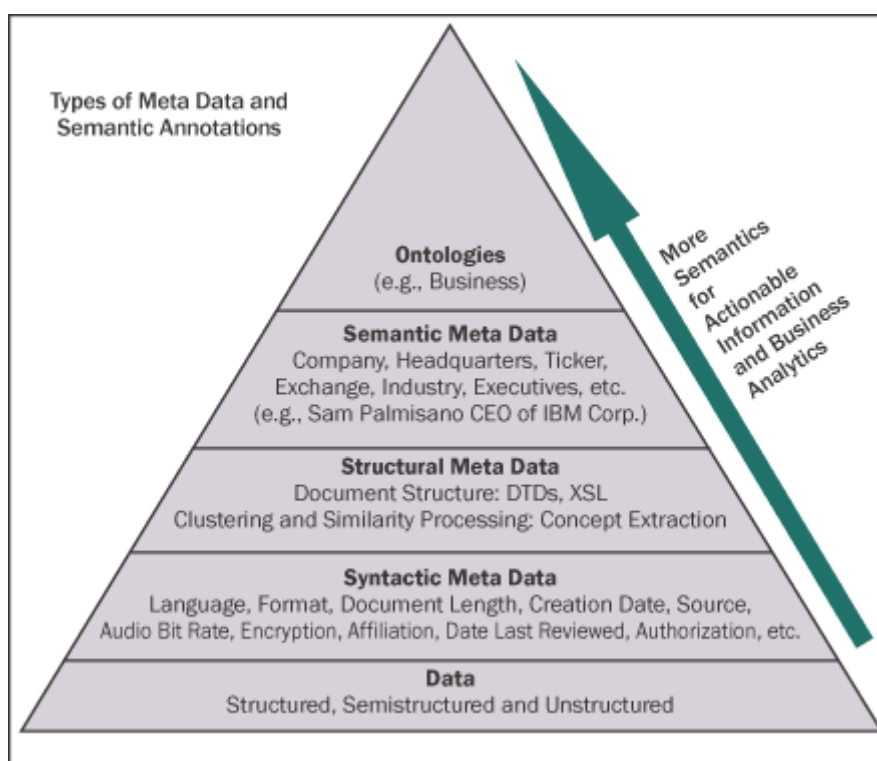
Z pohledu popisu a dokumentace dat je důležité nejen jeho množství, ale především kvalita takového popisu. Ta kromě jiného spočívá v přístupnosti popisu, která je přímo závislá na dodržování a podpora standardů a norem, včetně názvosloví.

13 V této pasáži (podobně i v dalších částech textu) jsou pod termínem data myšlena i data prostorová. Pokud by byly popisovány některé specifické vlastnosti prostorových dat, bude používán termín upřesněn.

14 Další definice termínu interoperabilita jsou k dispozici například v (Pražák et al., 2005).

15 Povšimněte si podobnosti s kartografií, konkrétně s kapitolami, které jsou věnované v různých publikacích jazyku mapy a mapovým znakům.

Příklad 2: Typy metadat a sémantického popisu



Ilustrace 1: Typy popisu dat (Sheth, 2003)

Předchozí ilustrace přibližuje různé typy a úrovně popisu dat:

1. Vlastní data mohou obsahovat některé základní informace o sobě samých. Používáme-li data zapsaná pouze pomocí textu (například značkovací jazyky), je možné tyto informace získat bez používání speciálního software. V případě binárních dat je nutné načtení takových struktur do systému, který je pro práci s danými daty přímo určený. S informacemi, které obsahují vlastní data dokáže většinou pracovat pouze člověk disponující specifickými znalostmi – jedná se o znalosti, které je možné získat na základě obsahu dat. Zpracování (jeho rychlost a jednoduchost) takových informací, jednoznačnost a pravdivost výsledků do značné míry ovlivňuje úroveň strukturovanosti dat.
2. Druhým stupněm popisu je dokumentace syntaktických pravidel. Pojem syntax (syntaxe) bude v tomto dokumentu chápán jako souhrn pravidel pro definování přípustné kombinace symbolů, které vyjadřují vlastní data. Informace související se syntaxí jsou nejčastěji k dispozici ve standardech příslušných formátů, ale syntaktická pravidla souvisí také například s použitým kódováním.

Příkladem jednoduché syntaxe je jazyk XML (Extensible Markup Language), kde jsou definovány prvky jazyka (elementy, atributy, procesní instrukce, entity apod.), jejich způsob zápisu a jeho omezení (spočívající například v používání omezené množiny znaků) a kombinace.
3. Nejtypičtějším zástupci třetí úrovně popisu dat, tzn. popisu struktury dat, jsou tzv. jazyky pro popis dokumentu¹⁶. Tyto jazyky jsou opět spojené se světem XML a ostatní značkovacích jazyků. Tzv. schémové jazyky popisují strukturu dokumentu – to znamená

¹⁶ Schémové jazyky, schema languages – nezaměňovat s konkrétním jazykem pro popis dokumentu W3C XML Schema.

způsoby kombinace konkrétních prvky a další omezení týkající se například datových typů, extrémních hodnot, kardinality apod.

4. Popis sémantiky dat je poměrně problematický. Velice často bývají do vlastních dat doplňovány sémantické informace například formou popisu nebo komentářů. Pokud zůstaneme v oblasti značkovacích jazyků můžeme sémantickou informaci vyjádřit pomocí vhodně zvoleného označení tagu (značku), který obklopuje vlastní data – uživatel pozná význam dat ohraničených tagem <adresa>, zatímco tag <I45WQ89b55UUn> je z pohledu významu neidentifikovatelný.
5. Ontologie představují nejvyšší formu popisu dat. Jedná se o strukturu schopnou popsat i vztahy mezi jednotlivými prvky nebo skupinami prvků dat. Jinými slovy slovy, zatímco ze struktury dat poznáme, že existuje vazba mezi prvky A a B (resp. lze vyčíst některé informace o vazbě, například kardinalitu nebo závislost), pomocí ontologií můžeme tuto vazbu specifikovat – je možné například definovat, že silnice vede nad prvkem železnice (více viz kapitola věnovaná ontologiím).

Příklad 3: Standardy pro popis dat

Pokud chceme popisovat dat a ptáme se „Jakým způsobem se mají správně data popisovat?“, existuje jediná správná odpověď – pomocí standardů. V oblasti standardů také není situace jednoduchá. Není možné říct, že pro jednu úroveň nebo formu popisu dat existuje jeden správný standard. Navíc se jednotlivé standardy vzájemně překrývají a i odborníci se často nedokáží shodnout, která z technologií je nejvýhodnější pro daný typ popisu dat.

Následující tabulka (vytvořená podle publikace Niemann, 2005) poskytuje základní přehled standardů a úrovní popisu dat ve formě datového modelu (všechny níže zmíněné modely a metody mohou být podrobněji popsány pomocí metadat, např. pomocí standardů ISO 19115, dalších dokumentů ze série ISO 191XX nebo standardem Dublin Core Metadata Initiative.).

Standard	Level of model
Controlled vocabularies & glossaries	Thesauruses
Relation models, mark-up languages	
ER model, DB Schema, XML schema languages	Taxonomies
Topic Maps	Conceptual models
RDF/S	
UML	
Ontologies	Logical theories
Description Logic	
First Order Logic	
Modal Logic	
Axiology	

Tabulka 1: Metody popisu dat

Shrnutí

Úkolem popisu dat je zajištění nebo alespoň podpora interoperability, tedy schopnosti vzájemné spolupráce systémů. Ve spolupráci systémů brání odlišný způsob zápisu dat, různá struktura dat a lišící se význam dat. Proto rozlišujeme syntaktickou, strukturální a sémantickou interoperabilitu. A také popis dat, resp. jejich vlastností, se zaměřuje na syntaxi dat, strukturu dat a sémantiku dat¹⁷.

Při práci s prostorovými daty se nemůžeme spoléhat pouze na inteligenci a představivost uživatel (ať se jedná o člověk nebo informační technologie). Je nutné poskytnout co možno největší množství relevantních informací, které umožní kvalitní, rychlou a maximálně zautomatizovanou interpretaci dat (porovnejte s termínem autorská slepota v kartografii).

¹⁷ V tomto případě vynecháváme nejabstraktnější vrstvu – pragmatiku dat. Důvodem je omezená (téměř neexistující) formalizace pragmatiky, kterou není možné zpracovávat pomocí automatizovaných procesů.

Metadata

Jak již bylo napsáno v úvodu, sémantická data mohou být popsány několika způsoby, například pomocí ontologií, UML modelů, ER modely, RDF, popisy ve formě textu, apod. Tato kapitola se zaměřuje jeden ze základních způsobů určení sémantiky prostorových dat. Důvodem pro vytvoření této samostatné kapitoly byla nejen jednoduchost používání metadat, ale i existence a úroveň implementace standardů a existence direktivy INSPIRE, které kromě jiného zavádí používání metadat popisujících geografická dat ve všech státech, které podléhají této direktivě.

Metadata¹⁸ jsou data popisující obsah, reprezentaci, rozsah (prostorový i časový), prostorový referenční systém, kvalitu a administrativní případně i obchodní aspekty využití digitálních dat. (Rapant, 2006)

Příkladem (často nevědomého) používání metadat v kartografii je tiráž, která patří mezi základní kompoziční prvky mapy. Tiráž mapy je definována jako soubor informací o různých aspektech tvorby a vlastnictví mapy (Voženílek, 2001). Mezi povinné prvky tiráže patří jméno autora nebo vydavatele mapy, místo a rok vydání (sestavení) mapy. Dále údaje uvedené v tiráži mohou specifikovat použité kartografické zobrazení, nakladatele a náklad, pořadí vydání, spolupracovníky podílející se na tvorbě mapy (redaktor, lektor), druh tisku, údaje o použitém papíru, copyright mapy a případná další práva spojená s používáním mapy, podkladové zdroje a další.

Metadatové záznamy poskytuje také řada webových stránek. Příkladem může být následující hlavička webové stránky, která obsahuje prvky popisující název stránek, jméno autora, stručný obsah stránek, klíčová slova, použité kódování a jazyk stránek:

```
<head>
<title>Mapy na webu</title>
<meta name="author" content="code: Jan Svoboda (Honza Svoboda);
jsvoboda@seznam.cz" />
<meta name="description" content="Server o kartografii na internetu" />
<meta name="keywords" content="mapy, kartografie, web, internet, www" />
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<meta http-equiv="content-language" content="cs" />
</head>
```

Jak je patrné z předchozích příkladů, jednotlivé prvky metadat mohou mít (a většinou také mají) stejný význam (např. označení autora), pouze jsou vyjadřovány pomocí jiných prostředků.

