

Metainformačné systémy a metadáta – aktuálne témy

Dagmar Kusendová

Katedra humánnej geografie, Prírodovedecká fakulta, Univerzita Komenského, Mlynská dolina,
842 15 Bratislava, Slovenská republika
kusendovva@fns.uniba.sk

Abstrakt. Cieľom príspevku je ukázať na význam metadát a súvisiacich aktuálnych tém v kontexte použitia geografických dát a budovania geoinformačných infraštruktúr a ich metadátových služieb.

Kľúčové slová: metadáta, sémantický web, Európsky metadátový rámec, INSPIRE

Abstract. The aim of the paper is to point out the significance of metadata in relation with current application of geographical data, creation of geoinformatic infrastructures and their metadata services.

Keywords: metadata, semantics web, Europe metadata framework, INSPIRE

1 Úvod

Metadáta, metainformačné infraštruktúry a ich služby tvoria prostredie, ktoré je určené na sprístupnenie potrebných dát pre tvorbu rozhodnutí a ďalšie účely. Ide o prostredie, ktoré by malo vytvoriť vhodné podmienky na prístup k dátovým zdrojom, ich efektívneho využitia, usporiadania a údržby tak, aby bolo pokryté čo najväčšie spektrum informačných potrieb.

Cieľom príspevku je ukázať na význam metadát a s nimi relevantných tém v kontexte použitia geografických dát (geodát), budovania geoinformačných infraštruktúr a metadátových služieb v prostredí internetu s dôrazom na regionálnu, t. j. európsku úroveň. Inšpiráciou k napísaniu príspevku bola práca B. Horákovej [8], ktorá vyčerpávajúco pokrýva uvedenú tematiku a tvorila ťažiskovú informačnú bázu príspevku.

2 Metadáta a metadátové štruktúry

Kvalitné informácie sú základom kvalitných rozhodnutí. Nositeľom kvality dátových zdrojov sú metadáta. Metadáta by mali komplexne charakterizovať ľubovoľný informačný objekt, ktorý je určený pre prenos, spracovanie alebo použitie z hľadiska charakteru, obsahu, kontextu štruktúry a ďalších charakteristík. Metadáta teda identifikujú, opisujú informačný objekt (napr. databázové zdroje) a dokumentujú jeho

„chovanie“, funkcie, a spôsobov manipulácie s ním a jeho využitia, ako aj väzby a vzťahy k iným informačným objektom. Objekt a jeho metadáta majú odlišný charakter a funkcie, ale musia z hľadiska reálneho využitia tvoriť jeden logický celok. Významným on-line dostupným zdrojom ucelenej terminológie pokrývajúcej metainformačné a ďalšie termíny je v ČR Česká terminologická databáza knihovníctva a informačnej vedy [15]. Podľa tohto zdroja možno rozlišovať tieto základné kategórie metadát z hľadiska ich obsahu a funkcie:

- administratívne (využívané na riadenie a správu informačných zdrojov, ako je napr. názov zdroja a jeho lokalizácia, dátum aktualizácie apod.),
- archivačné (vzťahujúce sa k procesom archivácie zdrojov a zabezpečujúce integritu a kontext metadátového opisu v čase),
- právne (dokumentujú právne nároky vzťahujúce sa k opisovanému zdroju, ako sú napr. autorské práva, kopírovanie zdroja apod.),
- používateľské (poskytujú charakteristiky pôvodného objektu tak, aby používatelia boli schopní pochopiť jeho obsah a použitie),
- technické (informujú ako sa systém chová, resp. čo je treba na jeho sprevádzkovanie – napr. parametre hardvéru, protokol HTTP apod.),
- štruktúrne (definujú vnútornú organizáciu objektu a sú potrebné na zobrazenie a usmernenie/navigáciu tohto objektu).

