

kázané látky. V mnoha evropských jazycích je možné popsat jeden jev nebo oblast více termíny, nebo naopak je jeden termín používán pro více oblastí a jevů, než je tomu v jiných jazycích. Překlad termínů do národních jazyků tedy není zdaleka jednoduchý a přímočarý a v mnoha případech je nutné velmi pečlivě rozhodování při volbě terminologie.

Český sociálněvědní archiv se v průběhu roku 2012 zařadil mezi ty archivy, které výhledově přijaly ELSST jako hlavní tezaurus pro popis svých datových souborů. K plnému zavedení tezauru do provozu je však ještě poměrně dlouhá cesta. V současné době byly podepsány dohody mezi Sociologickým ústavem AV ČR, v.v.i., a univerzitou v Essexu, která je držitelem licence pro tezaurus, a zároveň je dokončován překlad všech termínů, kterých je ale mnohem větší počet než v době, kdy byl ELSST zaváděn – dnes to pro představu znamená celkem 5653 termínů. Překlad je prováděn kvalifikovaným překladatelem, který je zároveň sociálním vědcem, což zaručuje adekvátní použití odborné terminologie. Výsledný seznam termínů pak bude odsouhlasen odborníky z různých oblastí českých sociálních věd, aby bylo zajištěno, že navržené termíny jsou skutečně v daných oblastech sociálních věd aktivně používány. Poté musí navržené termíny uzнат i mezinárodní konsorcium, které tezaurus ELSST spravuje. V další, velmi náročné fázi dojde k tomu, že celý datový katalog ČSDA bude popsán pomocí termínů ELSST a umožní tak klientům lepší vyhledávání pomocí mezinárodně uznané klasifikace. V neposlední řadě ale bude představovat také to, že český datový archiv bude přichystán na připravovanou integraci do budoucího mezinárodního datového katalogu, který je jedním z hlavních cílů projektu CESSDA.

Literatura

Balkan, Lorna, Ken Miller, Birgit Austin, Anne Etheridge, Myriam Garcia Bernabé, Pam Miller. 2002. „ELSST: a Broad-based Multilingual Thesaurus for the Social Sciences.“ Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02). Dostupné z: <<http://www.lrec-conf.org/proceedings/lrec2002/pdf/3.pdf>>.

EDDI 2012 – 4. EVROPSKÁ KONFERENCE TVŮRČŮ A UŽIVATELŮ METADATOVÉHO STANDARDU DDI

Tomáš Čížek (Sociologický ústav AV ČR)

Ve dnech 3. a 4. prosince roku 2012 se v norském Bergenu konala již čtvrtá evropská konference, věnovaná metadatovému standardu DDI (odvozeno od názvu sdružení, které standard vyvíjí – Data Documentation Initiative). Tento standard není čtenářům časopisu *Data a výzkum* zdaleka neznámý. Již v prvním čísle časopisu se mu věnoval článek Martina Vávry [2007] a třetí verzi DDI popsal Juraj Švec [2008]. Tento metadatový standard se nicméně neustále a zcela zásadním způsobem vyvíjí, a je tak nutná poměrně intenzivní komunikace mezi jeho tvůrci a uživateli. Konference EDDI je pak jednou z možností, jak se seznámit s novinkami ve vývoji DDI a také s příklady užití tohoto standardu v jednotlivých datových archivech.

Archivy a další pracoviště také vytvářejí množství softwarových aplikací pracujících s formátem DDI, které pak lze využít i v institucích, jež se na jejich tvorbě přímo nepodílely. Český sociálněvědní datový archiv, stejně jako ostatní menší archivy, si nemůže dovolit vývoj vlastních programů a aplikací, a tak je pro něj užitečné testování a úpravy stávajících softwarových řešení. Prezentace těchto programů a aplikací představovala významnou část z konferenčních příspěvků.

Podstatnou změnou ve formátu DDI byl přechod na verzi 3, která zásadním způsobem proměnila popis datového souboru od statického (dnes nazývaného DDI Codebook a číselně označeného 2) k dynamickému (DDI Lifecycle – DDI3), který umožňuje popsat datový soubor v čase, včetně různých změn a návazností na jiné datové soubory, což je velmi vhodné například k popisu panelových a jiných opakovaných výzkumů.

Český sociálněvědní archiv využívá DDI ve formě integrované do softwaru Nesstar. Jedná se tedy o verze 1.2.2 a některé části z verze 2.1, což jsou poměrně staré specifikace, které byly již podstatně vylepšeny v dalších verzích. Ani do budoucna se ale u systému Nesstar neuvažuje o přechodu na specifikaci 3 a vyšší, a to zejména z důvodu omezené kompatibility mezi verzemi „codebook“ a „lifecycle“. K použití nejnovějších verzí DDI je tedy nutná aplikace jiných programů, které moderních verzí DDI nativně využívají a které již byly mezinárodní komunitou vytvořeny a úspěšně implementovány. ČSDA tak začne zkušebně testovat na konferenci prezentované programy DdiEditor (vytvořený dánským datovým archivem) a archivační systém Questasy (vyvíjený v nizozemském archivu CentERdata). Výhodou těchto řešení je, že jsou poskytována bezplatně. Převod celého datového katalogu do nejnovějších verzí standardu DDI ale zatím v plánu není a tento potencionálně náročný krok bude zejména záležet na požadavcích organizace CESSDA, jejímž je ČSDA členem. Z ostatních zajímavých řešení, která byla na konferenci prezentována a která ČSDA do budoucna zvažuje zavést, jsou nové funkce systému Nesstar, jež umožňují vytvářet programové nadstavby tohoto systému a zpřístupnění jeho funkcí na dnes tolik populárních chytrých telefonech a tabletech.

Posledním významným tématem, kterému se konference věnovala, byla kompatibilita DDI se zaváděným standardem GSIM (General Statistical Information Model), který pro popis svých datových souborů plánují zavést národní statistické úřady. Vzhledem k podobnosti dat produkovaných sociálními vědami a oficiálními statistikami a vzhledem k tomu, že některé národní datové archivy slouží jako úložiště datových souborů oficiální statistiky, je nutné vývoj těchto formátů směřovat k co největší kompatibilitě. To je ovšem spíše záležitostí budoucího vývoje, i když nastoupená cesta se dá označit za poměrně slibnou.

Literatura

Švec, Juraj. 2008. „DDI 3.0: Kvalitativny posun v štandarde dokumentácie sociálnych dát.“ Informační článek. Data a výzkum – SDA Info 2 (2): 160–164.

Vávra, Martin. 2007. „Archivace sociologických dat.“ Data a výzkum - SDA Info 1 (1): 7–18.