18 V podobném smyslu se používají také termíny metainformace a metaznalosti. (Rapant, 2006)

Sémantika

Aby bylo možné co nejefektivněji využívat data i aplikace, je nezbytné pochopit význam datových sad v maximální možné míře. K tomu slouží nejen studium oborů, které data produkují (např. geografie¹⁹, sociologie, demografie, ekologie, biologie apod.), ale také využívání významových (sémantických) informací. Používání sémantických informací výrazně urychlí proces poznávání dat a díky tomu bude možné zamezit nejrůznějším nesrovnalostem při interpretaci a zpracování dat (konflikty mezi datovými sadami, omyly při využívání vizualizačních prostředků apod.). Příkladem nepochopení významu a smyslu dat může být například zpracování dat kvalitativního charakteru pomocí kartografických vyjadřovacích prostředků umožňujících znázornění kvantity. Obecně je možné doplnění a používání sémantických informací chápat jako snížení nejistoty při jakémkoli nakládání s daty, včetně prostorových dat. Zároveň může docházet ke zvýšení kvality rozhodovacích procesů závislých na datech obsahujících sémantické informace.

V předchozím textu bylo již zmíněno slovo „sémantika“. Účelem této kapitoly je daný termín vysvětlit, ukázat je propojení s daty (i prostorovými daty), včetně příkladů, a uvést výhody a nevýhody plynoucí z používání sémantiky

Termín „sémantika“

Obecná definice sémantiky: „Sémantika je nauka o významu jazykových jednotek.“ (podle <http://slovník-cizich-slov.abz.cz>). Termín sémantika se do oblasti IT dostal zprostředkovaně přes humanitní vědy – lingvistiku a logiku. Z pohledu humanitního nebo filosofického se termínem „sémantika“ zabývá například článek (Peregrin, 2005), který odkazuje na jiné zdroje a cituje další definice a popisy tohoto výrazu:

- Sémantika sa sústreďuje najmä na skúmanie vzťahov medzi jazykovými výrazmi a predmetmi, na ktoré sa tieto výrazy vzťahujú a na tie vlastnosti a vzťahy výrazov, ktoré súvisia a ich vzťahmi k týmto predmetom.
- Logiku dělíma na syntax (zabývající se otázkou dokazatelnosti) a sémantiku (zkoumající pravdivost).

Také sémantika v prostředí IT je popisována v mnoha publikacích. Jedním z textů věnujících se problematice sémantiky data je publikace (Sheth, 2001)²⁰ – článek je koncipován jako výstup z panelu zabývajícího se danou problematikou, přičemž v textu jsou zřetelné jednotlivé názory diskutujících a jejich postupné sblížení²¹. Na základě článku (Sheth, 2001) je možné sémantiku chápat jako mapování objektu reálného světa a modelu reprezentujícího skutečný objekt, přičemž tento model může být reprezentován v informačním systému například jako objekt nebo položka v databázi. Toto mapování zajišťuje sémantiku modelu prostřednictvím popisu nebo identifikace významu prvků. V publikaci (Sheth, 2001) je také sémantika definována a popsána pomocí následujícího schématu.

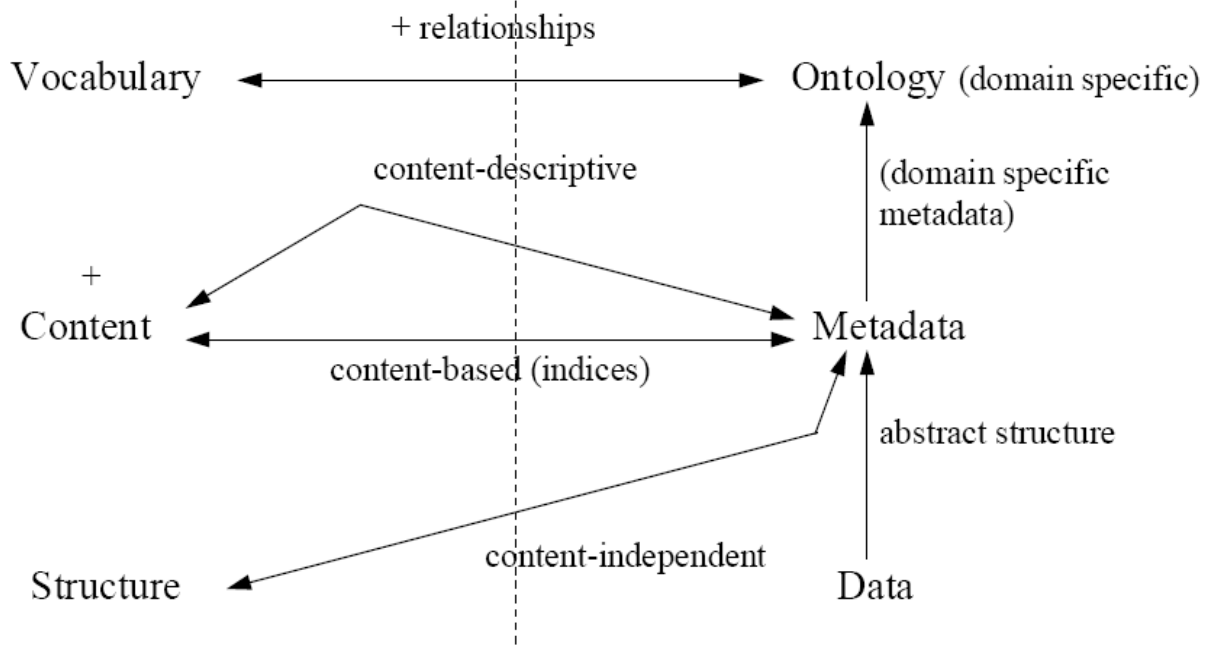
19 Krátký úvod do socioekonomické geografie a demografie je součástí předmětu Socioekonomická geografie pro geomatiku (KMA/SGG).

20 Studium problematiky vztahu sémantiky a dat je mnohem starší. První modely se podle (Mylopoulos, 2004) začaly objevovat v roce 1974.

21 Účelem tohoto textu není podrobný popis vývoje pojetí sémantiky, proto nebo přesto autor doporučuje publikaci Sheth, 2001 podrobně prostudovat.

What is semantics?

Where is semantics?



Ilustrace 2: Sémantika, podle (Sheth, 2001)

Obrázek ukazuje tři typy nebo úrovně sémantiky – strukturu dat, obsah dat a tzv. slovníky (ve smyslu striktně definované terminologie umožňující vzájemnou komunikaci mezi lidmi a systémy). S těmito vrstvami sémantiky jsou propojeny typy popisu dat, které mohou dané sémantické informace obsahovat.

Následující seznam obsahuje některé částečně interpretované myšlenky z výše uvedeného článku (Sheth, 2001), které mohou termín „sémantika“ více přiblížit.

- Sémantika (v nějaké formě) existuje stále, ale její používání je vázáno na konkrétní implementaci (systém musí chápat konkrétní význam).
- Sémantika pokrývá (determinuje) mnoho různých oblastí – interpretaci a využití dat a informací obecně i komunikaci, a to vše nejen mezi systémy, ale i mezi člověkem a systémem i mezi lidmi navzájem.
- Sémantika je závislá na člověka (který ji definuje), stroj pouze slouží k jejímu používání.
- Sémantika (ve smyslu sémantika dat) může být odvozena ze struktury dat, znalostních databází, jazyků pro popis dokumentu (schémat), programů, komunikačních protokolů apod.
- Sémantika dat může být čerpána také z popisů datových sad, právních dokumentů (zákony, vyhlášky, licence apod.) nebo norem a standardů.

Sémantiku v prostředí IT nepředstavuje monolitní strukturu, proto je možné ji dále členit a klasifikovat. Následující odstavce vychází především z dokumentů (Sklenák, 2004) a (Borgida, 2005) a ukazují jedno z možných dělení sémantiky:

- Implicitní sémantika

Implicitní sémantika vyplývá ze sdíleného porozumění, které je založené na konsensu lidí. Jinými slovy data mají tak jasný význam, že tento význam je zcela zřetelný bez nutnosti vysvětlování nebo dalšího popisu. Výhodou tohoto typu je rychlost zpracování informace. Dále není nutné přidávat k vlastní informaci přidávat další data, která zvyšují její celkový objem. Na druhou stranu je potřeba si uvědomit riziko mnohoznačnosti, kdy se uživatelé neshodnou na významu konkrétního pojmu. Problematické je také strojové zpracování dané informace.

- Explicitní sémantika

Explicitní sémantika je opakem sémantiky implicitní. V tomto případě rozlišujeme neformální a formální explicitní sémantiku.

- Neformální explicitní sémantika

Příkladem tohoto typu sémantiky může být například slovník nebo textový dokument. Opět se ukazuje, že je velice obtížné zpracování takové sémantiky pomocí automatizovaných procesů. Ale i na webu můžeme najít příklady neformální explicitní sémantiky – informací přidaných k původním datům, jejichž význam není zcela pevně stanoven pomocí formalizovaného zápisu. Příkladem mohou být například jazyky HTML (HyperText Markup Language) nebo UML (Unified Modeling Language) obsahující elementy, které mohou (ale při nesprávném používání nemusí) deklarovat význam daného elementu.

V případě HTML dávají laičtí tvůrci webových stránek často přednost vizuální stránce, před významovou. Jinými slovy například element `<h1>` vkládají do kódu nikoli proto, že znamená nadpis první úrovně (čili webová stránka by měla obsahovat pouze jeden takový element), ale proto, že je text vypsán velkým písmem (vizualizační vlastnosti by neměly záviset na použitém elementu, ale na stylovém jazyce).

- Formální explicitní sémantika

Formální explicitní sémantika přidává k předešlé variantě ještě formální (nebo formalizovaný) zápis sémantické informace. Opět je možné rozdělit formální explicitní sémantiku do dvou kategorií.

