

Editorial

Vážené čtenářky, vážení čtenáři,

prosincové číslo Sociowebu se věnuje tématu zdrojů a managementu sociálněvědních dat. První článek od Jindřicha Krejčího *uvádí čtenáře do problematiky managementu dat v sociálněvědním výzkumu*. Autor v něm pojednává o tom, jaký význam má spravování datových souborů a proč je vůbec potřeba se věnovat managementu dat systematicky v průběhu celého výzkumného projektu.

Druhý článek od Martina Vávry se věnuje problematice srovnatelnosti výzkumných dat a zejména tzv. principu ekvivalence, který předpokládá, že všem respondentům ve výzkumu se dostává stejného podnětu. Článek se zabývá především komparací dat v mezinárodních šetřeních a problémy spojenými s tím, že stejné znění otázky může být respondenty z různých kultur chápáno odlišně.

Ve třetím článku se Tomáš Čížek věnuje ochraně osobních údajů v sociálněvědním výzkumu. Autor se v příspěvku zabývá problémy možného zneužití údajů o respondentech, etickými kodexy jednotlivých národních a mezinárodních organizací sdružujících výzkumníky v sociálních vědách, a také standardy pro nakládání s osobními údaji.

Další dva články jsou zaměřené na popis konkrétních příkladů zdrojů dat pro sociálněvědní výzkum. Článek Pata Lyonse a Johany Chytilové se věnuje zdrojům dat v oblasti sociologie politiky. Tento stručný příspěvek poskytuje přehled o panelových šetřeních politických témat v ČR.

Poslední článek od editorky tohoto čísla je věnován zdrojům dat o imigrantech. Zde je rozebírán příklad dvou velkých českých národních šetření realizovaných v rámci mezinárodních srovnávacích projektů Labour Force Survey (LFS) a European Union Statistics on Income and Living Conditions (EU-SILC). Zmíněná rozsáhlá kontinuální šetření zahrnující desítky tisíc respondentů jsou analyzována z hlediska toho, jak dobře pokrývají populaci imigrantů v ČR.

Yana Leontiyeva

Yana.Leontiyeva@soc.cas.cz

Toto číslo Sociowebu vzniklo v rámci grantového projektu COMPDAT – Zdroje dat, výzkum standardů, kvality dat a metody harmonizace dat pro mezinárodní sociální komparativní výzkum a integraci do sítě CESSDA sponzorovaného MŠMT v programu INGO pod reg. č. LA09010.

« Teorie pro všechny »

»

Management sociálněvědních dat a cyklus života dat

Klíčová slova: . metodologie, management dat, sociálněvědní výzkum

Empirický výzkum se bez správy datových souborů sice neobejde, nebývá ovšem zvykem této problematice věnovat samostatnou pozornost. Správa dat probíhá v podobě jednotlivých, navzájem izolovaných úloh v rámci různých výzkumných fází. To ale může být problém, existuje totiž celá řada dobrých důvodů, proč se managementu výzkumných dat, věnovat systematicky a průběžně.

Úroveň správy dat má dopady na kvalitu dat, a ta je důležitým předpokladem relevance výsledných zjištění. Toto platí i obecněji, při vytváření funkčního prostředí pro výzkumnou praxi. Systematický management dat je především předpokladem prevence chyb a mylných výsledků, dokáže ale také uspořit spoustu času a zpřehlednit a zjednodušit výzkumnou práci. Mezi výzkumníky koluje bezpočet historek o trapných omylech způsobených použitím nesprávných dat, mnoho analýz zkrachovalo na nedostacích v datových souborech a mnohý badatel strávil dlouhé hodiny luštěním špatně dokumentovaných dat. Mnoho výzkumných informací se díky nepořádkům v datových souborech poztrácí.

Vyšší nároky na management dat, než při samostatné badatelské práci, jsou spojeny se záměry dlouhodobého uchování dat a sdílení dat mezi různými výzkumnými týmy. Právě požadavky na archivaci a otevřený přístup k sociálněvědním datům pro účely jejich sdílení a na druhé straně také využívání sekundární analýzy ve výzkumu ovšem získávají v sociálních vědách stále větší význam. S tím pak zákonitě

rostou i požadavky na profesionalitu správy dat ve výzkumném projektu. Tato větší náročnost však není samoučelná, ale přímo vyplývá z vývoje sociálněvědního empirického výzkumu a jeho metod.

V první řadě již dlouho a rychle roste objem dat, které sociálněvědní výzkum využívá. Nejen, že se zvyšuje počet a rozsah sociálněvědních výběrových šetření, ale ve společnosti všeobecně dochází k překotné digitalizaci a produkci dat využitelných pro analýzu. Zároveň dochází k rozvoji a zavádění nových informačních technologií umožňujících nové způsoby práce s daty, jejich přenos, propojování a také sdílení. Charakter současné výzkumné práce je také hodně ovlivňován rozšířením vědecké spolupráce, která bývá často založena na sdílení dat mezi různými výzkumnými týmy. V podstatě na této praxi stojí např. současný rozvoj mezinárodního komparativního výzkumu. Objem a dostupnost datových zdrojů a nové typy dat pak mají značný dopad na rozvoj metod analýzy a jejich zavádění zpětně generuje další požadavky na vytváření dat a podobu jejich zpracování.

Větší objem dat logicky vyžaduje systematictější přístup při jejich produkci i používání. Správa sociálněvědních datových zdrojů se v řadě aspektů nyní podobá situaci mnohých oborů přírodovědného a technického zaměření založených na zpracovávání velkých objemů digitálních informací. Spolu s aplikací informačních technologií pak dochází k přebírání některých pracovních postupů a standardů vyvíjených v oblasti informačních věd. Aby bylo možné data sdílet a využívat pro sekundární analýzu je však jednak potřeba databáze při jejich vytváření ošetřit a zdokumentovat tak, aby byly srozumitelné i mimo autorský výzkumný tým, a za druhé při realizaci výzkumných projektů, ať už data vytváří nebo ne, je nutné systematicky pracovat s širokým okruhem dostupných datových zdrojů.

