

praktické procvičení nebylo organizováno tak, abychom se konkrétní analytické postupy skutečně naučili. Odpolední cvičení měla předimenzovaný program, a tak nebylo možné v daném čase provést všechny analýzy. Další komplikací byl fakt, že naprostá většina účastníků kurzu včetně mě neměla dostatečné zkušenosti s prací v programu Mplus, ve kterém jsme měli analýzy provádět. Během odpoledních cvičení jsme se často cítili ztraceni, i z toho důvodu, že řešení úkolů nebylo dostatečně znázorněno a vysvětleno.

Tento krátký text bych chtěla ukončit poznámkou, že metodologická škola institutu GESIS je podle mého názoru velmi záslužným projektem, neboť nabízí vysoce kvalitní kurzy a přednášející a umožňuje studentkám a studentům metodologie prohlubovat své znalosti, setkat se s odborníky a vyslechnout si velmi zajímavé přednášky i mimo rámec jednotlivých kurzů. Doufám, že letošní první letní škola GESISu nebyla poslední a že zájemci z řad začínajících metodologů budou mít možnost účastnit se jí i v dalších letech.

ZAVEDENÍ TEZAUROU ELSST v ČSDA

Tomáš Čížek (Sociologický ústav AV ČR)

Český sociálněvědní datový archiv se v blízké době stane jedním z archivů, které pro popis svých datových souborů používají mezinárodní tezaurus ELSST (European Language Social Science Thesaurus). Tento tezaurus je v současné době přeložen do devíti národních jazyků (angličtiny, němčiny, francouzštiny, švédštiny, norštiny, španělštiny, řečtiny, finštiny a dánštiny) a do některých dalších jazyků překlad probíhá (italština, litevština a čeština).

Jedním z cílů organizace CESSDA je vytvoření nadnárodního datového katalogu, kde si uživatel bude moci vybrat data z celého souboru výzkumů, uložených v jednotlivých národních datových archivech. K tomu je ale potřeba, aby výzkumy v jednotlivých národních archivech byly popsány mezinárodně standardizovaným způsobem. Hlavně ke splnění tohoto cíle byl tezaurus ELSST vytvořen.¹

Začátky evropského jazykového tezauro pro sociální vědy se váží k projektu LIMBER (Language Independent Browsing of European Resources), probíhajícího v letech 2000 až 2011, v rámci něhož se spojili pracovníci německého, britského, řeckého, francouzského a norského datového archivu, aby vytvořili systém, ve kterém bude možné doplnit metadata (informace o datech) datových souborů takovým způsobem, že budou mezinárodně srovnatelná. To si žádalo vyřešení nejen technické stránky věci, tedy použití vhodného mezinárodně použitelného softwarového formátu pro zápis metadat, ale hlavně konstrukci společného slovníku, kterým data budou popsána. Tento úkol nebyl a dodnes není zdaleka triviální, protože velká různost evropských jazyků představuje minimálně překážku pro celoevropsky, a potažmo celosvětově, platný systém pro popis datových souborů vzniklých na půdě sociálních věd.

1 Organizace CESSDA již na svých stránkách v sekci „CESSDA catalogue“ ukazuje příklad toho, jak bude takový mezinárodní datový katalog vypadat. Současný katalog obsahuje 8152 studií z 16 evropských archivů. Budoucí katalog by měl zahrnovat pokud možno všechny uložené datové soubory z členských archivů organizace.

Jako základ tohoto slovníku-tezauru byl vybrán anglický thesaurus HASSET (Humanities and Social Science Electronic Thesaurus). Tento thesaurus byl pro potřeby popisu datových souborů v britském archivu „UK Data Archive“ rozvíjen a doplňován již více než dvacet let a představoval tak ideální základ pro konstrukci chystaného mezinárodního tezauru. V době tvorby tezauru ELSST obsahoval HASSET přes čtyři tisíce termínů ve zhruba dvě stě padesáti hierarchických uspořádáních (soustava nadřazených a podřazených termínů vázících se k určité oblasti nebo tématu). Aby bylo možné začít s konstrukcí nového mezinárodního tezauru, musel být podstatně zredukován velký počet termínů a hierarchií přítomných ve zdrojovém tezeauru HASSET. Skupina odborníků z archivů CESSDA vybrala následující hierarchická uspořádání: ekonomie, práce a zaměstnanost, politika, politické systémy, sociální problémy, diskriminace, postoje, znevýhodněné skupiny, politické instituce, etnické skupiny, životní podmínky, sociální struktura, demografie, sociologie, sociální zabezpečení, environmentální vědy, vzdělání, identita, národnost, rodiny, náboženství, analýza, metodologie. Po této redukci na zmíněné kategorie byl počet termínů určených k překladu zhruba jeden tisíc. Dalšími jednáními mezi archivy se tento počet zvýšil až na čtrnáct set termínů a počet hierarchických uspořádání na dvacet šest. Nicméně tato čísla platila pouze pro počáteční překlad a v dnešní době je již ustanovený a zavedený thesaurus termínově mnohem bohatší.

V praxi existuje více metod, jak vytvářet mnohojazyčný mezinárodní thesaurus. Thesaurus lze vytvořit úplně nově, bez přímých vazeb na existující tezaury, dále je možné přeložit již existující a fungující thesaurus nebo lze slučovat, uspořádat a přeložit více existujících tezaurů. V případě ELSST byla vybrána druhá varianta, tedy přeložení již existujícího tezauru. To s sebou nese celou řadu problémů, vázících se k překladu termínů a jejich uspořádání do hierarchií. Termíny a hierarchie vytvořené v jednom jazyce totiž nemusí být platné a srozumitelné v jiném jazykovém prostředí. Jako příklad může sloužit překlad slova „prime minister“ do ostatních jazyků. V případě například češtiny je odpovídající překlad tohoto slova „ministrský předseda“, který odpovídá anglickému významu slova. Ale již v případě německého prostředí je odpovídající termín „Bundeskanzler“ (v češtině „spolkový kancléř“), což je termín, který je od anglického originálu již velmi vzdálen. Možným řešením takových terminologických problémů pak byla implementace kulturně neutrálních termínů, jako je ve zmíněném příkladu termín „head of government“. Tyto termíny jsou v ELSSTu používány jako preferované, ostatní významy slova jsou případně v tezeauru také uchovány jako synonyma. Existují případy, kdy jednoslovný význam v cílovém jazyce vůbec neexistuje. To je příklad termínu „bezdomovectví“ v angličtině „homelessness“, který nemá jednoslovný ekvivalent ve španělštině. V takovém případě jsou použity nejbližší příbuzné termíny a opět jsou uvedena synonyma. Další problém představuje fakt, že překládaný termín může mít v jednom jazyce širší význam použití, než je tomu v jazycích ostatních. Příkladem může být slovo „drogy“, které v češtině a v některých jiných jazycích (například v němčině „Drogen“) znamená většinou ilegální látky, nikoliv léky, tak jak je tomu v anglickém významu slova „drugs“, které zahrnuje oba významy, tedy jak legální léčiva, tak neoficiální nebo přímo za-

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.