
Datové služby pro českou sociologii*

JINDŘICH KREJČÍ**

Sociologický ústav AV ČR, Praha

Data Services for Czech Sociology

Abstract: The article concerns the benefits of social data services and secondary data analysis and the past and current situation in data archiving in the Czech Republic. A debate concerning the formation of a social science data archive was opened in the late 1960s, but the ideas were brought to fruition in the late 1990s. The Sociological Data Archive (SDA) was founded in 1998 and it is the only institution that systematically provides access to data files from quantitative sociological surveys. The main access to the data library is provided on the internet. The SDA also pays great attention to promoting secondary analysis and the employment of existing data sources and co-operates in organising large research projects, especially the Czech participation in the ISSP. The SDA is a member of the CESSDA (Council of European Social Science Data Archives). This also means that the Archive can mediate access to materials stored in other social science data archives in Europe. The full inclusion of the SDA's services into an international network is connected to the adoption of international standards (DDI, XML technology), which is planned for the future. In recent years two qualitative data archives have also been established, the Czech Archive of Qualitative Data and Documents and the Digital Archive of Soft Data MEDARD. The Czech Statistical Office provides data services in the field of official statistics.

Sociologický časopis, 2002, Vol. 38 (No. 1-2: 125-138)

Význam archivu sociálních dat

Infrastrukturou výzkumu rozumíme zařízení a zdroje, které poskytují služby pro základní vědecký výzkum. V oblasti empirického sociálního výzkumu jsou to především knihovny a archivy materiálů, různá informační centra a databáze, metodologické zázemí výzkumu, výuková centra pro postgraduální vzdělávání, analytické služby, některé programy výzkumných šetření a archivy sociálních dat. Význam infrastruktury sociálního výzkumu v posledních desetiletích prudce roste, dochází k její koncentraci a propojování do mezinárodních sítí [viz Mochmann 1999, Warden 2001]. Důvodem je nejen růst objemu výzkumu a společenské poptávky po jeho výsledcích, ale také rozšiřující se mezinárodní spolupráce. Infrastruktura vytváří podmínky pro komunikaci mezi vědci navzájem a mezi vědou a společností a přispívá k harmonizaci a standardizaci výzkumných postupů, která je nutná pro srovnatelnost výsledků.

Důležitou součástí infrastruktury sociálního výzkumu jsou archivy sociálních dat, které pracují jako národní a mezinárodní centra. Jejich hlavním cílem je shromažďovat

*) Práce na této stati byla podpořena programem TSER Evropské komise v rámci projektu *Towards a European System of Social Reporting and Welfare Measurement*, registrovaného pod č. IC20-CT97-0048.

**) Veškerou korespondenci posílejte na adresu: Jindřich Krejčí, Sociologický ústav AV ČR, Jilská 1, 110 00 Praha 1, e-mail krejci@soc.cas.cz

elektronicky zpracovatelné soubory dat z empirických sociálních výzkumů a zpřístupnit je pro další, „sekundární“ analýzu. Archiv je zdrojem dat, vědeckých informací a souvisejících služeb a zároveň umožňuje další použití produktů skončených výzkumných projektů. Empirická šetření často vyžadují vysoké náklady, které jsou v řadě případů hrazeny z veřejných zdrojů. Je proto ve veřejném zájmu, aby data, jejichž informační hodnota obvykle přesahuje cíle jednoho projektu, byla využívána co nejvíce.

Datové archivy nefungují pouze jako pasivní knihovny datových souborů. Cílem jejich činnosti je všeobecná podpora sekundární analýzy sociálních dat [Mochmann a Guchteneire 1998, Mochmann 1998]. Vedle archivace a distribuce datových souborů, poskytují další služby pro sociální výzkum, jako jsou analytické služby, provozování knihoven analytických publikací, informační služby apod. Dále archivy poskytují technickou a organizační podporu pro výjimečné výzkumné projekty, tj. zejména pro projekty mezinárodního komparativního výzkumu a pro rozsáhlé dlouhodobé národní projekty. Důležitou součástí jejich činnosti je i vytváření komunikačních kanálů a s tím spojený vývoj informačních technologií.