- Formální explicitní sémantika pro zpracování člověkem

Tento typ sémantiky má nejčastěji podobu formální dokumentace nebo formální specifikace významu.

- Formální explicitní sémantika pro strojové zpracování

Základním cílem této varianty sémantiky je strojově zpracovatelný obsah – jinými slovy, automatický systém dokáže analyzovat význam obsahu a stanovit další postup zpracování. Pro zajištění automatické rozpoznávání významu existují dvě základní metody:

1. Procedurální přístup – systém provede konkrétní operace, pokud narazí na určitý příznak v datech.
2. Deklarativní přístup – sémantika je zachycena formální deklarací (typický přístup pro většinu prostorových dat).

Následující modelová situace ukazuje důležitost a význam sémantiky nebo naopak zmatky spojené s její absencí. Úkol, který má být splněn pomocí GIT technologií a prostorových dat zní: „Uved’

souřadnice místa, kde se nachází Praha.“

Na tuto zdánlivě jednoduchou otázku můžeme získat odpověď například pomocí encyklopedie Wikipedia (www.wikipedia.org). Výsledkem jsou souřadnice 50° 05' s.z.š. a 14° 25' v.z.d. Tazatel však velice snadno může označit odpověď za chybnou, a to hned z několika důvodů.

V zadání není specifikován pojem souřadnice. Jaké souřadnice mají být použity? Jaký souřadnicový systém a jaká referenční plocha? Jaká forma zápisu souřadnic – má se psát západní zeměpisná délka (případně v jakém jazyce), zkratka (případně jaká), nebo se západní délkové souřadnice uvádí se znaménkem mínus? Podobným způsobem je možno přistupovat i k zápisu zeměpisné šířky. Dále je nutné řešit, jak detailní má být uvedení polohy – na kolik desetinných míst se budou souřadnice uvádět?

To je ovšem pouze polovina problému. Podobně se můžeme ptát na termín „Praha“. Každému se určitě nejdříve vybaví hlavní město České republiky. Ale na Slovensku také leží obec Praha, stejně jako v USA ve státě Texas. Hora Praha (862 m n.m.) leží ve středních Brdech (druhá nejvyšší hora celého pohorí). Název Praha se používá také pro letiště, nádraží nebo rozhlasovou stanici. Loď Praha pluje po Brněnské přehradě. České aerolinie vlastní letadlo Praha (OK-WAA). Hotel Praha se kromě samotného města Praha nachází například také v Krnově, Broumově, Špindlerově Mlýně nebo Hřensku. Podobně bychom mohli najít také restaurace se jménem Praha (například v Harrachově, Janově nad Nisou nebo Dolních Dunajovicích; restaurace Praha také reprezentovala tehdejší Československo na světové výstavě EXPO 58 v Bruselu) a jiné subjekty.

Na závěr je třeba poznamenat, že téměř žádné pojetí „Prahy“ rozhodně reálně nepředstavuje bod (snad s výjimkou vrcholku výše uvedené hory), proto je nutné stanovit místo v/na Praze, ke kterému bude souřadnice vztažena.²²

Pokud bychom nalezení příslušného objektu doplnili jeho vizualizací pomocí kartografických prostředků, dostaneme se také do situace, kdy význam dat bude implikovat jejich kartografickou reprezentaci. Kartografický znak vytvořený pro objekt z výše uvedeného příkladu může mít

- figurální (bodovou) lokalizaci (např. hora, sídlo, loď, letadlo),
- liniovou (čárovou) lokalizaci (např. trasa letadla nebo lodi),
- areálovou (plošnou) reprezentaci (např. sídlo).

Je tedy jasné, že kontrola správnosti odpovědi na úvodní otázku je nutně spojena se znalostmi významu a dalších vlastností vstupních i výstupních dat, tzn. s popisem dat a sémantikou. Správná (rozumějte přehledná, jednoznačná a jasná) forma zápisu a používání dat je důležitá nejen pro člověka, ale především automatické systémy zpracovávající data.

Je potřeba si uvědomit, že sémantika dat není žádnou samospasitelnou technologií, která zajistí stoprocentní porozumění, komunikaci a pochopení různých datových sad. Skeptický pohled publikovaný v (Burkett, 2002) tvrdí, že jednak nemohou existovat unikátně pojmenované datové elementy, a navíc případné unikátní označení stejně nezaručí jednotné pochopení významu dat. Přesto by měla sémantika a její využívání podstatně zlepšit zpracování datových sad. Podobně „negativní“, ale na druhou stranu realistický pohled předkládá již citovaný článek (Peregrin, 2005) – jedná se o dvě tvrzení, která zdůrazňují omezenost pouhého formalizovaného zápisu pro vyjádření významu v oblasti IT:

1. Počítač může mít pouze syntax ('formální pravidla') .
2. Syntax nestačí na sémantiku (to jest zavedením 'formálních pravidel' nikdy nemůžeme udělit výrazům skutečné významy).²³

22 Jiné příklady „sémantických nedorozumění“ uvádí publikace (Sheth, 2001) nebo (Madnick, 2006).

23 Autor doufá, že oba výše uvedené výroky pouze doplňují mozaiku informací o sémantice a neodradí čtenáře od

Také je potřeba si uvědomit, že používání informací a dat v jednom kontextu (ve spojení s jedním typem popisu sémantiky) často limituje kontexty ostatní.

Sémantika a prostorová data

Úvodem pasáže věnované vztahu sémantiky a prostorových dat by bylo vhodné připomenout existenci výrazu sémantika v jedné z významných geovědních disciplín. V rámci kartografie je vyčleňována věda popisující teorii kartografických znaků a jejich používání – kartografická sémiologie, jejímž zakladatelem je francouzský kartograf J. Bertin. Kartografická sémiologie se dělí do dalších disciplín, které zkoumají následující oblasti:

- Syntaktika – vztahy znaků k sobě navzájem.
- Sémantika – vztahy znaků k obsahu toho, co označují.
- Sygmatika – vztahy znaků k funkci objektů, které určují.
- Pragmatika²⁴ – vztahy uživatelů znaků ke znakovým soustavám. (Kaňok, 1999)

Vlastní prostorová data představují model reálného světa. Jako v případě každého modelu i v případě modelu Země nebo jiných kosmických těles, případně jejich částí, je zapotřebí dopustit se určitých zjednodušení na úkor zdůraznění takových parametrů modelu, které jsou důležité pro představu a zpracování jevu nebo procesu zkoumaného daným modelem. Formu zjednodušení představuje i způsob nazírání reálného světa. Model OGIS (Open GIS) geodata (Gardels, 1996) pracuje s dvojí reprezentací reálného světa:

1. Izolované diskrétní prvky s konkrétními vlastnostmi jsou reprezentovány jako tzv. features.
2. Spojité povrchy popisující komplexní geografický jev (fenomén) jsou nazývány jako tzv. coverages.

Publikace (Rapant, 2006) oproti tomu nabízí tři základní pohledy:

1. Objektový náhled – reálný svět jako množina diskrétních objektů rozložených v prostoru.
2. Jevový náhled – reálný svět jako množina jevů popsanych rozložením hodnot vlastností v prostoru, případně v čase.
3. Procesní náhled – reálný svět jako množina procesů, které v něm probíhají a které ho zároveň utváří.

Aplikaci sémantiky v geografických dat ukazuje například konceptuální datový model OGIS geodata (Gardels, 1996). Tento model pracuje se třemi hlavními komponentami:

1. Prostorová komponenta (typ geometrie a časoprostorová lokalizace)
2. Sémantická komponenta (význam jednotlivých prvků)
3. Metadata (informace nutné ke korektní interpretaci dat)

častějšího využívání sémantických informací.

24 Syntax, sémantika a pragmatika bývají označovány jako tři komunikační dimenze.

Ontologie

Ontologie tvoří jednu z komponent sémantického webu, tedy tzv. druhé generace webu, která bude zaměřena na dosažení vyšší úrovně automatizovaného vyhledávání informací. Termín ontologie je však známý již z antické filosofie, kdy označoval nauku o bytí²⁵. Do informačních technologií ontologie pronikají v souvislosti s pokusy o co možno nejvěrnější popis reálného světa pomocí datových struktur. Podle T. Grubera je ontologie explicitní specifikace konceptualizace. Tato definice byla rozšířena W. Borstem, který ontologii považuje za formální specifikace sdílené konceptualizace (Svátek, 2002). Obě definice jsou díky použití neobvyklých pojmů dosti těžkopádné. Co vlastně jednotlivé výrazy v obou definicích znamenají?²⁶

- Explicitní – výslovný, přímý, jasný, zřetelný; otevřeně, přímo vyjádřený.
- Formální – vyhovující požadavkům, předpisům; zdůrazňující formu.
- Specifikace – přesný výčet, podrobnější rozepsání položek, zpřesnění, upřesnění.
- Sdílený – určený pro veřejné užívání.
- Konceptualizace – systém pojmů.

Jinými slovy můžeme ontologii definovat jako jasný, zřetelný a přímo vyjádřený přesný výčet pojmů daného systému. Jak již bylo uvedeno v jedné z počátečních kapitol tohoto textu, ontologie představuje nejvyšší stupeň popisu dat. Podobně jako ostatní prostředky pro popis dat i ontologie zabraňují více-smyslnosti a více-významovosti. Důvodem vytvoření a implementace ontologií je právě snížení této více-smyslnosti a umožnění sdílení a znovu používání existujících znalostních struktur, včetně popisu datových sad.

Z pohledu procesu modelování dat se ontologie řadí na úroveň konceptuálního modelu, což znamená, že jsou aplikačně nezávislé. Díky své koncepci lze ontologie považovat za nástroj, který umožňuje práci nejen pouze s daty, ale také s informacemi a se znalostmi. Zároveň je potřeba uvést, že ontologie představují pouze jednu z cest směřujících k efektivnějšímu a inteligentnějšímu popisu datových struktur. Existují případy (z pohledu dat, modelovaných procesů, nasazených systémů, časové náročnosti, komplexity apod.), kdy jejich nasazení je nevhodné nebo nedostatečně efektivní.