Požadavky na sdílení dat mají dopad na proměnu funkcí managementu dat. Při vytváření datového souboru je potřeba počítat s jeho archivací a zveřejněním a dopředu není jasné, kdo všechno, jak a k čemu bude data používat. To se pochopitelně odráží ve způsobu přípravy a ošetření datového souboru a v nárocích na jeho dokumentaci. Dopad je však hlubší, projevuje se nejen v jednotlivých úlohách, ale v celé koncepci správy dat.

Sociálněvědní výzkum, podobně jako jiné vědy, má podobu cyklu, kdy výsledky jednoho výzkumu se vrací zpět do výzkumného procesu jako východiska pro nové výzkumné úlohy. V prostředí založeném na sdílení datových zdrojů, jako se tomu děje v akademickém sociálněvědním výzkumu, v tomto cyklu hraje významnou roli právě opětovné použití výzkumných dat [viz např. Humprey 2006]. Data tak získávají nový význam pro přenos a reprodukci znalostí a tyto jejich funkce je potřeba zohlednit ve způsobu jejich správy. Vhodná východiska pro to poskytují tzv. koncepty „cyklu života dat“.

Správa digitálních informací je v tomto případě inkorporována do cyklického systému vědeckého poznávání. Příkladem takového pojetí je model životního cyklu výzkumu definovaný Charlesem Humpreym [ibid.], který je zobrazen v grafu 1, a jeho aplikace na správu dat v grafu 2.

Management životního cyklu dat pak představuje komplexní metodiku založenou na modelu, který zachycuje průběh využití dat jako celek strukturovaný do fází s různými cíli, úlohami a aktéry, ve kterém jsou vztahy mezi jednotlivými prvky určeny životním cyklem vědeckého poznávání. Takové přístupy aplikují např. koncepce a metodiky datového managementu vypracované britským archivem UK Data Archive [Eynden, Corti et al. 2011, Eynden, Bishop et al. 2010] nebo americkým konsorciem ICPSR [2009]. Fáze datového managementu z modelu, který používá výše zmíněné konsorcium ICPSR pro depozitory a uživatele služeb svého datového archivu, jsou zobrazeny v tabulce 1. Výzkumný projekt zde začíná rešerší existujících dat a management dat pak prochází celým výzkumným procesem.

Literatura:

Eynden, Veerle Van den, Libby Bishop, Laurence Horton, Louise Corti. 2010. *Data Management Practices in the Social Sciences*. Colchester: UK Data Archive (UKDA). On-line: <http://www.data-archive.ac.uk/media/203597/datamanagement-socialsciences.pdf> (navštíveno 9. 11. 2011).

Eynden, Veerle Van den, Louise Corti, Matthew Woolard, Libby Bishop, Laurence Horton. 2011. *Managing and Sharing Data*. Colchester: UK Data Archive (UKDA). On-line: <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf> (navštíveno 9. 11. 2011).

ICPSR. 2009. *Guide to Social Science Data Preparation and Archiving. Best Practice Throughout the Data Life Cycle*. 4-th Edition. Ann Arbor: Inter-university Consortium for Political and Social Research (ICPSR), University of Michigan. On-line: <http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf> (navštíveno 9. 11. 2011).

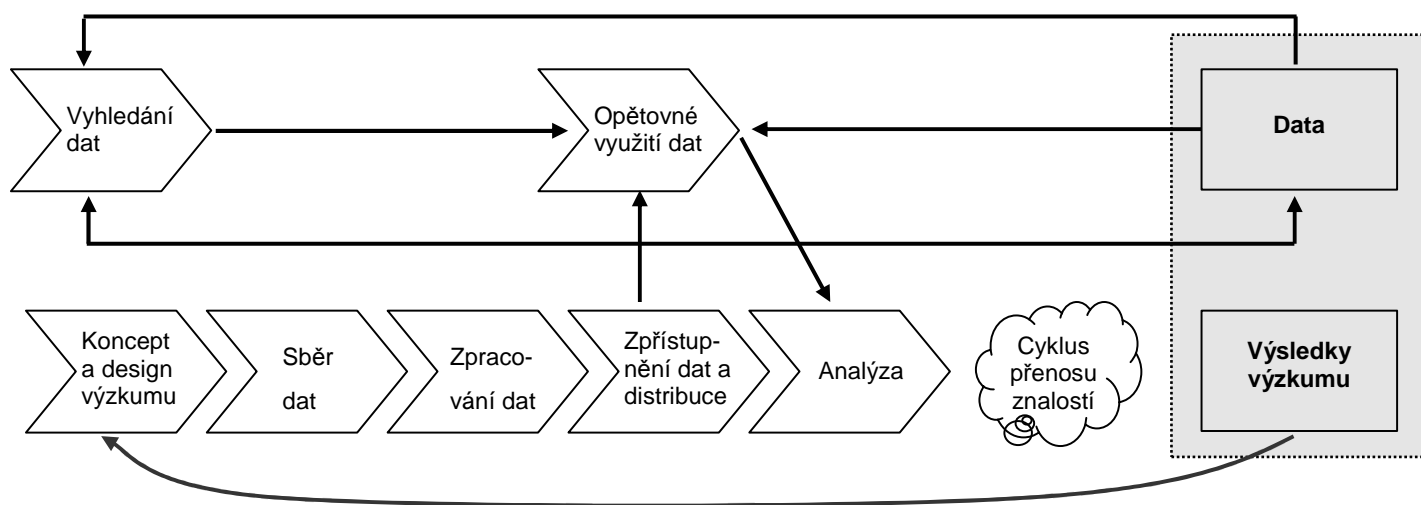
Humprey, Charles 2006. e-Science and the Life Cycle of Research. Nepublikovaný elektronický dokument. On-line: <http://datalib.library.ualberta.ca/~humphrey/life-cycle-science060308.doc> (navštíveno 9. 11. 2011).