Archiv usnadňuje použití empirického materiálu mimo výzkumný tým, který ho vytvořil. Uložená data a dokumentace jsou zabezpečena proti ztrátě nebo znehodnocení a zachována pro budoucnost. Existence archivu umožňuje veřejnou kontrolu nad kvalitou dat a vytváří tlak na zkvalitnění výzkumu. Publikované analýzy a interpretace je možné ověřit. Archivy slouží jako zdroj studijního materiálu pro výuku na vysokých školách a pro postgraduální studium. Datové archivy představují důležitou infrastrukturu pro mezinárodní komparativní výzkum a longitudinální výzkum [viz Lane 1990]. Vzhledem k rozvinuté vzájemné spolupráci a vytváření mezinárodních sítí datových organizací se archivy stávají centry mezinárodní spolupráce.

V Severní Americe a v západní Evropě byly první archivy zakládány v 60. letech, což souviselo s rozvojem možností elektronického zpracování dat. V posledním desetiletí se způsob fungování datových archivů změnil a jejich význam vzrostl. Stalo se tak s přispěním dvou faktorů: „globalizace v sociálním výzkumu“, kdy prudce vzrostl objem mezinárodní spolupráce, a rozvoji informačních technologií. Díky rozvinuté mezinárodní spolupráci mezi archivy a díky internetu má nyní výzkumník s vynaložením minimálních nákladů k dispozici přístup k obrovskému kvantu dat a dalších vědeckých informací uložených na celém světě. Nemusí přitom opustit vlastní kancelář, tento přístup mu zprostředkuje jeho počítač.

Následující text podává přehled datových služeb, které má k dispozici česká sociologie, a které jsou spojeny zejména se vznikem Sociologického datového archivu Sociologického ústavu Akademie věd ČR.

Archivace sociálních dat v ČR v evropském kontextu

Český empirický sociální výzkum má i v celoevropském měřítku poměrně dlouhou tradici. První empirická šetření proběhla už ve 30. letech minulého století a první program systematického šetření postojů byl založen už v roce 1946 v Československém ústavu pro výzkum veřejného mínění. Dopad komunistického režimu na další vývoj sociálních věd byl bohužel katastrofální. V 50. letech byla sociologie považována za buržoazní pavědu a všechny aktivity označené za sociologický výzkum byly zrušeny. Výběrová šetření byla nadále prováděna v oblasti socio-ekonomického výzkumu v rámci oficiálních statistik statistického úřadu.

Prudký rozvoj empirického výzkumu nastal po uvolnění režimu, zejména v druhé polovině 60. let.¹ Období „normalizace“ vrátilo situaci zpět. Některé aktivity v oblasti sociálního výzkumu byly položeny pod přímou kontrolu ÚV KSČ (např. výzkumný program bývalého Kabinetu pro výzkum veřejného mínění), jiné byly zcela znemožněny. Nadále byla ale prováděna výběrová šetření socio-ekonomického charakteru, včetně rozsáhlých projektů šetření veřejného ekonomického mínění, proběhlo několik šetření s tematikou rodiny a podařilo se zorganizovat i stratifikační výzkumy navazující na předchozí výzkumné aktivity (Třídní a sociální struktura 1978 a 1984).

Po převratu v roce 1989 došlo k rychlému nárůstu činnosti v oblasti empirického sociálního výzkumu a k zapojení ČR do mezinárodních výzkumných aktivit. Rozvoj výběrových šetření byl dán nejprve samotným faktem uvolnění restriktivního prostředí a dále pak růstem veřejné potřeby informací z výzkumů. Ta se neustále zvyšuje vzhledem k postupné stabilizaci demokracie, ekonomickému růstu a zapojování země do evropských a světových struktur.

Historie českého sociálního výzkumu a jeho současný rozvoj naznačuje potřebnost infrastruktury v oblasti datových služeb, ale i komplikace spojené s jejím budováním.