Z pohledu klasifikace rozeznáváme několik typů členění ontologií (podle Svátek, 2002):

1. Členění podle „historických paradigmat“

1. Terminologické či lexikální ontologie lze ztotožnit s pokročilými tezaury, používanými v knihovnictví a dalších oborech orientovaných na textové zdroje. Jejich charakteristickým rysem je ústřední role termínů, které již nejsou dále (formálně) definovány. Používané relace mají z velké části taxonomický charakter (vymezení vztahu obecnějšího a speciálnějšího termínu), vedle toho bývá vyjádřena synonymie, meronymie (vztah termínů označujících celek a jeho část) a další relace obecného charakteru. Nejznámější terminologická ontologie je nepochybně lingvisticky zaměřený WordNet (<http://wordnet.princeton.edu/>); z něj byl odvozen např. Sensus (<http://www.isi.edu/natural-language/projects/ONTOLOGIES.html>) nebo vícejazyčná varianta EuroWordNet (<http://www.illc.uva.nl/EuroWordNet/>).

Podle (Benda, 2005) patří tento typ (označovaný jako taxonomie) a typ číslo tři

25 Ontologiemi se kromě jiných zabývali také významní antičtí filosofové Sokrates nebo Aristoteles.

26 Význam jednotlivých slov (kromě konceptualizace) byl vyhledáván na stránkách ABZ.cz: slovník cizích slov (<http://slovník-cizich-slov.abz.cz>).

(množina odvozovacích pravidel) mezi nejtypičtější ontologie.

2. Informační ontologie jsou rozvinutím databázových konceptuálních schémat. Hrají roli nadstavby nad primárními (strukturovanými, např. relačně-databázovými) zdroji, pro které zabezpečují jednak konceptuální abstrakci potřebnou pro pojmové dotazování, jednak vyšší úroveň kontroly integrity než běžné nástroje.
3. Znalostní ontologie navazují na výzkum v oblasti reprezentace znalostí v rámci umělé inteligence. Ontologie jsou zde chápány důsledně jako logické teorie, a jejich vazba na reálné objekty (instance) je oproti informačním ontologiím relativně volná. Třídy (koncepty) a relace jsou systematicky definovány prostřednictvím formálního jazyka.

2. Členění podle míry formalizace

Přestože je formalizace, jak již bylo řečeno, do jisté míry definiční vlastností ontologií, smysluplné využití mají i „ontologie“ zcela neformální či „semi-formální“. Jde zpravidla o glosáře, v nichž jsou jednotlivé pojmy vysvětleny přirozeným jazykem (volnou či strukturovanou formou). Ontologie vyjádřené ve formálních jazycích pak lze dále rozlišovat podle formálně-logických vlastností daného jazyka, jako je úplnost a rozhodnutelnost; tyto vlastnosti vycházejí z vlastností logického kalkulu, na kterém je jazyk založen, např. deskripční logiky. Většina formálních ontologií v sobě ovšem svým způsobem zahrnují i ontologii neformální. Jednotlivé konstrukty bývají totiž vybaveny dokumentačních položkou, umožňující vyjádřit obsah přirozeným jazykem.

3. Členění podle předmětu formalizace

1. Doménové ontologie jsou typem daleko nejfrekventovanějším. Jejich předmětem je vždy určitá specifická věcná oblast, vymezená širěji (např. celá problematika medicíny nebo fungování firmy) či úžeji (problematika určité choroby, poskytování úvěru apod.). Příklady doménových ontologií se širokým záběrem jsou Enterprise Ontology (<http://www.aiai.ed.ac.uk/~bat/>) nebo Beer Ontology (<http://www.purl.org/net/ontology/beer.owl>).
2. Generické ontologie usilují o zachycení obecných zákonitostí, které platí napříč věcnými oblastmi, např. problematiky času, vzájemné pozice objektů (topologie), skladby objektů z částí (mereologie) apod. Někdy se ještě výslovně vyčleňují tzv. ontologie vyšší úrovně („upper-level“), které usilují o zachycení nejobecnějších pojmů a vztahů, jako základu taxonomické struktury každé další (např. doménové) ontologie; nejnovějším výsledkem tohoto směru je SUMO – Standard Upper Merged Ontology (<http://suo.ieee.org/SUO/SUMO/index.html>). Ontologie typu common-sense („přirozeného rozumu“) mohou naopak obsahovat řadu velmi specifických, avšak relativně doménově-nezávislých znalostí, které lidé používají v každodenním životě. Nejznámějším příkladem je Cyc (<http://www.cyc.com/>).
3. Jako úlohové ontologie jsou někdy označovány generické modely znalostních úloh a metod jejich řešení. Na rozdíl od ostatních ontologií, které zachycují znalosti o světě („tak, jak je“), se zaměřují na procesy odvozování. Mezi úlohy tradičně zachycené pomocí takových znalostních modelů patří např. diagnostika, zhodnocení („assessment“), konfigurace, nebo plánování.
4. Aplikační ontologie jsou nejspecifičtější. Jedná se o konglomerát modelů převzatých a adaptovaných pro konkrétní aplikaci, zahrnující zpravidla doménovou i úlohovou část (a tím automaticky i generickou část).

Ontologie se skládají z následujících prvků (zpracováno podle Svátek, 2002):

- Třídy

Třídy neboli koncepty, typy, kategorie, či rámce představují základní stavební kámen každé ontologie. Třída má mnoho společných vlastností se stejnojmenným prvkem z oblasti objektově orientovaných jazyků, například dědičnost, ale v jiných se liší (např. nezahrnuje procedurální metody). Rozlišují se třídy definované (určené podmínky nutnosti i postačitelnosti) a primitivní (pouze nutné podmínky nebo zcela bez definovaných podmínek). V ontologiích tvoří souhrn tříd většinou uspořádanou stromovou hierarchii, která se nazývá taxonomie. Tato stromová struktura však bývá často reliktem z jiných typů způsobů modelování dat, protože ontologie disponují širší škálou možností, například vícenásobnou dědičností.

- Vlastnosti

Termín vlastnost disponuje největším počtem synonym. V popisu ontologií se můžeme setkat s pojmy jako relace, typ relace, funkce, sloty (většinou se vztahuje pouze na binární relace), role, „semantic link type“ nebo atributy. Navíc ještě často dochází k záměně se stejnojmenným pojmem například z oblasti značkovacích jazyků. V ontologiích vlastnosti nepředstavují podrobnější popis dat, ale definuje vztahy mezi jednotlivými třídami a individui. Vlastnosti (navzdory svému názvu) představují plnohodnotný prvek ontologie, který není závislý na konkrétní třídě (jako například v objektově orientovaných jazycích). Dalším specifickým ontologií je možnost tvorby hierarchie vlastností.

- Individua

Individua, individuály, konstanty, objekty nebo instance reprezentují reálné objekty, které již v rámci ontologie nebudou dále členěny. Některé ontologie tento prvek nevyužívají, resp. jej nahrazují třídou – neboť ontologie by měly představovat koncepty a nikoli reálné jevy nebo objekty. Na druhou stranu používání individuí přibližuje ontologie jiným formám a úrovním popisu dat (například databázovým schématům nebo UML). Navíc volba, zda má být daný prvek třídou nebo individuem často závisí na subjektivním pohledu a charakteru ontologie.

- Axiomy

Axiomy neboli pravidla představují další typ logických formulací. Axiomy umožňují vyjádřit například ekvivalenci tříd nebo vlastností, subsumpci (podřazení konkrétnějšího pojmu obecnějšímu termínu) tříd nebo vlastností, disjunkčnost tříd, rozklad třídy na podtřídy apod. V některých jazycích jsou axiomy definovány jako součásti tříd nebo vlastností (např. v Ontolingua nebo DAML + OIL), v jiných jazycích stojí zcela samostatně.

- Facety

Facety neboli meta-sloty či omezení na sloty představují vlastnosti relací (vlastností). Příkladem může být již zmíněná hierarchie vlastností. Dalšími možnostmi je definování symetrie, inverzní vlastnosti, tranzitivitu nebo funkčnost. Do této skupiny se řadí také stanovení definičního oboru (domain) a oboru hodnot (range), které se vymezují pomocí tříd nebo individuí.

- Primitivní hodnoty a datové typy

Obor hodnot dato-typového slotu bývá vymezen některým základním datovým typem (string, integer, float...), číselným nebo alfanumerickým intervalem, nebo výčtem. Někdy se primitivní hodnoty označují i jako dato-typové instance a samotné datové typy pak jako dato-typové třídy, např. číslo 4 pak bude instancí datového typu integer.

- Souhrnné informace

Souhrnné informace je možné rozdělit do dvou skupin. Jednu část představují metadata ontologie, údaje o autorovi, verzi ontologie, dokumentaci apod. Kromě těchto popisných údajů mohou souhrnné informace obsahovat importované a vložené externí ontologie.

Jak zapsat ontologie

Aby ontologie byly skutečně funkční, je zapotřebí vytvořit formalizovaný systém pro jejich zápis, zpracování a sdílení. Podobně jako v jiných částech informačních technologií (např. zápis prostorových dat, jazyky pro popis dokumentu apod.) i zde existuje a v minulosti existovalo velké množství nejrůznějších jazyků a formátů určených pro tvorbu ontologických systémů. V této části uvádíme stručný přehled těchto jazyků, přičemž dvěma nevýznamnějším jazykům OWL a RDF jsou věnovány samostatné pasáže. Následující abecední seznam vychází především z publikace (Svátek, 2002), kde se zájemce může setkat s detailnějším popisem jednotlivých jazyků a formátů.

- CyC (www.cyc.com, www.opencyc.org²⁷)

Název jednoho z nejstarších pokusů o ontologický systém je odvozený ze slova „enCYClopedia“. Z hlediska kódování a zápisu je používán specifický jazyk CyCL vycházející z jazyka LISP. Cyc se skládá z tzv. mikroteorií, které představují vzájemně provázaná tvrzení.

- DARPA²⁸ Agent Mark-up Language (DAML)
Úkolem projektu DAML (www.daml.org) bylo na základě syntézy výsledků dosavadního výzkumu v oblasti ontologií vytvoření sémantického jazyka pro RDF, který nabízel větší možnosti než jazyka RDFS. Výsledkem byl jazyk DAML-ONT.
- DAML+OIL

Jak vyplývá ze zkratky tento jazyk vznikl propojením výše uvedených formátů DAML-ONT a OIL. Základem jazyka DAML+OIL jsou třídy reprezentované buď svým jménem (pojmenované třídy) nebo určitým logickým výrazem (anonymní třídy). Pro tvorbu logických výrazů vymezujících třídy se používají konstruktory (intersectionOf, unionOf, complementOf, oneOf, toClass, hasClass, hasValue, minCardinalityQ, maxCardinalityQ, CardinalityQ), které je možné libovolně kombinovat. Vlastní ontologie je tvořena axiomy vybudované nad definovanými třídami. Existuje celkem 11 typů axiomů – subClassOf, sameClassAs, disjointWith, subPropertyOf, samePropertyAs, inverseOf, transitiveProperty, uniqueProperty, unambiguousProperty, sameIndividualAs, differentIndividualFrom.