Ionescu, Sanda, Mary Vardigan. 2008. "Putting DDI 3.0 to Work for You! Part I." Presented at the annual meeting of the International Association of Social Science Information Service and Technology (IASSIST), Palo Alto, USA, May 2008. On-line: <http://www.ddialliance.org/node/248> (navštíveno 9. 11. 2011).

Jindřich Krejčí

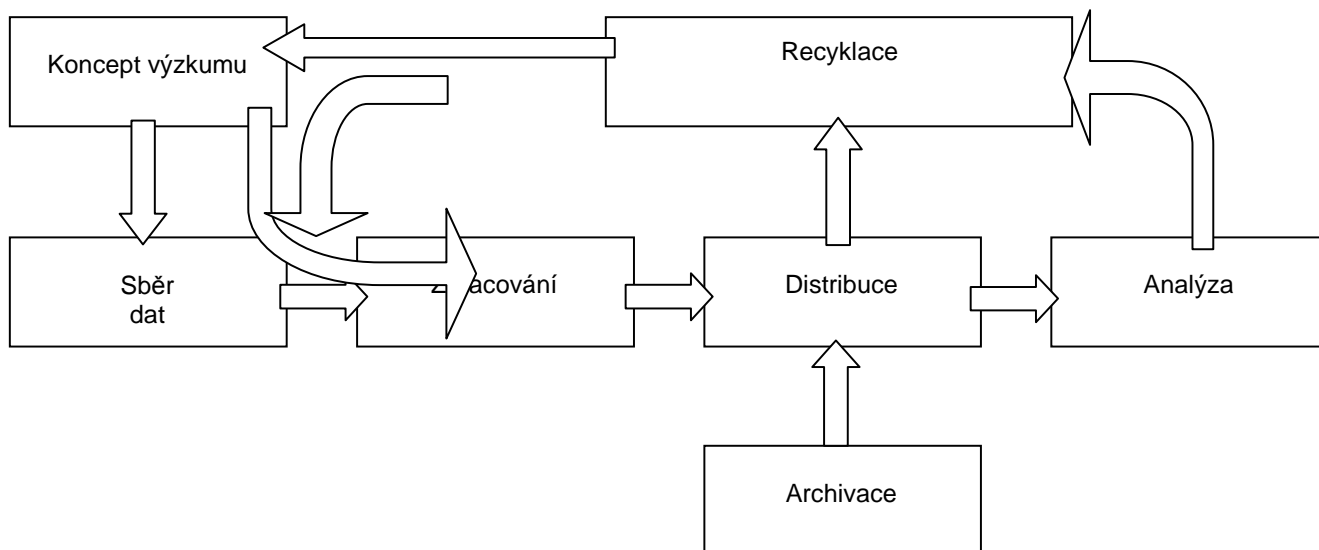
Jindrich.Krejci@soc.cas.cz

Graf 1. Výzkumná data a model životního cyklu poznávání



Zdroj: Humprey [2006].

Graf 2. Životní cyklus dat



Zdroj: Ionescu, Vardigan [2008].

Tabulka 1. Fáze managementu sociálněvědních dat podle americké organizace ICPSR

Fáze 1	Plánování a psaní projektu	<ul style="list-style-type: none"> - Rešerše existujících dat a metadat. - Rozhodnutí, zda projekt vytvoří nový datový soubor. - Řešení specifických otázek archivace (informovaný souhlas respondentů, ochrana osobních údajů, autorská práva atd.). - Identifikace potenciálních uživatelů dat. - Odhad nákladů archivace.
Fáze 2	Zahájení projektu a management dat	<ul style="list-style-type: none"> - Vytvoření plánu managementu dat. - Rozhodnutí o formě dokumentace a jejím obsahu. - Provedení pretestu a pilotáže materiálů a metod.
Fáze 3	Sběr dat a vytvoření datového souboru	<ul style="list-style-type: none"> - Respektování pravidel nejlepší praxe šetření z hlediska vytváření datového souboru a shromažďování podkladů pro dokumentaci. - Vytváření datového souboru, zajištění integrity dat, vytvoření struktury souboru, definice proměnných, pojmenování proměnných a hodnot, seskupování proměnných, kódování, ošetření chybějících hodnot, transformace, zajištění ochrany osobních údajů respondentů. - Dokumentace dat, nastavení dokumentačního standardu, technické zajištění dokumentace.
Fáze 4	Analýza dat	<ul style="list-style-type: none"> - Management zdrojového datového souboru, pracovních a jiných odvozených souborů; verze a edice datového souboru. - Transformace dat, vytváření odvozených proměnných. - Práce se strukturou datového souboru. - Zálohování datových souborů.
Fáze 5	Příprava dat pro sdílení	<ul style="list-style-type: none"> - Zajištění ochrany osobních údajů pro účely sdílení. - Určení formátů pro depozici dat. - Určení uživatelských práv. - Zajištění požadavků archivu na depozici..
Fáze 6	Depozice dat	<ul style="list-style-type: none"> - Zajištění vhodného formátu dat a metadat. - Realizace depozice.
Fáze 7	Archivace	<ul style="list-style-type: none"> - Vyhodnocení datového souboru. - Dodatečné prozkoumání ochrany osobních údajů. - Zpracování dat pro archivaci, prozkoumání kvality dat. - Popis dat. Vytvoření metadat v požadovaném formátu. - Příprava dat pro on-line analýzu, vylepšení datového souboru. - Zabezpečení dat a metadat. - Vytvoření přístupu k datům. - Vytvoření východisek pro uživatelskou podporu.

Zdroj: ICPSR [2009].