V oblasti oficiálních statistik zabezpečuje datové služby Český statistický úřad. Za určitých podmínek mohou být primární data z výběrových šetření statistického úřadu zpřístupněna výzkumníkům i mimo tuto instituci. Poskytování nezpracovaných datových souborů však nepatří k prioritám statistického servisu, který je zaměřen spíše na distribuci výsledků svých výzkumů.

V oblasti akademického výzkumu datové služby dlouho chyběly. Většina dat zůstávala v držení jednotlivých výzkumných týmů. Projekty systematické archivace dat byly i v případě jednotlivých institucí raritou. Za této situace nebyl dostupný žádný systematický přehled o existujících datech. Případný zájemce o další analýzu musel vhodný datový soubor nejprve vypátrat a pak se dohodnout s autorským týmem. Okamžitému použití často bránila nedostatečná dokumentace souboru či nevhodný formát dat a dotazníku.

Samotná myšlenka archivace dat z československých výběrových šetření v oblasti akademického výzkumu má ovšem velmi dlouhou tradici. Potřeba a možnosti uchovávat a dále rozšiřovat elektronicky zpracovatelné datové soubory prostřednictvím instituce datového archivu se objevila již na konci 60. let [viz Illner 1968]. Právě 60. léta byla zároveň obdobím počátků rozvoje datových služeb obecně a situace u nás nevykazovala nijak výrazné zpoždění. První evropský archiv, německý Centrální archiv v Kolíně nad Rýnem, byl založen v roce 1960. Období normalizace slibně se rozvíjející situaci změnilo. Idea vybudování českého datového archivu, který by sloužil sociologům, se stala ne-reálnou a naopak se objevila obava z možnosti zneužití uložených dat režimem. Na konci 70. let byl přesto založen tematický archiv zaměřený na výzkumy sociální stratifikace z let 1978 a 1984 [Matějovský a kol. 1979]. Ten byl ovšem v souvislosti s rušením bývalého Ústavu pro filozofii a sociologii ČSAV v roce 1990 zničen. Část materiálů byla zachována jeho uživateli a je dnes uložena v Sociologickém datovém archivu.

¹) Nelze nevzpomenout Výběrové šetření vertikální sociální diferenciací a mobility obyvatelstva ČSSR realizované pod vedením Pavla Machonina, které založilo tradici opakovaných výzkumů sociální stratifikace, jež trvá do současnosti [viz Tuček 2000].

Po roce 1989 se myšlenka vybudování širěji zaměřeného českého datového archivu opět stala populární a v tomto směru bylo zvažováno několik projektů a byla definována východiska pro jeho činnost. Potřebné finanční prostředky a úsilí se ale podařilo koncentrovat až v rámci projektu Sociální trendy. Na jeho základě byl v září 1998 v Sociologickém ústavu AV ČR pro veřejnost otevřen Sociologický datový archiv (SDA) [viz Krejčí 1998]. Trvalost zařízení nyní, po skončení projektu garantuje Sociologický ústav AV ČR.

Vedle SDA, který se specializuje na data z kvantitativních výzkumů, byly založeny též dva archivy kvalitativních dat. Český archiv kvalitativních dat a dokumentů při Masarykově univerzitě v Brně (viz Reference na internetu: Český archiv [viz Katrňák 1999]) vznikl též v rámci projektu Sociální trendy a s krátkým zpožděním po SDA. Digitální archiv měkkých dat MEDARD (viz Reference na internetu: MEDARD [viz Alan et al. 2000]) je součástí pražského Virtuálního institutu a byl založen v roce 2000.