- eXtensible Ontology Language (XOL²⁹)
Základem jazyka XOL (<http://www.ai.sri.com/pkarp/xol/>) se stala odlehčená verze jazyka OKBC, tzv. OKBC-Lite. Hlavní změnou oproti výše uvedeným ontologickým jazykům je propojení ontologií a značkovacích jazyků, konkrétně XML. XOL umožňuje definovat strukturu a hierarchii tříd, včetně pouze binárních vazeb.
- Ontobroker
Jazyk Ontobroker (<http://ontobroker.semanticweb.org/>) je založen na stejném principu jako jazyk SHOE – to znamená, že jde o doplňování sémantických informací do webových

27 Pouze uvolněná část Cyc.

28 Defense Advanced Research Projects Agency – organizace, která zajišťovala vývoj jazyka DAML.

29 Pod zkratkou XOL se v oblasti ontologií vyskytují ještě dva termíny -- Ontology Exchange Language a XML Based Ontology Language. Přičemž ve většině publikací se objevuje v souvislosti se zkratkou XOL především první uvedený termín, včetně uvedených webových stránek.

stránek. Odlišná je architektura jazyka, který je na rozdíl od předchozího příkladu (SHOE) centralizovaná – je založena na existenci ontologické serveru. Server poskytuje informace účastníkům, kteří k němu mají přístup a zároveň umožňuje sběr informací z anotovaných stránek. Proto Ontobroker nemusí na rozdíl od jazyka SHOE definovat vlastní sadu elementů – jedná se pouze o obohacení HTML o atributy.

- Ontolingua

Jazyk Ontolingua (<http://www.ksl.stanford.edu/software/ontolingua/>) se od výše uvedeného formátu Cyc liší svým věkem (je zhruba o šest let mladší) a svou otevřeností. Na rozdíl od převážně komerčně zaměřeného Cyc je Ontolingua otevřená, protože jejím prvotním úkolem bylo propojení, výměně a sdílení ontologií v rámci jednotlivých odborných komunit. Vlastní jazyk Ontolingua byl vytvořen jako nadstavba jazyka KIF (viz níže) a využívá podobně jako Cyc syntaxi jazyka LISP. Ontolingua umožňuje pomocí vymežujících podmínek (axiomů) definovat třídy, relace a funkce.

- Ontology Inference Layer (OIL)

Jazyk OIL (<http://www.ontoknowledge.org/oil/>) představoval zcela novou možnost popisu podmínek v ontologiích. Pro tento účel byla použita deskripční logika (omezená predikátová logika).

- Open Knowledge Base Connectivity (OKBC)

OKBC (<http://www.ai.sri.com/~okbc/>) mělo počátkem devadesátých let dvacátého století představovat ekvivalent k Open Database Connectivity (ODBC) – to znamená, že OKBC má představovat univerzální rozhraní pro přístup ke znalostním systémům, které by umožňovalo komunikaci mezi takovými systémy. Základní prvky, jejichž předávání mezi aplikacemi OKBC řeší, jsou převzaty převážně z jazyka Ontolingua.

- Operational Conceptual Modelling Language (OCML)

Jazyk OCML (<http://technologies.kmi.open.ac.uk/ocml/>) vznikl jako reakce na omezené možnosti jazyka Ontolingua pro přímý vývoj konkrétních programových aplikací (tak, aby se modely nemusely konvertovat do jiného jazyka). Prakticky se jedná o rozšíření deklarativní části jazyka Ontolingua o prvky převzaté z procedurálních jazyků a expertních systémů (cykly, podmínky apod.) – často se používá v souvislosti OCML termín „operacionální Ontolingua“. Proto OCML zápis libovolné aplikace, nikoli pouze ontologie – OCML lze označit jako plnohodnotný programovací jazyk

- Simple HTML Ontology Extension (SHOE)

SHOE (<http://www.cs.umd.edu/projects/plus/SHOE/>) představuje prvním jazyk určený pro doplnění sémantiky (metadata o objektech, ontologie, definující sémantiku těchto metadat) k webovým stránkám. Jazyk SHOE je velice jednoduchý – zachycuje pouze třídy a relace bez dalšího odlišení.

Ontologie – příklady a ukázky

Ontologie se velice často skrývají za produkty, které jsou běžně využívány návštěvníky webových stránek. Jedná se například o katalogy internetových obchodů, kategorie ve vyhledávacích nebo metadata na webových stránkách. Literatura pojednávající o ontologiích poskytuje poměrně rozsáhlý seznam příkladů využití ontologií v různých oblastech lidské činnosti. Příklady jsou k dispozici například v publikaci (Svátek, 2002).

Následující příklad ukazuje velice možné využití ontologií v kartografii, případně obecně v geoinformace. Úkolem je pomocí ontologie sestavit jednoduchý autoatlas, který bude obsahovat

pouze základní prvky – silniční komunikace, železniční tratě a vodní toky, dále budou na silničních komunikacích označena riziková místa, tj. body, kde dochází ke křížení výše uvedených liniových prvků. K těmto rizikovým místům budou připojena ještě varování pro řidiče.

Pro tvorbu této konkrétní ontologie je používán software Protégé verze 3.4 (<http://protege.stanford.edu/>) a jazyk OWL (Web Ontology Language), proto i používaná terminologie je závislá na těchto prostředcích.

1. V prvním kroku vytvoříme základní datový model liniových prvků – třída (Class) `Liniove_prvky`, která obsahuje třídy `Silnice` (dále dělenou na podtřídy `Dalnice`, `I_trida`, `II_trida`, `III_trida`), `Zeleznice` (podtřídy `Koridor`, `Elektrifikova`, `Neelektrifikovana`) a `Vodni_tok`.

Tento model ještě neukazuje možnosti ontologií, protože stejné schéma obsahují pouze vnořené třídy je možné vytvořit podle jakéhokoli modelovacího jazyka, např. UML, E-R modelu nebo například pomocí jazyka pro popis dokumentu. Síla ontologií spočívá především v modelování vztahů.

Přesto je možné již v tomto úvodním kroku nastavit tzv. disjoints – tzv. vylučující třídy. Jinými slovy můžeme zapsat – jestliže je liniový prvek železnici, pak nemůže být silnicí. Podobně lze nastavit vylučující třídy pro jednotlivé typy silnic a železnic – tomto případě je potřeba dát si pozor na fakt, že elektrifikovaná železnice může být rychlostním koridorem, ale rozhodně nemůže být neelektrifikovanou železnici (podobně i v dalších případech).

2. V dalších krocích se zaměříme na odvozování tříd pomocí restrikcí. K tomu je nutné nejdříve definovat vlastnosti (Property). Pro naše účely vytvoříme vlastnosti `Lezi_na`, `Protina`, `Vede_nad`, `Vede_pod`. Tyto vlastnosti budou deklarovány v doméně `Liniove_prvky`.

V rámci definování vlastností je možné také inverzní vlastnosti – `Lezi_na` zatím svoji vlastnost nemá, vlastnost `Protina` je sama sobě inverzní a `Vede_nad` je inverzní vlastností k `Vede_pod` (a naopak).

3. Třetí krok znamená vytvoření třídy `Rizikove_body`. V tomto bodě konečně použijeme restrikcce, kdy jejich prostřednictvím popíšeme fakt, že veškeré rizikové body leží na silnicích. Zároveň budou z tohoto příkladu patrné možnosti dědičnosti, protože každá další podtřída vytvořená ve třídě `Rizikove_body` bude mít nastavenou vlastnost `Lezi_na Silnice`.
4. Pokusíme se nadefinovat rizikový bod `Brod`. Jaké informace máme o tomto bodu? `Brod` leží na silnici (tato vlastnost je zděděna od rodičovského prvku), ale nesmí to být dálnice a také `brod` leží tam, kde silnice protíná vodní tok. Následující body ukazují formy zápisu těchto deklarácí:

1. `Lezi_na only Silnice`
2. `Lezi_na only not Dalnice`
3. `Protina some Vodni_tok`

Podobným způsobem je možné definovat i další rizikové body na silnicích:

- Křižovatka – silnice protíná jinou silnici, kromě dálnice.
- Mimoúrovňová křižovatka – silnice vede nad nebo pod jinou silnicí
- Železniční přejezd – silnice, kromě dálnice, protíná železniční trať (s výjimkou rychlostního koridoru)
- Most – silnice vede nad libovolným prvkem

- Podjezd pod železnicí – silniční komunikace prochází pod železniční tratí

Je však potřeba si uvědomit, že přes nespornou složitost stojí ontologie na začátku celého procesu tvorby datového modelu nebo dokonce celé aplikace. Na ontologie navazují další nástroje sémantického webu, databáze, znalostní báze, software agents apod. Na druhou stranu kvalitně vytvořená otevřená ontologie ušetří velké množství práce (a tedy i nákladů) při dalších navazujících činnostech.

Interoperabilita a standardizace

Jako bylo popsáno v úvodu tvorba SDI představuje velice aktuální problém současných geoinformačních věd. Budování SDI (kromě jiného) znamená umožnit kombinování dat a služeb od různých poskytovatelů. Proto je důležitá interoperabilita data a služeb, které může být dosaženo pouze prostřednictvím standardizace formalizovaných popisů dat i služeb. Důsledkem fungující SDI bude především snížení nákladů na data i služby, protože uživatelé nebudou nuceni vlastnit a spravovat (aktualizovat) rozsáhlé sady prostorových dat a robustní procesní systémy, ale budou si kupovat nebo pronajímat pouze ty části SDI, které budou skutečně potřebovat.

Spolupráce na poli GIT

Tato kapitola chápe interoperabilitu také jako výsledek spolupráce na poli standardizace. Z tohoto důvodu budou následující odstavce obsahovat stručný popis vybraných subjektů přispívajících k vlastní standardizaci nebo k rozvoji standardů v oblasti GIT a prostorových dat.