»

Princip ekvivalence v mezinárodních sociologických výzkumech

Klíčová slova: metodologie, management dat, harmonizace dat

V loňském prosincovém čísle Sociowebu se Jindřich Krejčí a Yana Leontiyeva věnovali mezinárodnímu komparativnímu výzkumu v sociologii, respektive harmonizaci dat v tomto typu výzkumů. Tento článek na téma navazuje. Popisují zde *princip ekvivalence*. Tento princip, zjednodušeně řečeno, vyjadřuje to, že všem respondentům se dostává v sociologickém

dotazníkovém šetření stejného podnětu. [1] Bez toho by byla srovnatelnost údajů od jednotlivých respondentů v různých zemích a kulturách nemyslitelná. Tento princip se nevztahuje pouze k mezinárodnímu výzkumu. Platí ovšem, že právě ve výzkumech, na kterých participuje více zemí, vyvstávají problémy s dosažením ekvivalence nejsilněji. Johnson (1998: 3-6) uvádí 52 konkrétních významů pojmu ekvivalence v sociálněvědním výzkumu. V zásadě lze ovšem podle Johnsona rozlišit dva základní typy ekvivalence. První nazývá interpretativní ekvivalence, druhý pak procedurální ekvivalence. Zatímco interpretativní ekvivalence se vztahuje k subjektivně pojaté srovnatelnosti významů (otázek, škál atd.), procedurální ekvivalence odkazuje k „objektivní“ srovnatelnosti položek v dotazníku či metrik jednotlivých škál mezi jednotlivými kulturními skupinami (Johnson 1998: 30). Dokonalá interpretativní ekvivalence v dotazníkové šetření

předpokládá, že všichni respondenti chápou podněty výzkumu stejně (stejně rozumí otázkám, stejně chápou a také používají škály, na nichž vybírají odpověď). Její ověřování je vždy obtížné – právě proto, že se zaměřuje na subjektivní porozumění. Ověření toho, zda respondenti rozumí stejně podnětům v dotazníku, je trochu jiný problém, než vymýšlení různých postupů, klasifikací, standardů a převodníků, které mají zajistit procedurální ekvivalenci.

Právě „technicky“ pojatá procedurální ekvivalence má k harmonizaci dat (která je formalistická a na pravidlech založená ze své podstaty) velmi blízko. Nyní už by mělo být zřejmé, že harmonizace nemusí vždy přispívat interpretativní ekvivalenci. Jde například o situaci, kdy je stejné znění otázky respondenty z různých kultur pochopeno odlišně. Položky dotazníku mají v jednotlivých národních verzích stejné škály, doslovný překlad je proveden kvalitně, jak je to jen možné, a jeho správnost „potvrzena“, například zpětným překladem do původního jazyka, který provede jiný překladatel, neseznámený s originální podobou dotazníku. A přesto může vzniknout problém. Uvedme si konkrétní příklad. V roce 1998 byla ve výzkumu ISSP (kterého se každoročně zúčastňuje několik desítek zemí, včetně například Japonska) zaměřeném na téma náboženství zařazena otázka: „Který z těchto výroků je nejbližší Vašemu názoru na bibli?“. [2] Následovalo několik výroků vyjadřujících určitý postoj. Přestože by nepochybně bylo možné vytvořit přesný překlad této otázky a nabízených odpovědí do japonštiny, japonský tým se rozhodl otázku vyřadit, a to s odůvodněním, že „v Japonsku nejrozšířenější náboženství, šintoismus a buddhismus nemají texty, které by byly ekvivalentní křesťanské Bibli“. Zde je zřejmé, že zařazení této otázky by pro drtivou většinu japonských respondentů nebylo relevantní, neboť nemají vztah ke křesťanské kultuře a konstrukce interpretačně ekvivalentní otázky relevantní pro japonskou společnost nebyla možná. Vezměme si ale jinou otázku ze stejného výzkumu, která zněla: „Věříte v peklo?“. [3] Podle dat z ISSP 1998 36,7 % Japonců zcela nebo spíše v peklo věří. Tento výsledek je jistě možné srovnávat s údaji za další země zařazené v tomto výzkumu – není žádný důvod se domnívat, že došlo k nějaké chybě při procesu harmonizace, [4] která by takové srovnání znemožňovala. Do jaké míry bylo dosaženo interpretativní ekvivalence je ovšem již složitější otázka a aby na ní bylo možno odpovědět, je potřeba posoudit odpovědi japonského vzorku na další otázky a především získat podrobnější informace o japonské náboženské scéně. Obecně je možno říci, že interpretativní ekvivalence je ohrožena vždy, když jsou velmi diferencované a kulturně specifické jevy převáděny na nějaký „společný rámec“ – tedy harmonizovány. To neznamená kritiku harmonizace – bez ní by nebylo možné srovnání vůbec provádět. Jde spíše o výzvu k opatrnosti při interpretování výsledků.

Literatura:

Johnson, T., 1998. „Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research“. *ZUMA-Nachrichten spezial*:1-39. Dostupné na http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial3.pdf

[1] Pojem ekvivalence se objevil vedle klasických otázek reliability a validity a je s nimi také v mnoha ohledech provázán.

[2] „Which of these statements comes closest to describing you feelings about the Bible?“ v originálním dotazníku, z něhož se pořizovaly jednotlivé jazykové varianty.

[3] „Do you believe in hell?“ v originálním dotazníku.

[4] Pokud k takové chybě (může to být například chyba v překladu otázky nebo použití odlišné škály pro odpovědi) dojde, představuje dobrou praxi při archivaci a distribuci dat všechny takové chyby zaznamenat v rámci metadat, nebo v metadatech na nějaký dostupný soubor s těmito odchylkami odkázat. Pokud tým, který sestavuje spojený mezinárodní datový soubor, dojde k závěru, že chyba ohrožuje srovnatelnost dat, pak by měl příslušnou proměnnou nebo dokonce celý národní soubor z mezinárodních dat vyřadit. Je nutno konstatovat, že v rámci výzkumů, které zde zmiňujeme (ISSP, EVS, ESS), se toto vše děje.