Z hľadiska úrovne abstrakcie sa metadáta dajú členiť podľa [8, 12] na:

- syntaktické, ktoré poskytujú podrobnosti o dátovom zdroji, slúžia na ich katalogizáciu a kategorizáciu (napr. formát súboru, použitý jazyk, autorizácia, dátum vzniku zdroja a ďalšie),
- štrukturálne, ktoré dokumentujú štruktúru dokumentu pre potreby jeho uloženia, spracovania alebo prezentácie (napr. schéma XLS),
- sémantické, ktoré opisujú kontextové informácie s dôrazom na odvetvové/doménové špecifikácie.
- ontológie, ktoré predstavujú najvyššiu formu metadát a sú základom sémantického webu.

Metadáta by mali byť komplexné a kompletne pri opise zdroja s cieľom objaviť informačný zdroj, oceniť a vymedziť jeho vhodnosť, resp. spôsobilosť k použitiu, umožniť prístup k zdroju, zabezpečiť jeho použitie, umožniť výmenu dát a informácií a riadenie informačných zdrojov. Za najvýznamnejšie prínos použitia metadát považuje Horáková [8] to, že poskytujú informácie pre dátové katalogy a umožňujú efektívnu výmenu a implementáciu externých dát.

2.1 Logika a sémantika metadát

Potrebnú logiku a sémantiku poskytujú metadátam, ako aj príslušným systémom, metadátové štandardy a štandardy z oblasti ontológie. Najmä ontologické štandardy by mali definovať spoločný, resp. unifikovaný systém definícií pojmov zrozumiteľných človeku ako aj počítaču. Mali pomôcť zaistiť nutnú mieru súčinnosti (interoperability) v získavaní vyššej využiteľnosti poskytovaných geografických dát a informácií, a to najmä pri automatickej identifikácii a nájdenia dátového, resp. informačného zdroja, jeho obsahu, štruktúry, použiteľnosti apod. Najčastejšie poskytujú potrebnú sémantiku a logiku v danej oblasti záujmu, resp. profesie

(doméne), napr. v geodézii, geoinformatike apod. V oblasti štandardizovanej ontológie sú neformálne a poloformálne glosáre a diagramy postupne nahrádzané strojovými ontologickými jazykmi s využitím prostriedkov ontologického inžinierstva [13]. Aj keď v súčasnosti existuje viacero doménových ontológií, v oblasti geoinformatiky je tento proces len v počiatkoch a využívajú sa najmä selekčné jazyky a riadené slovníky a z nich osobitne tezaury, ktoré tvoria sémantickú sieť konceptov spájajúce slová s ich bližšími a vzdialenejšími a súvisiacimi pojmami. Podrobnejšie sa problematike použitia ontológií v geoinformatike venujú práce [6] a [14].

2.2 Riadené slovníky a ich použitie v metadátových systémoch

„Riadený slovník je všeobecné označenie pre viac alebo menej štruktúrovaný zoznam termínov, z ktorého sú vyberané prípustné hodnoty pre opis zdroja na úrovni jednotlivých metadátových položiek. Opis informačných zdrojov sa neuskutočňuje voľne, ale prostredníctvom tzv. selekčných jazykov vo zvolenom formáte, čo by malo zaisťovať systémovosť a konzistencia opisu.“ [8, s. 34]. Uvedená definícia identifikuje riadený slovník ako súbor predmetových hesiel s definovanou hierarchickou štruktúrou termínov a synonymím alebo iných vzťahov medzi pojmami. Aj keď sa tieto slovníky v praxi definujú užšou, najčastejšie profesijnou skupinou používateľov, pre ich širšie použitie treba zabezpečiť, aby boli všeobecne dostupné, najlepšie prostredníctvom webovských aplikácií. Pri stanovení hodnôt vlastností zdroja, ktoré sú prezentované metadátovou položkou v konkrétnom metadátovom rámci alebo schéme, ako je napr. schéma značkovacieho jazyka XML alebo schéma normy ISO 19139, sa vyžaduje použitie pojmov (hesiel) len z tohto rámca/schémy, čo výrazne podporuje interoperabilitu samotných metadát. Príkladom riadených slovníkov, ktoré sa bežne používajú napr. v ČR sú OKEČ, CZ-NUTS (Český statistický úrad), GEMET (teraurus pre oblasť životného prostredia) [7].