Datem svého vzniku se český archiv SDA zařadil na závěr první vlny rozvoje archivační činnosti ve střední a východní Evropě. První národní databanka v této oblasti byla založena v Maďarsku v roce 1985. Datový archiv maďarského konsorcia TÁRKI však zůstal na dlouhou dobu osamocen. Až v devadesátých letech byly založeny archivy v Estonsku (v roce 1996) a ve Slovinsku (v roce 1997). Po nich následoval už český archiv SDA. Specifickým způsobem jsou zajištěny datové služby v Polsku. Institut sociologických studií Univerzity ve Varšavě je členem mezinárodní datové organizace ICPSR (Inter-university Consortium for Political and Social Research, viz Reference na internetu: ICPSR) a datové služby jsou částečně zajištěny jejím prostřednictvím. V 90. letech byly též založeny datové služby ve Finsku, Řecku a Irsku. Datové organizace nyní pracují ve všech zemích Evropské unie s jedinou výjimkou, a tou je Portugalsko. Evropské datové archivy jsou sdružené v Radě pro evropské datové archivy v oblasti sociálních věd CESSDA (Council of European Social Science Data Archives, viz Reference na internetu: CESSDA) a tvoří spolu rozsáhlou mezinárodní síť datových služeb. CESSDA nyní též podporuje nové aktivity, které směřují k vybudování datových služeb na Slovensku, v Rumunsku, Bulharsku, Lotyšsku, Litvě a v Rusku [viz Hausstein 2001, Reference na internetu: GESIS Branch Office].

Archivy kvalitativních dat v podobě národních institucí začaly ve světě vznikat až v 90. letech. Nejznámější a nejvlivnější organizací tohoto druhu je britský archiv Qualidata při Univerzitě v Essexu, který byl založen v roce 1994. Oba české kvalitativní archivy vznikly na přelomu století v souladu s novými trendy v budování infrastruktury sociálního výzkumu.

Projekt Sociologického datového archivu (SDA)

Jak již bylo řečeno, datový archiv SDA vznikl jako jeden z produktů projektu Sociální trendy (podpořen GA ČR). Rozsáhlý „komplexní“ projekt byl zaměřen na sociální výzkum, vzdělávání, publikaci „sociálních zpráv“ a archivaci. Od roku 1999 je SDA samostatným oddělením Sociologického ústavu, a jeho provoz je hrazen z rozpočtu ústavu. V letech 1999 až 2000 byl systém služeb archivu rozvinut do současné podoby díky projektu Dobudování archivu sociologických dat (podpořen GAČR). Od jara 2001 je archiv SDA členem Evropské rady archivů sociálních věd CESSDA.

Hlavním cílem archivu SDA je veřejné zpřístupnění dat z projektů sociologického výzkumu v ČR, dostupných výzkumů veřejného mínění a mezinárodních projektů s českou účastí pro akademické a nekomerční účely. Stejně jako v případě většiny starších a

větších zahraničních datových organizací činnost SDA směřuje k všeobecné podpoře sekundární analýzy dat. K cílům proto patří též technická podpora speciálních výzkumných projektů a podpora využívání existující infrastruktury v oblasti sekundární analýzy sociálních dat. Archiv nemá odborníky v oblasti informatiky. Rozvoj informačních technologií byl dosud zajištěn externí spoluprací a v budoucnu se předpokládá zapojení do mezinárodních projektů v této oblasti (viz dále).

Archivovaná data v SDA

Archiv shromažďuje datové soubory vzniklé na základě státního financování v Sociologickém ústavu AV ČR a v jiných českých institucích zabývajících se sociologickým výzkumem, soubory z mezinárodních výzkumů získané na základě dohod o výměně dat a soubory z dostupných výzkumů veřejného mínění. Datová knihovna nyní obsahuje přibližně 250 titulů datových souborů. Vzhledem k tomu, že české a anglické verze souborů jsou uloženy zvlášť, je však počet výzkumů, jejichž data jsou archivována ve skutečnosti o něco nižší. Archivace probíhá postupně. Cílem je shromažďovat jak starší, tak aktuální datové soubory. Archiv ale zatím nemá kapacitu shromáždit všechny dostupné soubory.