Martin Vávra

martin.vavra@soc.cas.cz

»

Ochrana osobních údajů respondentů v sociálně-vědním výzkumu

Klíčová slova: metodologie, management dat, ochrana osobních údajů

Sociálně-vědní výzkum může být, prakticky samozřejmě jako jakákoliv jiná lidská aktivita, dobrý sluha, ale špatný pán. V historii se objevuje mnoho příkladů úmyslného i neúmyslného zneužití výzkumných aktivit z hlediska státních struktur i samotných výzkumníků. Pro vládnoucí vrstvu je za každého společenského uspořádání nutné vědět, co si myslí veřejnost, informace o veřejném mínění pak mohou být využívány k společensky pozitivním činům, ale také jako prostředek k ovládnutí veřejnosti nebo k její manipulaci. Pro společenského vědce samotného je pak samo provádění sociálně-vědního výzkumu také velmi citlivou záležitostí značně závislou právě na politické situaci v dané společnosti.

Zneužití údajů získaných sociálně-vědním výzkumem může být různé, obecně se jedná o využití údajů k účelům, které respondentovi nebyly předem známy, případně může dojít k jeho nechtěné identifikaci a spojení názorů a osobních údajů s konkrétní a vyhledatelnou osobou. Použitím takových informací dochází k narušení práva člověka na soukromí nebo může dojít přímo k využití údajů proti zájmům respondenta. Je samozřejmě nutné dodat, že k takovým zneužitím v sociálních vědách dochází pouze v mízivém procentu případů a v drtivé většině případů je to omylem, nedopatřením a neúmyslně. Cílem je samozřejmě, aby k nim nedocházelo vůbec. Právě kvůli možnému zneužití údajů o respondentech i možnému politickému zneužití s sebou nese výzkum v oblasti sociálních věd velké nároky na dodržování etických standardů. Tyto nároky bývají formálně popsány jak v učebnicích sociálního výzkumu, tak v kodexech jednotlivých národních a mezinárodních organizací sdružujících výzkumníky v sociálních vědách. Například organizace ISA (mezinárodní sociologická asociace) ve svém etickém kodexu uvedeném přímo na hlavní webové stránce o etice výzkumu sběru dat říká následující [1]:

2.3.1. Sociologové by měli otevřeně sdílet metody výzkumných postupů a zdroje svých dat tak, jak je tomu ve vědě obvyklé.

2.3.2. „Bezpečnost, anonymita a soukromí subjektů výzkumu a informátorů by měly být přísně respektovány, jak v kvantitativním, tak v kvalitativním výzkumu. Výzkumníky získané zdroje osobních informací by měly být důvěrné, pokud si sám informátor nepřeje a nesouhlasí být citován. Pokud jsou informátoři snadno identifikovatelní, výzkumník je povinen poučit je o možných následcích publikování dat a závěrů výzkumu.

2.3.3. Sociologové, kteří pracují s osobními daty musí zachovávat podmínky diskrétnosti, za kterých byla data sebrána. Mohou ovšem používat data shromážděná v historických archivech, soukromých i veřejných za podmínky dodržení právních podmínek daných zemí a podmínek akceptovaných mezinárodní vědeckou komunitou a také podmínek archivů.

2.3.4. Souhlas subjektů výzkumu a informátorů by měl být získán předem. Skrytý výzkum by měl být vyloučen v principu, možný je pouze v případech, pokud požadované informace nebo přístup k takovým zdrojům informací jsou blokovány držiteli moci v dané společnosti.

Dodržení etických standardů je však pouze jednou z povinností sociálně-vědního výzkumníka. Prakticky po celém světě jsou nějakým způsobem omezována práva výzkumníků sbírat osobní údaje o respondentech přímo ze zákona. Nejvíce rozvinutá je tato ochrana v členských státech Evropské unie. Zákony na ochranu osobních údajů musely být v členských zemích Evropské unie přijaty na základě směrnice číslo 95/46/EC přijaté v říjnu roku 1995. Směrnice byla přijata, aby chránila základní práva a svobody, zejména právo na soukromí. V České republice byla tato směrnice

základem pro zpracování zákona č. 101/2000 Sb. na ochranu osobních údajů, který byl přijat v roce 2000 v souvislosti s probíhajícím přístupovým řízením do Evropské unie. Tento zákon, od roku 2000 vícekrát novelizovaný, se jednoznačně dotýká i sociálně-vědního výzkumu a samozřejmě i problémů spojených s přístupem k takovému typu dat. Zpracováním dat se podle zákona rozumí: „*jakákoliv operace nebo soustava operací, které správce nebo zpracovatel systematicky provádějí s osobními údaji, a to automatizované nebo jinými prostředky. Zpracováním osobních údajů se rozumí zejména shromažďování, ukládání na nosiče informací, zpřístupňování nebo jinými pozměňování, vyhledávání, používání, předávání, šíření, zveřejňování, uchovávání, výměna, třídění nebo kombinování, blokování a likvidace*“. Zákon vymezuje osobní údaje a citlivé osobní údaje způsobem, jak to ukazuje následující citace přímo z pasáže zákona.

Definice osobních údajů podle zákona 101/2000 Sb.

Pro účely tohoto zákona se rozumí

a) osobním údajem jakákoliv informace týkající se určeného nebo určitého subjektu údajů. Subjekt údajů se považuje za určený nebo určitelný, jestliže lze subjekt údajů přímo či nepřímo identifikovat zejména na základě čísla, kódu nebo jednoho či více prvků, specifických pro jeho fyzickou, fyziologickou, psychickou, ekonomickou, kulturní nebo sociální identitu,

b) citlivým údajem osobní údaj vypovídající o národnostním, rasovém nebo etnickém původu, politických postojích, členství v odborových organizacích, náboženství a filozofickém přesvědčení, odsouzení za trestný čin, zdravotním stavu a sexuálním životě subjektu údajů a jakýkoliv biometrický nebo genetický údaj subjektu údajů,