2.3 Metadátové štandardy

Úloha sémantik je v metadátach geografických dát nesporná najmä v súvislosti s tým, že každá tvorba dát (či už manuálna, poloautomatická alebo automatická), aplikácia alebo databáza má svoj dátový slovník, termíny, koncepty, klasifikácie, taxóny. Metadáta, ktoré ich opisujú, by mali vnašať potrebnú sémantiku nielen v rámci jednej domény, ale aj naprieč viacerých domén. To sa dá dosiahnuť len na základe dôsledného používania otvorených medzinárodných štandardov skôr typu *de facto* než *de jure*. Dikertívne presadzované štandardy sa v praxi využívajú minimálne na rozdiel od dobrovoľne akceptovaných, vytvorených na základe dohôd, ktoré sú výsledkom prirodzeného vývoja a reálnych požiadaviek praxe. Aktuálny vývoj potvrdzuje, že tento druh štandardov sa stáva súčasťou noriem, ktoré sa postupne uplatňujú aj v praxi.

V súčasnosti sa celosvetovo vyprofilovala len malá skupina metadátových štandardov, ktoré tvoria základ pre vývoj ďalších, najmä doménových štandardov, tzv. aplikačných (komunitných) profilov. Základom pre tvorbu a aplikáciu metadát na

všeobecný opis informačných zdrojov je v súčasnosti Dublin Core (súčasť platnej normy ISO 15836:2003 [10] a na opis geografických dát norma ISO 19115:2003 [2].

2.4 Sémantický web

V súvislosti s rozvojom informačno-komunikačných štruktúr, a najmä internetu a jeho heterogenity, rastie význam metadát vo funkcii štruktúrovaných dát pre počítaťové spracovanie informácií o webovských zdrojoch, resp. geografických dátach a službiach vo forme sémantického webu [16]. Sémantický web je relatívne nový fenomén, ktorý je založený na „sémantických metadátach“. Predpokladom praktického použitia metadát týmito počítačovými nástrojmi je však tvorba korektných a konzistentných metadát s tvorbou používateľských rozhraní a programových aplikácií na ich jednoduchú tvorbu. Tieto prostriedky by výrazne pomohli „zavedeniu“ a použitiu všeobecne uznávaných a rešpektovaných štandardov, ktoré poskytujú automatizovanú syntax zápisu metadát a smetadátových položiek.

Najdôležitejšími všeobecne uznávanými štandardmi, ktoré tvoria technologický základ sémantického webu sú najmä tieto štandardy: eXtensible Markup Language (XML), resp. XML schémy [18] a Resource Description Framework (RDF, RDF schémy) organizácie W3C [17]. Metadáta odpovedajúce týmto štandardom by mali postupne zabezpečiť, že každý zdroj bude „vybavený“ charakteristikami, ktoré umožnia používateľom internetu pracovať v sieti WWW ako s relačnými databázami a dopytovať sa na ich obsah prostredníctvom jazykov podobných SQL.

3 Európsky metadátový rámec – nielen INSPIRE

Aktivity zamerané na využitie geoinformačných dátových zdrojov a najmä tvorbu geoinformačných metadátových infraštruktúr akcelerujú nielen celosvetovo, ale aj v európskom priestore. V oblasti normalizácie sú medzinárodné normy (ISO) postupne prijímané do sústavy európskych noriem (EN) prostredníctvom normalizačných organizácií (CEN – Európsky výbor pre normalizáciu, CENELEN – Európsky výbor pre normalizáciu v elektronike, ETSI – Európsky inštitút pre normalizáciu v telekomunikáciách). EN normy sú ďalej prijímané a postupne harmonizované a transponované do normalizačného rámca jednotlivých členských štátov.