The Open Source Geospatial Foundation (OSGeo)

Důležitou mezinárodní organizací zabývající se podporou vývoje open source software z oblasti geoinformačních technologií a jeho propagací je nadace OSGeo – The Open Source Geospatial Foundation (<http://www.osgeo.org/home>). Tuto organizaci podporují přední světoví výrobci komerčního GIT software, například společnosti Autodesk, PCI Geomatics nebo Lizardtech. Konkrétní aktivity OSGeo jsou v českém jazyce popsány na webových stránkách OSGeo Local Chapter (http://grass.fsv.cvut.cz/wiki/index.php/OSGeo_Local_Chapter):

- Poskytování zdrojů pro projekty nadace - infrastruktura, právní podpora,
- Poskytování volně dostupných geodat - free software je nepoužitelný bez dat.
- Podpora používání open source software v průmyslu (nikoliv pouze software nadace) - PR, kurzy.
- Prosazování implementace otevřených standardů a na nich postavené interoperabilitě v projektech nadace.
- Zajištění vysokého stupně kvality v projektech nadace a budovat tak "značku" nadace.
- Větší zpřístupnění projektů nadace koncovým uživatelům - např. tvorba balíčků, dokumentace.
- Poskytování podpory pro větší nasazení projektů nadace v oblasti vzdělávání.
- Podpora komunikace a spolupráce mezi komunitami OSGeo na úrovni programovacích jazyků (Java/C/Python) a operačních systémů (Win32, Unix, MacOS).
- Podpora a pomoc při vývoji projektů nadace díky internacionalizaci programů a přesahu mezi komunitami.
- Organizace roční OSGeo konference, pravděpodobně ve spolupráci s dalšími projekty (např. EOGeo).
- Cena Sola Katze za služby komunitě OSGeo.

Přístupnost dat a její omezení

Licence dat a software

Data i software téměř vždy nestojí sami o sobě, ale jsou propojeny s nějakými pravidly, která specifikují a omezují jejich používání, možnosti modifikace, začlenění do vlastních aplikací a další distribuci. Tato pravidla bývají označovány jako licence. Následující odstavce se zabývají problematikou licencí ve spojení s prostorovými informacemi a obecně s GIT. Kapitola nebude zaměřena na všechny typy licencí, ale pouze na tzv. otevřené licence. Důvodem je jejich značné rozšíření, zatím především na poli software (ale budou ukázány také případy datových šířených pod těmito licencemi). Otevřené nebo svobodné licence tak přispívají k vyšší popularitě GIT, jejímž důsledkem může být (a často také je) volání uživatelů po dodržování standardů, přičemž právě technologie a data distribuované pod některou z popisovaných licencí představují důležitý prvek v šíření standardů. V neposlední řadě je nutné zmínit také ekonomickou stránku používání otevřených technologií a dat. Ekonomické výhody do značné míry ilustruje rozvoj těchto technologií a dat v oblastech státní správy mnoha zemí světa, ale i podpora státních i soukromých organizací při vývoji těchto systémů.

V textu budou zmíněny budou jednotlivé typy licencí, jejich odlišnosti a velice frekventované termíny „open source“ a „free“.

Open source

Definice open source (Open Source Definition)³⁰, kterou publikuje a využívá organizace Open Source Initiative (www.opensource.org), stanovuje, která softwarová licence může být označena jako open source (v současnosti se v odborných kruzích poměrně široce diskutuje, zda je vhodnější a tedy i správnější označení free software nebo open source software). Definice se skládá z deseti základních bodů, které jsou popsány v následující tabulce (překlady definic podle publikace Arnošt, 2001):

Tabulka 2: Definice open source

1. Free Redistribution	Licence nesmí omezovat prodej nebo jinou distribuci programu jako součásti programového balíku obsahujícího software z různých zdrojů; licence by za takový prodej neměla vyžadovat autorský nebo jiný honorář.
2. Source Code	Produkt musí obsahovat zdrojový kód a musí umožňovat distribuci jak ve zdrojové, tak v binární ("zkompilované") podobě; pokud program není šířen včetně zdrojových kódů, musí být dobře popsána možnost jejich získání, a to za přiměřený poplatek (pokrývající náklady), nebo v případě Internetu zdarma; zdrojový kód nesmí být zamlžen; přechodné formy (např. výstup preprocesoru nebo překladače) nejsou dovoleny.
3. Derived Works	Licence musí umožňovat tvorbu odvozených prací a musí jim umožňovat,

30 Tato definice byla založena na bázi Debian Free Software Guidelines (autorem byl Bruce Perens). V současnosti je na stránkách Open Source Initiative dostupná verze uveřejněná 7.7.2006, jejímž autorem je Ken Coar.

	aby byly šířeny pod stejnou licencí jako původní produkt.
4. Integrity of The Author's Source Code	Licence může omezovat distribuci změněné formy zdrojového kódu pouze v případě, že je umožněno šíření tzv. záplat (patch files) spolu se zdrojovým kódem; licence musí výslovně povolit šíření programu přeloženého ze změněného zdrojového kódu; licence může vyžadovat, aby odvozené práce nesly jméno nebo verzi odlišné od původního programu.
5. No Discrimination Against Persons or Groups	Licence nesmí diskriminovat osoby nebo skupiny osob.
6. No Discrimination Against Fields of Endeavor	Licence nesmí omezovat použití programu v určité sféře; nesmí například omezovat použití programu v komerčním prostředí nebo v genetickém výzkumu.
7. Distribution of License	Práva přiložená k programu musí platit pro všechny, bez nutnosti dalších přídavných licencí.
8. License Must Not Be Specific to a Product	Práva přiložená k programu nesmí záviset na existenci programu v určitém programovém balíku; pokud je program z balíku vyřazen a je používán nebo šířen v souladu s licencí, všichni, ke kterým se program dostane, by měli mít stejná práva jako ti, kteří dostanou program jako součást programového balíku.
9. License Must Not Restrict Other Software	Licence nesmí klást omezení na software, který je šířen společně s licencovaným programem; licence nesmí například trvat na tom, aby všechny programy distribuované na stejném médiu splňovaly podmínky Open Source software
10. License Must Be Technology-Neutral	Žádné ustanovení licence nesmí být založeno na jakékoli individuální technologii nebo stylu rozhraní.

Podle publikace Open Source GIS Software: Myths and Realities (Camara, 2004) je většina open source produktů v oblasti GIT vytvářena jako individuální projekty (53%) nebo jako projekty vzniklé pod hlavičkou nějakého sdružení či konsorcia (41%). Jako méně používaný, ale především v oblasti GIT velmi efektivní model je označen termínem „networked team“ – tato varianta přináší podle autorů zmíněné publikace nejvyšší míru inovací a funkcionality.

Tabulka 3: Typy open source projektů

Způsob vzniku open source produktu	Příklady GIT aplikací
Individuální projekty	GDAL, GeoServer, QGIS...
Networked team	GRASS, GeoTools, MapServer, R...

Konsorciální projekty	PostGIS, GeoNetwork, Thuban, JUMP...
-----------------------	--------------------------------------

Individuální projekty jsou především doménou knihoven pro zpracování geoprostorových dat (GDAL, OGR, Proj4...). Ty byly většinou tvořeny jednotlivci, jako jsou Frank Warmerdam nebo Gerald Evenden.

Ostatní aplikace vznikly a vznikají většinou pod křídly velkých projektů, které jsou financovány zčásti veřejných a zčásti ze soukromých zdrojů. Příkladem může být software GRASS, kde rozhodující podíl na jeho tvorbě měla US Army (konkrétně Construction Engineering Research Laboratory /USA-CERL/, Champaign, Illinois), ale na dalším vývoji se podílely i jiné federální instituce, univerzity (např. University of Illinois, Champaign, Illinois) a soukromé společnosti. V současnosti existují dvě koordináční centra vývoje – Baylor University, Texas a Università degli studi di Trento, Itálie.

Jako příklady projektů vzešlých z univerzitního prostředí je možné uvést databázi Postgres (University of California, Berkeley) nebo MapServer (University of Minnesota) – projekt byl financován organizacemi NASA a Minnesota Department of Natural Resources (MNDNR) a podílely se na něm i další subjekty (Refractions Research, GeoConnections or Blue Sphere), které dále participovaly na vzniku „konkurečního“ projektu Geoserver. Na první projekt navázala kanadská firma Refractions Research, která na bázi databáze Postgres vyvinula další nekomerční software PostGIS. Podobných příkladů kooperace by se dala najít řada, například systém GeoNetwork (Food and Agriculture Organization /FAO/, World Food Programme /WFP/ a United Nations Environment Programme /UNEP/), JUMP (Vivid Solutions, Inc, GeoConnections, British Columbia Ministry of Sustainable Resource Management, Canadian Centre for Topographic Information – Sherbrooke, Ontario Ministry of Natural Resources, Refractions Research) nebo OpenJUMP (nekomerční subjekty – JUMP Pilot Project, Projet SIGLE, Pirol; komerční společnosti – Lat/Lon GmbH, Integrated Systems Analysts, Inc; univerzity – University of Zurich, EPFL, Vysoká škola báňská – Technická univerzita Ostrava, Hafen-City University Hamburg,...).

Publikace (Meng, 2003) uvádí velké množství subjektů využívajících open source softwarové produkty. Jedná se často o významné komerční společnosti jako například Amazon.com, Cisco, RedHat, Yahoo, Apple Computer, Hotmail, Deutsche Telekom, Pixar, Sony, DreamWorks, Boeing nebo IBM. Mnohé z těchto společností se také podílí na jeho vývoji open source aplikací. Open source software je podporován i vládami mnoha států světa (např. Čína, Korejská republika, Německo, Francie, Velká Británie, USA, Malajsie, Singapur, Taiwan nebo Peru).

Předchozí odstavce ukazuje na velmi širokou podporu, provázanost a částečnou koordinovanost jednotlivých open source projektů a iniciativ. Tento fakt odporuje jednomu z častých argumentů odpůrců open source software, kteří souhrnně označují open source nálepkou krátkodobých individuálních projektů bez záruky dalšího vývoje a uživatelské podpory.