zdroj www.uoou.cz, sekce právní předpisy

Pohledem na výše uvedené definice lze zjistit, že se zákon vztahuje prakticky na všechny výzkumné aktivity v oblasti sociálních věd. Dokonce lze říci, že sociální věd se přímo týká nakládání s citlivými údaji. Pro výzkumníky z toho vyplývá nutnost splnění formálních povinností. Je nutné si při každém výzkumném rozhovoru vyžádat souhlas informátora (dle zákona subjektu údajů), v případě citlivých údajů (při výzkumu tedy téměř vždy) musí tento souhlas být písemný. Ačkoliv se zákon může zdát přísný, vlastně se celkem shoduje s tím, co po výzkumnících požadují i etické kodexy, kterými se mají při výzkumu řídit. Pokud chce výzkumník sbírat osobní údaje, musí se registrovat na Úřadu pro ochranu osobních údajů. Je také nutno data spolehlivě ochránit před zneužitím, a to především jejich bezpečným uložením s možností přístupu pouze pro oprávněné osoby. Formálně se splnění těchto povinností řeší použitím tzv. informovaného souhlasu, což je jistý typ smlouvy, která tyto problémy řeší. Existuje mnoho různých forem, jak by měl vypadat, a neexistuje nějaký univerzální prototyp této smlouvy. Vždy by měla

být sestavena na základě konkrétního výzkumného projektu. Například podle ESDS (Economic and Social Data Service – významný datový archiv ve Velké Británii) existují tyto základní body, které musí informovaný souhlas splňovat:

Souhlas musí být poskytován svobodně s dostatečným zdůrazněním toho, k čemu je poskytován.

Mezi oběma stranami výzkumu musí existovat aktivní komunikace - co je od účastníků výzkumu požadováno a proč je jejich účast vyžadována.

Je potřeba rozlišovat mezi souhlasem k samotnému výzkumu a souhlasem ke sdílení poznatků třetími stranami, nebo k publikaci.

Musí být podrobně popsáno, jak bude s osobními údaji nakládáno, jak budou data zabezpečena, jakým způsobem budou data anonymizována a kdo k nim bude mít přístup.

Mlčení neznámá souhlas - souhlas nelze odvozovat z neposkytnuté odpovědi na dopis, nebo pozvánku k účasti na výzkumu

Pro výzkumníka je dodržení všech těchto náležitostí důležité. Lze se sice vyhnout potížím se získáváním souhlasu při okamžité anonymizaci (zbavení dat veškerých údajů, které by mohly vést k identifikaci respondenta), nebo by se osobní údaje vůbec nesbíraly, respondent by však pak byl pro výzkumníka naprosto neznámým člověkem a je otázkou, nakolik je vůbec taková výzkumná situace reálná.

Literatura:

Bitrich, Tomáš 2001 Diskuze k zákonu o ochraně osobních údajů v kvalitativním výzkumu. *Biograf* č.25, 121-122.

Čížek, Tomáš 2006 Archivace a sdílení kvalitativních dat. *SDA-Info*. Roč. 8, č. 1, s. 13-15.

Tyl, Tomáš. 2004: Některé právní aspekty kvalitativního výzkumu. *Biograf* (35): 31 odst. Dostupné na adrese <http://www.biograf.org/clanky/clanek.php?clanek=3513>

Zákon č. 101/2000 Sb. O ochraně osobních údajů

[1] volný překlad části textu, originál je dostupný na adrese: http://www.ucm.es/info/isa/about/isa_code_of_ethics.htm

Tomáš Čížek

Tomas.Cizek@soc.cas.cz

»

Přehled panelových šetření politických

Klíčová slova: metodologie, politika, zdroje dat

Většina výběrových šetření v sociologii je průřezová, tj. respondenti jsou v rámci šetření dotazováni pouze jednou. Tyto jednorázové výzkumy poskytují obraz společnosti v jednom bodě v čase a neumožňují přímé studium procesů sociálních a politických změn. K tomu, abychom z reprezentativního vzorku dospělé populace mohli zjišťovat změny na individuální úrovni, musíme respondenty vyzpovídat nejméně dvakrát, což je základní logika panelových šetření. Data z průřezových výzkumů mohou být použita k nepřímému vyvozování postojových změn, ale jsou k těmto účelům méně vhodná než data z panelových šetření (Vinopal 2009; Wooldridge 2001). Je důležité zmínit, že v České republice jsou realizovány významné panelové výzkumy jako EU - SILC nebo SHARE, jejichž výstupy jsou užitečné pro studium některých politických témat, nicméně jejich hlavním zaměřením je výzkum sociálních nerovností a stárnutí, nikoliv politiky. [1]

Výlučně politická témata nebyla prostřednictvím panelu v České republice zkoumána často. V roce 1996 vznikl panelový výzkum pro účely studia senátních voleb, který umožňoval studium voličské účasti, výběru kandidáta a porovnání postojů před a po volbách. O tomto výzkumu je pojednáno v článku o zdrojích dat ze senátních voleb v tomto čísle Socioweby.

Dalším panelovým šetřením politických témat byl výzkum o čtyřech vlnách ve městě Klatovy (obec s 23 tisíci obyvateli v Plzeňském kraji), provedený v období od září 1999 do listopadu 2000. Tato unikátní, Lazarsfeldem inspirovaná (1944: 155) studie hledala spojitost komunikace mezi jednotlivci s formováním politických postojů (model dvoustupňového toku informací). Výzkum, vedený Prof. PhDr. Hynkem Jeřábkem CSc. (ISS FSV UK), začínal na rozlehlém souboru dvou tisíc respondentů, který se později v důsledku důvěrně známého efektu úbytku respondentů v panelu, zmenšil na pět set zkoumaných jednotek. Výzkumníci se zde zaměřili na stabilitu politických postojů a hodnot místních obyvatel a použili standardní nástroje k měření levo-pravé politické orientace (viz Jeřábek 1999; Schubert 2010). Data v současné době nejsou dostupná pro vědecké a studijní účely.