V roku 2004 CEN zahájil oficiálne štandardizačné práce na vytvorenie Európskeho metadátového rámca. V rámci svojej sekcie CEN/ISSS (Information Society Standardization System) organizuje pracovné stretnutia WS/MII-DC (Workshop on Metadata for Multimedia Information – Dublin Core) odborníkov a organizácií z rôznych krajín Európy (SR nemá zastúpenie), kde pracujú na tvorbe technických pravidiel, dokumentácie, špecifikácie a odporúčaní pre oblasť metadát vo forme tzv. CEN Workshop Agreement-ov. O význame geografických dát a informácií svedčí skutočnosť, že bola prvou doménou, ktorou sa sekcia zaoberala.

Iniciatíva Európskej komisie INSPIRE má ambíciu tiež pripraviť vhodné technické, právne, organizačné podmienky a prostredie pre vytvorenie Európskej geoinformačnej infraštruktúry na báze národných infraštruktúr. Jednou z priorit, dnes už právoplatnej Smernice Európskeho parlamentu a Rady o vybudovaní

geoinformačnej infraštruktúry v Spoločenstve [2], sú metadáta spolu s kreovaním metainformačných systémov a služieb ako základných prvkov infraštruktúry, ktorých konečným cieľom je bezbariérové sprístupnenie geoinformačných zdrojov členských štátov Európskej únie. V správach expertných skupín INSPIRE [9] sa konštatuje existencia množstva kvalitných geografických dátových zdrojov v regionálnej a lokálnej mierke, ich širšiemu využitiu bránia najmä: absencia, resp. neúplnosť a nízka kvalita niektorých dát, nedostatočná dokumentácia, nekompatibilita, izolovanú platnosť vyhľadávacích a prístupových služieb a prílušných systémov, bariéry v zdieľaní a opätovnom použití dát. V podobnom duchu sú závery aj niektorých ďalších aktivít [11].

Uvedené nedostatky by sa mali postupne odstraňovať formou transpozície a implementácie opatrení Smernice Európskeho parlamentu a Rady 2003/98/ES o opakovanom použití informácií verejného sektora [5] a ďalších relevantných smerníc najmä pre oblasť verejnej správy a normalizácie (Smernica Európskeho parlamentu a Rady 98/34/ES o postupe poskytovania informácií v oblasti noriem a technických predpisov [4]) na úrovni členských štátov, teda aj SR. V tomto kontexte bola metadátovej norma ISO 19115:2003 už publikovaná ako európska norma EN ISO 19115:2005 a v súlade so smernicou 98/34/ES bola transponovaná do normalizačných sústav členských štátov EÚ. Cieľom týchto krokov je pripraviť najmä technický rámec pre tvorbu prostredia na zdieľanie dát vo verejnej správe. V rámci EÚ, je uvedená norma dôrazne odporúčaná v súvislosti s iniciatívou INSPIRE a nadväznými aktivitami súvisiace s tvorbou geoinformačných infraštruktúr v Európe (EÚ).

4 Záver

V závere treba súhlasiť s B. Horákovou [8, s. 31], že akademický výskum v oblasti metadátovej a ontologických systémov, sémantických metadát a vývoja relevantných programových nástrojov najmä vo forme sématického webu, ktorý významné posúva aj vývoj metadát, zatiaľ predbieha prax. Realizované praktické aplikácie takmer nevyužívajú výsledky výskumu, najčastejšie sa obmedzujú na terminologické ontológie bez širšieho využitia normalizačných nástrojov. Ide však o krátkozrakú politiku, nakoľko bez ich použitia bude dosiahnutie interoperability v oblasti geoinformačných zdrojov len veľkým želaním.