Na základě publikací (Meng, 2003), (Wheeler, 2004), (Ven, 2008) je možné shrnout obecné výhody a nevýhody open source aplikací (ve skutečnosti se jedná o výčet vlastností open source software, které mohou pro některé typy uživatelů představovat přednost, pro jiné zase zbytečnou zátěž):

Tabulka 4: Výhody a nevýhody open source

Výhody	Nedostatky
Uživatelé mají přímé propojení s vývojáři a designéry, což vede k snadnému a rychlému odstraňování chyb a distribuce aktualizovaných	Většinou neexistuje komfortní uživatelská podpora jako v případě komerčních produktů.

verzí.	
Open source produkty inklinují k platformní nezávislosti a implementaci mezinárodních standardů, což vede k širším možnostem interoperability, které v konečném důsledku znamenají nižší náklady na hardware a ostatní software.	„Standardy“ definované komerčními produkty jsou často (zvláště v oblasti GIT) více využívány než otevřené mezinárodní standardy.
Open source software je většinou „hardwarově nenáročný“ – je možné ho využívat i na technicky starších zařízeních.	Na rozdíl od komerčních produktů open source produkty často nereflektují novinky na poli hardware.
Open source aplikace je možné modifikovat a zařazovat jako moduly nebo knihovny do vlastních aplikací.	Komerční projekty tvoří často homogenní softwarové produkty – což může být výhodné z pohledu některých uživatelů a také z pohledu komunikace mezi jednotlivými částmi programu.
Cena	Nulová cena aplikací je často vyvážená potřebou využívání speciálních služeb (školení, nákup studijních materiálů).

Open Source nástroje jsou v současnosti voleny především z důvodu nižší ceny (většinu prostředků je možné investovat do nákupu kvalitních datových sad nebo služeb), vysoké kvality a především možnosti zásahu do programového kódu a tudíž přizpůsobení požadavkům uživatele. Z pohledu nasazení open source z oblasti GIT jsou v současné době nejdůležitější serverové technologie (MapServer, Geoserver), které stojí v pozadí mnoha komerčních i otevřených aplikací poskytujících geoprostorová data prostřednictvím kartografických interpretačních metod.

Licence

Licenční politika, kterou můžeme chápat jako stanovení limitů pro poskytování a další používání software, dokumentů nebo datových sad, představuje jeden z charakteristických rysů, podle nějž je možné popsat a hodnotit vhodnost software pro dané účely.

Reprezentativní výběr nekomerčních programů a dat zaměřených na GIT je k dispozici na webové stránce www.freegis.org. V současnosti (6.3.2009) obsahují webové stránky freegis.org

- 349 softwarových produktů,
- 25 datových sad,
- 19 dokumentů,

- 10 projektů.

Následující tabulka obsahuje typy licencí a jejich zastoupení ve zkoumaném vzorku software a jiných GIT produktů.

Tabulka 5: Licence GIT produktů

Licence	Počet produktů
GNU GPL	182
GNU LGPL	66
MIT	28
Public Domain	25
BSDish	16
GPL (v2), GNU FDL	12
MPL 1.1, Apache License	7
X11-Style	6
GNU GPLv3	3
Creative Commons, Artistic License, AFL 2.1, GEOTRANS License	2
CeCILL V.2, vhelmaps License, SVG Toolkit license, shptrans license, Partly Public Domain, QPL, OpenMap License, NCSA-Type license, MapServer License, kdem License, IPW License, GSLIB License, GIS Viewer license, AVPython License (MIT-like), Apache License/non-free	1

Jak je patrné z předchozí tabulky dominantní roli mezi GIT technologiemi, resp. mezi jejich způsoby šíření hrají čtyři hlavní licence:

1. GNU GPL (<http://www.gnu.org/licenses/licenses.html#GPL>)

Licence GNU GPL (GNU General Public License) umožňuje přístup ke zdrojovému kódu, který může být libovolně modifikován, používán ve vlastních aplikacích a dále šířen. Dodržování těchto pravidel (tzv. copyleft) představuje základní princip GNU GPL. Pod touto licencí jsou poskytovány například programy jako GRASS, JGRASS, PostGIS, QGIS, JUMP, OpenJUMP, Thuban, gvSIG, Geoserver, OpenSVGMapserver, GeoNetwork, GNU GaMa or MySQL Spatial Extensions.

2. GNU LGPL (<http://www.gnu.org/copyleft/lesser.html>)

GNU LGPL (GNU Lesser General Public License, dříve GNU Library General Public License) byla již podle názvu odvozena z licence GNU GPL. Licence GNU LGPL je méně striktní v definování různých omezení, proto umožňuje propojení s komerčními aplikacemi. GNU LGPL je často využívána v případě nahrazení closed source (uzavřeného) kódu otevřenou aplikací s podobnými funkcemi. Pod touto licencí jsou šířeny programy jako MapGuide Open Source, deegree, uDig, SAGA, Mapyrus, Mapnik, library GEOS or GeoTools.

3. MIT (<http://www.opensource.org/licenses/mit-license.php>)

Text licence MIT (správně se označuje jako Licence Expat) je velmi stručný, jednoduchý a

především plně kompatibilní s licencí GNU GPL. Tato licence se vztahuje například na některé hojně používané knihovny jako například GDAL (licence X/MIT), OGR (dříve byly obě knihovny šířeny pod BSD Licence) nebo Proj4 (MIT-Style licence). Také způsob poskytování a využívání populárního UMN Map Server je odvozen z licence MIT.

4. Public Domain

Produkty distribuovány pod touto platformou jsou zbaveny veškerých autorských práv a omezení. To znamená, že kód může zařazen do jakýchkoli aplikací, včetně proprietárních a komerčních. Na úrovni Public Domain jsou distribuována především data nebo knihovny, ale vztahuje se též na některé programy, například digvu nebo Spherkit. Zastoupení této licence v oblasti GIT postupně klesá.

Struktura licence používané pro distribuci geografických dat je poněkud odlišná od předchozího případu, kde mezi GIT převažují softwarové produkty. Data popsána na serveru freegis.org jsou nejčastěji poskytována jako Public Domain (14 případů). Ostatní licence mají zastoupení výrazně nižší:

- GNU GPL: 4
- Bez popsané licence: 3
- Creative Commons: 2
- GNU FDL: 1
- Partly Public Domain: 1

Licence GNU Free Documentation License (GNU FDL³¹; <http://www.gnu.org/copyleft/fdl.html>³²) představuje dalšího zástupce skupiny copyleft licencí. Tato forma licence, jak vyplývá z jejího názvu, se zaměřuje především na textové dokumenty, konkrétně dokumentaci ke nejrozličnějšímu software. Podmínky jsou podobné jako v případě GNU GPL – v textu musí být uvedeni všichni autoři (i v současnosti neaktivní) a všechny změny, odvozená díla musí být licencována stejným způsobem, nesmí se používat technické prostředky zabraňující distribuci nebo úpravy dokumentů.

Licence Creative Commons (CC; <http://creativecommons.org>³³) souvisí se stejnojmennou americkou neziskovou organizací specializující se na legální šíření autorských děl (fotografie, video, zvuk, texty apod.) mezi širokou veřejností. Licence CC vzniká odlišným způsobem od podobných produktů. Nejedná se o jedinou licenci, ale o licenční schéma, které je schopné na základě splnění nebo nesplnění čtyř konkrétních vlastností (viz tabulka na konci odstavce) generovat konkrétní licenci. Teoreticky je k dispozici 16 různých licencí, ale prakticky se používá pouze šest základních (<http://creativecommons.org/licenses/>).

Tabulka 6: Vlastnosti používané v licenci Creative Commons

Název vlastnosti	Popis vlastnosti
Attribution	Rozmnožování, rozšiřování, vystavování a sdělování díla a z něj odvozených děl pouze při uvedení autora.
Noncommercial	Rozmnožování, rozšiřování, vystavování a sdělování díla a z něj odvozených děl pouze pro nevýdělečné účely.
No Derivative Works	Rozmnožování, rozšiřování, vystavování a sdělování díla pouze v původní

31 Používají se také zkratky GFDL nebo FDL.

32 Neoficiální český překlad verze 1.2 na stránkách serveru root.cz – <http://knihy.root.cz/kniha/cesky-preklad-gnu-free-documentation-license-1-2/>.

33 Česká verze Creative Commons – <http://www.creativecommons.cz/>.

	podobě.
Share Alike	Rozmnožování, rozšiřování, vystavování a sdělování odvozených děl pouze za podmínek identické licence.

Závěr

Úkolem předchozích kapitol mělo být zdůraznění a vysvětlení důležitosti popisu dat. O tom, že popis dat, včetně sémantiky, představuje velice důležitý nástroj zlepšující kvalitu vyhledávání informací, svědčí i následující tabulka. Ta srovnává vyhledávání informací založené na různé úrovni popisu dat – klíčových slov (velmi elementární popis dat) a ontologií (jeden z nejsofistikovanějších typů popisu dat).

Tabulka 7: Srovnání vyhledávání informací na základě klíčových slov a ontologií (Svátek, 2002)

	Klíčová slova	Ontologie
Úspěšnost – obecné dotazy	20-44%	92-94%
Úspěšnost – specifické dotazy	71-83%	83-91%

Jaké operace mohou pozitivně ovlivnit správně dokumentovaná (popsaná) prostorová data?

- Porozumění významu prostorových dat jak ze strany uživatelů, programátorů nebo zpracovatelů, tak ze strany technologických prostředků, včetně aplikace postupů umělé inteligence.
- Rychlost a přesnost vyhledávání informací, včetně možnosti zadávání konkrétnějších a přesnějších dotazů.
- Kontextové zpracování a vizualizace prostorových dat.
- Rychlejší přenos a zpracování dat (na základě dotazů budou zpracovávána a distribuována pouze požadovaná data).
- Přístupnost vlastních datových sad a zprostředkovaně také aplikací využívajících prostorová data.
- Internacionalizace a multilingualita dat.