Ambicióznějším panelovým výzkumem byl projekt, zaměřený na politické postoje a utváření mediální agendy, který realizovalo Centrum pro výzkum veřejného mínění (CVVM) během dvanácti týdnů v období od dubna do července 2008. Výběrový soubor sestával z 650 respondentů, kteří na adresu agentury každý týden poštou odesílali vyplněný dotazník. Sběr dat v panelových výzkumech prostřednictvím pošty není v současné době úplně běžný, neboť většina výzkumů používá ke sběru dat internet. V analýze změn náklonnosti k politickým

stranám (identifikace s politickou stranou) se ukázalo, že sociálně-psychologická či sociálně-identitní základna pro stabilní podporu stran, která je podle mnoha vědců považována za klíčový prvek stabilní demokracie, je v České republice silná. I tak je však jednotlivců, pociťujících k politické straně vazby, v celém elektorátu menšina (Linek a Lyons 2009). Rovněž tato data dosud nebyla archivována v Českém sociálně-vědním datovém archivu (ČSDA).

Panelová šetření politických témat si zatím v České republice nezískala velkou popularitu, což je částečně způsobeno i jejich vysokou organizační a finanční náročností. Výzkumníci, zabývající se trendy v politických postojích v ČR, si tak zatím musí vystačit s výstupy těch několika málo panelových šetření, která na našem území dosud vznikla.

Literatura:

Jeřábek, H. (ed.). 1999. Utváření postojů obyvatel českého města. Pracovní texty / Working papers 99:7. Praha: Sociologický ústav AV ČR.

Lazarsfeld, P., B. Berelson, and H. Gaudet. 1944. *The People's Choice*. New York: Columbia University Press.

Linek, Lukáš, Pat Lyons. 2009. „Stabilní postoj nebo konzistentní odpovídání? Prozkoumání stability stranické identifikace v ČR“ In Markéta Škodová, Vlastimil Nečas (eds.): *Veřejná a mediální agenda: komparativní analýza tematizace veřejné sféry*, Praha: Professional Publishing, str. 98–123. ISBN 978-80-7431-009-6.

Schubert, J. 2010. Klasická testová teorie reliability v metodologii výběrových šetření. *Data a výzkum - SDA Info*, 4(2): 105-122.

Vinopal, Jiří. 2009. „Konstrukce panelu respondentů a datový soubor.“ Pp. 124-130 in M. Škodová, V. Nečas (eds.). *Veřejná a mediální agenda: komparativní analýza tematizace veřejné sféry*. Praha: Professional Publishing.

Wooldridge, J.M. 2001. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

[1] Více informací o těchto panelových šetřeních naleznete čtenáři na následujících webových stránkách: European Union Statistics on Income and Living Conditions (EU-SILC) - http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc

Survey on Health and Aging in Europe (SHARE) - <http://www.share-project.org/>

Pat Lyons

pat.lyons@soc.cas.cz

Johana Chyliková

johana.chylikova@soc.cas.cz

»

Údaje o imigrantech ve výzkumech LFS a EU-SILC

Klíčová slova: metodologie, migrace, zdroje dat

Oficiální statistiky, ač jsou jakkoliv spolehlivé, nemohou poskytnout vyčerpávající údaje o imigrantech a jejich integraci. Proto přicházejí na řadu výzkumy, které podchycují nejen ve statistikách chybějící nebo neúplné demografické a jiné důležité ukazatele (například vzdělání a profese imigrantů v ČR, jejich rodinný status, počet dětí apod.), ale také další významné aspekty života imigrantů.

Vzhledem k častým národním odlišnostem a omezením v administrativních zdrojích dat jsou v některých zemích považována za vhodný zdroj širších informací o migrantech velká výběrová šetření s několikatisícovým vzorkem respondentů. Ta jsou organizována národními statistickými úřady v rámci velkých mezinárodních projektů. V tomto krátkém příspěvku se budeme věnovat dvěma největším šetřením realizovaným v rámci mezinárodních projektů Labour Force Survey (LFS) a European Union Statistics on Income and Living Conditions (EU-SILC).

Mezinárodní šetření Labour Force Survey je v České republice harmonizováno s Výběrovým šetřením pracovních sil (VŠPS), realizovaným na národní úrovni od konce roku 1992. Připomeňme, že se jedná o rozsáhlé čtvrtletní šetření všech členů domácností (starších 15 let) v bytech vybraných náhodným (v ČR dvoustupňovým stratifikovaným) výběrem. Každé čtvrtletí se ve výběrovém vzorku ocitne přibližně 25 tis. domácností. Toto čtvrtletní šetření zjišťuje celou škálu informací o ekonomických aktivitách dospělé populace a je prováděno ve všech členských zemích EU a v několika dalších zemích.

Podle požadavků evropských institucí musí LFS jako jedno z nejdůležitějších šetření zaměřených na zaměstnanost sloužit jako základní zdroj informací o integraci cizinců, zejména co se týče začlenění na trh práce. V mnohých zemích je toto šetření, vzhledem ke svému rozsahu a periodicitě, skutečně považováno za nejvhodnější nástroj ke zjišťování jinak obtížně dostupných sociálních statistik nejen domácího obyvatelstva, ale i imigrantů (např. vzdělanostní úroveň, získaná kvalifikace a pozice na trhu práce, složení domácností atd.). V zemích, jako jsou například Rakousko, Lucembursko, Německo a Holandsko, jsou data LFS používána dokonce jako hodnotná alternativa k oficiálním statistikám pro odhad skutečného počtu imigrantů v zemi. Nicméně v českém kontextu přináší použití tohoto datového zdroje pro statistiku migrace několik zásadních omezení.