Referencie

1. CEN. CWA 15245. EU e-Government METADATA FRAMEWORK. Dostupné na WWW: <ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa15245-00-2005-Apr.pdf>,2007.
2. ČSN ISO 19115: Geografická informace – Metadata. Věstník Českého normalizačního institutu, 10/2004.
3. Directive 2007/2/EC of the European Parliament and of the Council of 14.3. 2007 establishing an Infrastructure for Spatial Information in the European Community

- (INSPIRE) Dostupné na WWW:
<[http://www.cenia.cz/web/www/web-pub2.nsf/\\$pid/MZPMSFGR0A25/\\$FILE/1_10820070425en00010014.pdf](http://www.cenia.cz/web/www/web-pub2.nsf/$pid/MZPMSFGR0A25/$FILE/1_10820070425en00010014.pdf)>,2007.
4. Directive 98/34 of the European Parliament and of the Council laying down a procedure for the provision of information in the fields of technical standards and regulations and of rules on information services, OJ L 204 of 21.7.1998, amended by Directive 98/48 of the European Parliament and of the Council, OJ L 217 of 5.8.1998. Dostupné na WWW:
<http://europa.eu.int/comm/enterprise/tris/consolidated/index_en.pdf>,
<http://europa.eu.int/eur-lex/pri/en/oj/dat/1998/l_204/l_20419980721en00370048.pdf>,2007.
 5. Directive of the European Parliament and of the Council on the re-use of public sector information (Smernica Európskeho parlamentu a Rady 2003/98/ES, o opakovanom použití informácií vo verejnom sektore). 17.11.2003. Dostupné na WWW: <<http://europa.eu.int/eur-lex/cs/dd/docs/2003/32003L0098-CS.doc>>,2007.
 6. Fonseca, F., Egenhofer, M., Agouris, P., and Câmara, G. (2002) Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS 6(3): pp. 231-257. Dostupné na WWW:
<http://www.personal.psu.edu/faculty/f/u/fuf1/publications/Fonseca_TGIS_2002.pdf>,2007.
 7. GEMET. General Multilingual Environmental Thesaurus. European Environment Information and Observation Network (EIONET). Dostupné na WWW:
<<http://www.eionet.eu.int/GEMET>>,2007.
 8. Horáková, B. Metainformační infrastruktura v evropském a národním měřítku. Vysoká škola báňská – Technická univerzita Ostrava, Hornicko-geologická fakulta, 2006, Ostrava (habilitační práce).
 9. INSPIRE. Selection of Experts for Implementing Rules Drafting Teams. JRC- Institute for Environment and Sustainability, Ispra. 2005-07-29. Dostupné na WWW: <http://www.ec-gis.org/inspire/reports/ir_dt_selection_experts_final_v5.pdf>,2007.
 10. ISO 15836:2003. Information and documentation - The Dublin Core metadata element set. Dostupné na WWW:
<<http://www.niso.org/international/SC4/n515.pdf>>,2007.
 11. MADAME. Methods for Access to Data and Metadata in Europe. Summary Final Report. March 2001. 2005. Dostupné na WWW:
<http://www.shef.ac.uk/~scgisa/MADAMENew/finalpublic.pdf> >,2007.
 12. Sklenák V.: Metadata, sémantika a sémantický web. INFORUM 2004: 10. konference o profesionálních informačních zdrojích Praha, 25.–27. 5. 2004. Dostupné na WWW:
<http://www.inforum.cz/inforum2004/pdf/Sklenak_Vilem1.pdf>,2007.
 13. Svátek V.: Ontologie a WWW. Dostupné na WWW:
<<http://nb.vse.cz/~svatek/onto-www.doc> >,2007.
 14. Visser Ubbo: Using Environmental Information Efficiently: Sharing Data and Knowledge from Heterogeneous Sources. Center for Computing Technologies.

- University of Bremen. Dostupné na WWW:
<<http://www.cs.vu.nl/~heiner/public/tzifinal.PDF>>,2007.
15. Výklad termínů z České terminologické databáze knihovnictví a informační vědy (TDKIV). Dostupné na WWW:
<http://www.nkp.cz/o_knihovnach/Slovník/index.htm>,2007.
 16. W3C. OWL (Web Ontology Language). Dostupné na WWW:
<<http://www.w3.org/TR/owl-features/>>,2007.
 17. W3C. RDF Schema. Dostupné na WWW: <<http://www.w3.org/TR/rdf-schema/>>,2007.
 18. W3C. XML Schema. Dostupné na WWW:
 19. <<http://www.w3.org/XML/Schema>>,2007.