Bibliografie

- . ANNONI, A., FRIIS-CHRISTENSEN, A., LUCCHI, R. & LUTZ, M. (2008). Requirements and Challenges for Building a European Spatial Information Infrastructure: INSPIRE. In van Oosterom & P., Zlatanova, S. Creating Spatial Information Infrastructures. Towards the Spatial Semantic Web. CRC Press, Taylor & Francis Group, London. ISBN 978-1-4200-7068-2.
- . ARNOŠT, P., 2001. Co je to Open Source software [online]. Root.cz, 22.8.2001 [cit 6.3.2009]. URL: <http://www.root.cz/clanky/co-je-to-open-source-software/>.
- . BENDA, L., 2005. Je sémantický web řešením? [online]. In Vesmír 84, 2/2005. URL: <http://www.vesmir.cz/clanek/je-semanticky-web-resenim>.
- . BORGIDA, A., MYLOPOULOS, J., 2005. Data Semantics Revisited [online]. Springer Berlin / Heidelberg, p. 9-26, 2005 [cit 7.3.2009]. ISSN: 0302-9743 (Print) 1611-3349 (Online). ISBN: 978-3-540-24576-6. URL: <http://www.springerlink.com/content/d4m5vtmtc38fxnwb/fulltext.pdf>.
- . BURKETT, W., 2002. The Myths of "Standard" Data Semantics [online]. In XML Journal, SYS-CON Media, Inc., 24.10.2002 [cit. 10.3.2009]. URL: <http://xml.sys-con.com/node/40511>.
- . CAMARA, G., ONSRUD, H., 2004. Open Source GIS Software: Myths and Realities [online]. In: Julie M. Esanu and Paul F. Uhler, Eds, Open Access and the Public Domain in Digital Data and Information for Science: Proceedings of an International Symposium. Washington, The National Academies Press, 2004 [cit 6.3.2009]. ISBN 0-309-09145-4. URL: http://www.dpi.inpe.br/gilberto/papers/camara_open_source_myths.pdf.
- . CRAGLIA, M. et al, (2003). GI in the wider Europe [online]. GINIE consortium, University of Sheffield, 2003 [cit 5.3.2009]. URL: http://www.ec-gis.org/ginie/doc/ginie_book.pdf.
- . ČADA, V. & MILDORF, T. (2005). Delimitation of reference geodata from land data model. GIS Ostrava 2005. Ostrava: VŠB - TUO, 2005. s. 1-12. ISSN 1213-239X.
- . ČERBA, O., CHARVÁT, K., JEŽEK, J., KAFKA, Š. & MUSIL, M. (2009). Spatial data interoperability makes ICT use more efficient.
- . Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). (2007)
- . GARDELS, K., 1996. The Open GIS approach to distributed geodata and geoprocessing [online]. Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, January 21–25. [cit 5.3.2009]. URL: http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/gardels_kenn/ogismodl.html.
- . Geographic information — Metadata. Information géographique — Métadonnées . ISO 19115:2003. (2003), ISO.
- . JEDLIČKA, K.; BŘEHOVSKÝ, M.; ŠÍMA, J., 2003 Úvod do geografických informačních systémů [online]. Plzeň: Západočeská univerzita, 2003 [cit 4.3.2009]. 116 s. URL: <http://gis.zcu.cz/studium/ugi/elearning/index1.htm>.
- . KAŇOK, J., 1999. Tematická kartografie. 1. vyd. Ostrava: Přírodovědecká fakulta

- Ostravské univerzity, 1999. 318 s. ISBN 80-7042-781-7.
- . MENG, T.T., 2003. The Case for Open Source: OSS vs Proprietary Software [online]. Cospa Knowledge Base, 14.4.2003 [cit 6.3.2009]. URL: http://kb.cospa-project.org/retrieve/3355/tan_tze_meng.pdf.
 - . MADNICK, S., ZHU, H. 2006. Improving data quality through effective use of data semantics. Data Knowl. Eng. 59, 2 (Nov. 2006), 460-475.
 - . MYLOPOULOS, J., 2004 Data Semantics Revisited: Databases and the Semantic Web [online]. In DASFAA'04, Jeju Island, Korea, March 17-19, 2004 [cit 7.3.2009]. URL <http://aitrc.kaist.ac.kr/~dasfaa04/doc/JM-DASFAA.pdf>.
 - . NEUMANN, J., 1996. Geografická informace. Český výkladový a anglicko-český a česko-anglický překladový slovník. Praha: Ministerstvo hospodářství ČR, 1996. ISBN 80-212-0130-4.
 - . NIEMANN, B. (2005). Data Reference Model Implementation Through Iteration and Testing Version 1.0. Retrieved November 19, 2009 from <http://web-services.gov/DRMITIT10172005.doc>.
 - . PEREGRIN, J. 2005. Co je to sémantika? [online]. P. Sousedík (ed.): Jazyk, logika věda, Filosofie, Praha, 2005, s. 51-62 [cit 11.3.2009]. URL: <http://jarda.peregrin.cz/mybibl/PDFTxt/499.pdf>.
 - . PRAŽÁK, J. et al., 2005-2009. Terminologický slovník zeměměřičství a katastru nemovitostí [online]. Zdíby: Výzkumný ústav geodetický, topografický a kartografický, Terminologická komise Českého úřadu zeměměřického a katastrálního, 2005-2009 [cit 4.3.2009]. URL: <http://www.vugtk.cz/slovník>.
 - . RAPANT, P., 2006. Geoinformatika a geoinformační technologie. Vysoká škola báňská – Technická univerzita Ostrava, Ostrava, 2006. ISSN 80-248-1264-9.
 - . SHETH, A., 2001. Data Semantics: what, where and how? [online]. [cit 7.3.2009]. URL: <http://www.dke.univie.ac.at/semanticweb/history/ws0405/S95.pdf>.
 - . SHETH, A., 2003. Semantic Meta Data for Enterprise Information Integration [online]. In Information Management Magazine, 2003 [cit 20.3.2009]. URL: <http://www.information-management.com/issues/20030701/6962-1.html>.
 - . SKLENÁK, V., 2004. Metadata, sémantika a sémantický web [online]. In INFORUM 2004: 10. konference o profesionálních informačních zdrojích, Praha, 25.–27. 5. 2004 [cit 11.3.2009]. URL: http://www.inforum.cz/pdf/2004/Sklenak_Vilem1.pdf.
 - . SVÁTEK, V., 2002. Ontologie a WWW [online]. In DATAKON 2002, Brno, 19.-22.10.2002 [cit 7.4.2009]. URL: <http://nb.vse.cz/~svatek/onto-www.pdf>.
 - . ŠÍMA, J., 2003. Geoinformační terminologie pro geodety a kartografy. 1. vyd. Zdíby: Výzkumný ústav geodetický, topografický a kartografický, 2003. 87 s. ISBN: 80-85881-20-9.
 - . ŠLAPÁK, O., 2003. Data, informace, znalosti [online]. In E-LOGOS ELECTRONIC JOURNAL FOR PHILOSOPHY / 2003 [cit 23.5.2009]. ISSN: 1211-0442. URL: <http://nb.vse.cz/kfil/elogos/miscellany/slapa103.pdf>.
 - . VEN, K., VERELST, J., MANNAERT, H., 2008. Should You Adopt Open Source Software? [online]. IEEE software, May/June 2008 [cit 6.3.2009]. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4497765.

- . VOŽENÍLEK, V., 2001. Aplikovaná kartografie I. Tematické mapy. 2. vyd. Olomouc: Univerzita Palackého, 2001. 187 s. ISBN: 802440270X.
- . WHEELER, D.A., 2004. Why Open Source Software / Free Software (OSS/FS)? Look at the Numbers! [online]. RobotCub.org, 7.11.2004 [cit 6.3.2009]. URL: <http://www.robotcub.org/index.php/robotcub/content/download/290/1049/file/Why%20Open%20Source%20Software.pdf>.

Přílohy

Požadavky a zdroje

Následující seznam obsahuje témata a náměty na zdroje informací, které jsou společně s předchozím textem důležité pro úspěšné složení zkoušky z předmětu Databázové systémy GIS 2 (KMA/DBG2).

- Prostorová data, klasifikace dat, prostorové a konceptuální modely – zopakovat podle studijních materiálů z předmětu KMA/UGI.
- Otevřené výměnné formáty pro prostorová data – standardy GML, KML a LandXML.
- Jazyky pro popis dokumentu – Jiří Kosek. XML schémata (dostupné na internetu); DSDL – materiály webových stránek www.dsdl.org a články Erica van der Vlista na téma DSDL (např. z konference XML Prague)..
- Metadata – prostudovat směrnici INSPIRE a další příslušné materiály a návrhy, standard Dublin Core a strukturu ISO 19XXX (doporučuji dokument „Relation between ISO 19115 and ISO 19119 and the elements of the INSPIRE draft metadata implementing rules“).
- Sémantika v kartografii, kartografická sémiologie – zopakovat podle studijních materiálů z předmětu KMA/TKA.
- Publikace Vilém Sklenák. Metadata, sémantika a sémantický web.
- Ontologie – tutoriály editoru ontologií Protégé, text Vojtěcha Svátka. Ontologie a WWW (případně podobné publikace), diplomová práce Radka Stuchlíka – Ontologie a sémantický web.
- Standardy – RDF, RDF Schema, OWL.
- Spolupráce v GIT – seznámit se s následujícími organizacemi a aktivitami a jejich činností: ISO, W3C, OGC, INSPIRE, EUROGI, Eurogeographics, CAGI.
- Přístupnost dat a map – článek O.Čerby z konference GIS... Ostrava 2008, diplomová práce K.Špicelové.
- Harmonizace dat – důvody harmonizace, principy harmonizace, typy harmonizace (výsledky projektů zabývajících se harmonizací dat, např. RISE, Orchestra, Humboldt; aktivity INSPIRE; články a prezentace, jejichž autorem je Clemens Portele).
- Zdroje prostorových dat – seznámit se pojmy (principy, příklady, důvody používání apod.) thesaurus (thesaurus), gazetteer, katalogová služba.

Dokument Data pro geoinformační technologie a jejich popis je určený především pro výuku předmětu Socioekonomická geografie pro geomatiku – KMA/SGG (dříve Databázové systémy GIS 2 – KMA/DBG2), který je přednášen na Katedře matematiky Fakulty aplikovaných věd Západočeské univerzity v Plzni.

Dokument je možné používat pro vlastní potřeby. Pokud tak učiníte, dodržujte prosím podmínky licence. Dokument je šířený pod licencí GNU FDL (Free Documentation License; <http://www.gnu.org/copyleft/fdl.html>).

Otakar Čerba, 2009-2010

Poslední aktualizace: 26.1.2010