V rámci LFS jsou standardně sbírány údaje o zemi narození respondentů a jejich státním občanství, případně o typu nebo délce pobytu v dané zemi. Navíc každý kalendářní rok je ke standardnímu dotazníku VŠPS připojen tzv. Ad-hoc modul, který se zaměřuje na určité téma nebo subpopulaci (například osoby se zdravotními potížemi, rodiny s dětmi, důchodce, mladé lidi na začátku pracovní kariéry apod.). V roce 2008 se Ad-hoc modul LFS věnoval situaci imigrantů a jejich potomků na trhu práce. [2] Nehledě na speciální tematické zaměření na mezinárodní migraci, toto šetření LFS realizované v ČR (podobně jako i ostatní šetření VŠPS) do značné míry podhodnotilo zastoupení imigrantů ve výběrovém vzorku. V roce 2009 z více než 36 tisíc respondentů mělo „pouze“ 354 dotazovaných (méně než 1 %) jiné než české občanství. Nejvíce byli zastoupeni občané Slovenska (126 respondentů), Ukrajiny (67), Vietnamu (33), Polska (22), Ruska (20) a Německa (13). [2] Pro srovnání je na tomto místě vhodné uvést oficiální statistiky, podle nichž pobývalo na konci roku 2008 na území ČR 397 536 osob s cizí státní příslušností starších 15 let. Podíl cizinců na celkovém obyvatelstvu ve věku 15 let a více tedy byl v daném období kolem 4,4 %. [3]

Jedním z hlavních vysvětlení, proč česká data LFS nejsou vhodná pro analýzu údajů o imigrantech, je samotný fakt, že se jedná o šetření domácností, které nepokrývá hromadná ubytovací zařízení, kde bydlí podstatná část imigrantů, zejména ekonomických. Dalším omezením je relativně nízké zastoupení cizinců v české populaci, takže nastává větší riziko výběrové chyby. V neposlední řadě může hrát roli i neochota cizinců odpovídat (ať již z důvodů jazykové bariéry či z obav sdělovat osobní údaje). Výsledkem působení těchto faktorů je několikanásobné podhodnocení (4-5krát nižší) zastoupení cizinců ve výběrovém souboru (a následných přepočtech na celkový stav obyvatel) ve srovnání s administrativními zdroji dat. Z tohoto důvodu se také strukturní údaje o cizincích v LFS (věk, vzdělání, státní občanství) liší od výstupů z administrativních zdrojů dat. Využití těchto výstupů pro mezinárodní srovnání může být poměrně zavádějící.

Problémy s neadekvátním zastoupením imigrační populace má také další rozsáhlý mezinárodní projekt Šetření příjmů a životních podmínek domácností (Životní podmínky), který je součástí

Evropského šetření EU-SILC. Toto šetření je organizováno jako rotační panel domácností zaměřený na osoby ve věku 16 a více let. Celkový počet respondentů dotázaných v ČR byl v jednotlivých letech až 27 tisíc. Součástí šetření Životní podmínky je každoročně soubor otázek věnovaných specifickému tématu. Na rozdíl od LFS nebyl v rámci šetření EU-SILC prozatím organizován žádný modul věnovaný tématu mezinárodní migrace. Šetření EU-SILC je bohatým zdrojem informací o sociální situaci českých domácností, jejich příjmech, bydlení a pracovních a zdravotních podmínkách dospělých členů těchto domácností. Vzhledem k poměrně širokému spektru otázek by mohla data z šetření EU-SILC posloužit jako důležitý zdroj jindy nedostupných a velmi žádaných informací o naturalizovaných cizincích a cizích státních příslušnících pobývajících v ČR. Jak ukazuje tabulka 1, realita je jiná. Ani v tomto rozsáhlém šetření nenajdeme dostatečně velký vzorek pro podrobnější analýzy, protože populace imigrantů, obzvláště těch, kteří se přistěhovali na území ČR poměrně nedávno, je v datech výrazně podhodnocena. [4]

Obě výše zmíněná evropská šetření (LFS a EU-SILC) prozatím nejsou z hlediska dat o imigrantech v ČR příliš užitečná. Přestože ze strany EU jsou v současné době iniciovány poměrně rozsáhlé aktivity k využívání dat z těchto dvou šetření k monitoringu integrace migrantů, v očích statistiků je tato snaha vnímána s velkou nevolí, neboť jsou si vědomi zkresení, které zmíněná výběrová omezení přinášejí. Kromě jiného také proto, že poskytují údaje především o úspěšně integrovaných cizincích.

[1] V upravené podobě se tomuto tématu bude LFS opět věnovat v roce 2014.

[2] Údaje pochází z frekvenčních tabulek z dat VŠPS poskytnutých Českým statistickým úřadem.

[3] Zdroj: Český statistický úřad

[4] Na druhou stranu data EU-SILC mohou být velice zajímavým zdrojem o současné situaci slovenské menšiny, která, ač je v ČR početná, zatím není příliš probádaná.

Yana Leontiyeva

Yana.Leontiyeva@soc.cas.cz

Tabulka 1. Imigranti v České republice ve výzkumech EU-SILC, 2007-2009

	2007		2008		2009	
	Země narození	Občanství	Země narození	Občanství	Země narození	Občanství
Slovensko	434	59	498	61	438	66
Ukrajina	68	23	76	23	73	29
Polsko	50	20	60	27	53	21
Rumunsko	26	1	26	1	23	1
Ruská federace	25	16	28	16	19	9
Ostatní oblasti Asie	25	20	28	20	15	14
Ostatní země	76	18	102	25	84	26
Nezjištěno	3676	3675	4179	4179	3537	3537
Celkem jiné země, než Česká republika, %	3,05%	0,68%	3,04%	0,64%	3,03%	0,71%
Česká republika	18679	19227	21936	22581	19060	19599
Celkový počet respondentů	23059		26933		23302	

Zdroj: Data EU-SILC 2007-2009 Česká republika, vlastní vypočet

« Vydává Sociologický ústav AV ČR, v.v.i., dne 1.12.2011 » « Šéfredaktorka: Renata Mikešová »
 « Redakční rada: Daniel Čermák, Radka Dudová, Jana Chaloupková, Yana Leontiyeva, Pat Lyons, Petra Guasti, Natalie Simonová, Eva Mitchell, Petr Sunega, Iva Štohanzlová » « Technická redaktorka: Jana Slezáková »
 « Adresa: SOCIOweb, Sociologický ústav AV ČR, v.v.i., Jilská 1, 110 00 Praha 1, tel./fax: +420 222 221 662, e-mail: socioweb@soc.cas.cz » « ISSN 1214-1720 »

« © Sociologický ústav AV ČR, v.v.i., Praha »