

CESSDA EXPERT WORKSHOP, TAMPERE 2012: AKVIZICE DAT

Martin Vávra (Sociologický ústav)

Sdružení evropských sociálněvědních datových archivů CESSDA uspořádalo na začátku října 2012 ve finském Tampere workshop zaměřený na strategie akvizice dat. CESSDA pořádá workshopy, zaměřené vždy na určité konkrétní téma relevantní pro odborníky z oblasti sociálněvědních datových služeb, každoročně a lze říci, že letos se vrátila na začátek. Na začátek sledu kroků, které patří k práci s datovými soubory, a zároveň na začátek cyklu workshopů, neboť ten úplně první, pořádaný roku 1987 v Dánsku, se zabýval právě získáváním dat do archivu.

Během workshopu se ukázalo, že problémy se získáváním dat jsou, asi celkem pochopitelně, v jednotlivých zemích velmi podobné. Nejdříve je ovšem nutné se o jejich existenci dozvědět. To znamená, nějakým způsobem si udržovat přehled o prováděných výzkumech v sociálních vědách a samozřejmě i výzkumech vznikajících v jiných oblastech, pokud jsou z hlediska sociálních věd relevantní, a především o datech, která jsou jejich výsledkem. Takové hledání je snadnější, pokud existuje databáze, v které jsou údaje o všech z veřejných zdrojů podporovaných výzkumech – jako je v našem případě Centrální evidence projektů (CEP)¹. I v takovém případě je ovšem dobré provádět kontinuálně rešerše odborné časopisecké produkce a využívat sociální sítě. Jen tak je možné se dozvědět například o datech vzniklých před rokem 1993, ze kterého pocházejí nejstarší záznamy v CEP.

Je samozřejmě otázka, o jaká data by měly vůbec archivy usilovat. Na jedné straně jakési pomyslné škály „vědeckého významu dat“ jsou data z velkých mezinárodních šetření, jako je ISSP nebo EVS. Na druhou stranu můžeme postavit různé ankety, mnohé studentské výzkumy, zkrátka velkou množinu výzkumů, u kterých můžeme mít velké pochybnosti o jejich kvalitě a využitelnosti dat z nich pocházejících. Jak se ukázalo během semináře, na tom, kde na této škále by měl archiv stanovit hranici, lze jen obtížně najít shodu. Například zástupce německého Zentral Archivu zdůrazňoval, že archivy by se měly soustředit na získávání a uchovávání dat z výzkumů, o jejichž kvalitě nejsou pochyby, a poukázal právě na výše zmíněná mezinárodní šetření. Oproti tomu záběr českého ČSDA, ale i jiných archivů, je širší a zahrnuje například i výzkumy prováděné na různých specifických populacích, jako jsou studenti či třeba starostové obcí.

I když už o datech víme, zdaleka není vyhráno. V dnešní době elektronické komunikace je sice jednoduché potenciální depozitory oslovit, ovšem zároveň platí, že rozeslání e-mailů či klasických dopisů je většinou málo efektivní. Tvůrci dat buď vůbec neodpoví, nebo archivaci dat odmítnou, anebo zahájí jednání, které ovšem nikam nevede. Pouze menšina případů je úspěšně zakončena archivací a následným zpřístupněním dat.

Ve více příspěvcích na workshopu bylo zmíněno, že existuje něco jako „snadná cesta“ a „těžká cesta“ získávání dat. Onu snadnou cestu představuje většinou získávání dat od kolegů výzkumníků ze spolupracujících institucí – ať už to jsou

1 Dostupná z: <http://www.isvav.cz/prepareProjectForm.do>

univerzity nebo velké výzkumné organizace na poli sociologie. Efekt zde přinášejí osobní vazby na výzkumníky pracující ve stejném oboru, což samozřejmě platí pouze za předpokladu, že pracovníci archivu jsou také sociální vědci, což je ovšem (nejen) v případě ČSDA splněno. Další výhodou představuje to, že archivy mají co těmto lidem a institucím nabídnout. Například školení v práci se systémem Nesstar, ve kterém lze data prohledávat a on-line analyzovat, případně, v případě univerzit, spolupráci na výuce metodologických kurzů. Těžká cesta oproti tomu vede k výzkumníkům a týmům, kteří jsou v nesociologických institucích, ale přesto udělali (třeba zcela jednorázově) výzkum, jehož data jsou pro sociální vědy významná. Zde je potřeba počítat s tím, že proces získávání dat bude dlouhý a bude zahrnovat poskytování informací o konkrétním archivu i o smyslu archivace a zpřístupňování dat pro sekundární analýzu obecně, což jsou znalosti, které sociologové ve většině případů, díky informační aktivitě archivů, již mají.

Tuto zkušenost sdílejí archivy v tak navzájem odlišných zemích, jako je Finsko nebo Irsko. Lze říci, že mezi zeměmi se příliš neliší postupy jednotlivých archivů jako spíše oficiální politiky a to, jak podporují, či nepodporují zpřístupňování dat z výzkumu, a také něco, co bychom mohli nazvat „zvyklostí“ vládnoucí v poli sociálních a humanitních věd. Například pracovníci dánského archivu mají jednoznačnou výhodu v tom, že se mohou při své práci odkázat na normy říkájící, že data, jež jsou výsledkem výzkumu prováděného z veřejných peněz, mají být dostupná. Německý archiv se oproti tomu může opírat pouze o doporučení data po ukončení výzkumu zveřejnit, jež vydaly německé grantové instituce.

A co se týče oněch zvyklostí – jak poznamenala zástupkyně finského archivu, nelze očekávat výrazné zlepšení stavu bez toho, že by bylo poskytování dat pro další využití považováno ve vědecké komunitě za „merit“, za něco, co je stejně tak potřebné a úctyhodné jako například publikování článků ve vědeckých časopisech. Je tedy potřeba podporovat tuto kulturu vzájemného sdílení dat. Mnohokrát na workshopu také zaznělo slovo důvěra – nestačí formální dohody mezi depozitory a archivem na straně jedné a mezi archivem a uživateli na straně druhé, je potřeba, aby archiv byl vnímán jako důvěryhodný. To může zahrnovat získání nějakého formálního certifikátu, jako je například Data Seal of Approval, ale především to znamená, že díky své aktivitě ve vztahu k depozitorům i uživatelům dokáže tuto důvěru získávat a udržovat.

Závěry

Jak tedy situaci zlepšit? Rozhodně je dobré podporovat, nebo dokonce iniciovat formální opatření, která stanoví zpřístupňování dat vzniklých z veřejných prostředků jako povinnost. Zároveň ovšem archivy musí být aktivní, to znamená dávat o sobě vědět, komunikovat s potenciálními depozitory, vyhledávat data a vytvářet nástroje, které zabezpečí, že předávání dat proběhne bez problémů a efektivně – například různé webové formuláře, které umožní uploadovat meta-data i datové soubory. To asi nejdůležitější – tedy to, že poskytování dat by mělo být chápáno jako něco záslužného – je ovšem zároveň to nejobtížnější. Samotné archivy zde těžko způsobí nějakou rychlou změnu, je to otázka vývoje celého vědeckého pole. Mohou ovšem svou činností k takové změně přispívat.