První fáze získávání a zpracovávání dat byla zaměřena zejména na mezinárodní komparativní projekty, na kterých Česká republika od roku 1990 participovala, a na výzkumy monitorující hlavní tendence ve vývoji sociální struktury. V další fázi se SDA zaměřil na data z výzkumů vzniklých v Sociologickém ústavu AV ČR od roku 1990. V současné době probíhá projekt zpracování a archivace dat z pravidelných šetření bývalého Institutu pro výzkum veřejného mínění (IVVM). Řadu datových souborů do archivu přinesla i spolupráce s vysokými školami, státními výzkumnými ústavy a též s některými komerčními agenturami.

V současné době jsou v archivu k dispozici např. data z následujících významných projektů:

- České výzkumy v rámci mezinárodního programu ISSP z let 1992 až 2000: Sociální nerovnosti (ISSP 1992), Životní prostředí a postoje k lokální politice (ISSP 1993), Rodina a role „gender“ (ISSP 1994), Národní identita (ISSP 1995), Role vlády (ISSP 1996), Pracovní orientace (ISSP 1997), Náboženství (ISSP 1998, sběr dat v roce 1999), Sociální nerovnosti a spravedlnost (ISSP 1999), Životní prostředí (ISSP 2000). Jedná se o kompletní datové soubory z českých šetření, tj. včetně specifických národních proměnných.
- Mezinárodní spojené datové soubory ISSP z let 1985 až 1996. Datové soubory obsahují spojené národní moduly z šetření programu ISSP a identifikační proměnné ve standardní podobě.
- Social Consequences of Transition 1995: data z mezinárodního šetření SOCO, které bylo věnováno sociálním dopadům transformace v ČR, Maďarsku, Polsku, na Slovensku a ve východních zemích Německa.
- Sociální spravedlnost 1991, 1995 a 1999: československá data z rozsáhlého mezinárodního projektu a data z českých opakování výzkumu v letech 1995 a 1999.
- Ekonomická očekávání a postoje I.-XI.: data z dlouhodobého projektu realizovaného v letech 1990 až 1993 v Česku i na Slovensku a v letech 1994 až 1998 pouze v ČR.

- Stratifikační výzkumy: Třídní a sociální struktura 1978 a 1984, Sociální stratifikace ve východní Evropě po roce 1989 (Bulharsko, ČR, Maďarsko, Rusko a Slovensko v roce 1993, Polsko v roce 1994).
- Deset let transformace: rozsáhlý výzkum provedený v ČR v roce 1999.
- Volební výzkumy: Exit Poll 1992 a 1996, 24 hodin před volbami do PSP v roce 1996, 24 hodin před volbami do Senátu v roce 1996, IVVM 1998, Trendy 1998 (duben a květen).
- Strategie a aktéři sociální transformace a modernizace: ČR, Slovensko a Polsko 1995.
- Data z výzkumů týmu České pohraničí z let 1990 až 1999.
- Data z longitudinálního výzkumu panelu rodičů a dětí Rodina 1989 až 1998.
- Data z výzkumů Muži a ženy na trhu práce 1991 a 1995 a Muži a ženy s vysokoškolským diplomem (1998).

V archivu jsou postupně zpřístupňována data z pravidelných výzkumů veřejného mínění bývalé agentury IVVM a ze současných výzkumů Centra pro výzkum veřejného mínění (CVVM), které navazují na tento program. Archiv obsahuje také data z řady dalších projektů a přístup k množství datových souborů může být zprostředkován na základě mezinárodní spolupráce.

Přístup k datům

Hlavní přístup ke službám datového archivu je vybudován na internetu (viz Reference na internetu: SDA), kde je umístěn elektronický katalog dat. Ten obsahuje přehled o archivovaných souborech s možností vyhledávání, základní informace o výzkumných projektech, datových souborech a jejich proměnných a dotazníky, případně kódovníky v elektronické podobě dostupné ke stažení. Informace prezentované na internetu jsou obnovovány periodicky. Nabídka archivovaných souborů nemusí být vždy úplná a některé informace o souborech a některé materiály nelze elektronickou cestou poskytnout. Archiv je proto možné kontaktovat s konkrétními dotazy na datové soubory a jejich dokumentaci.

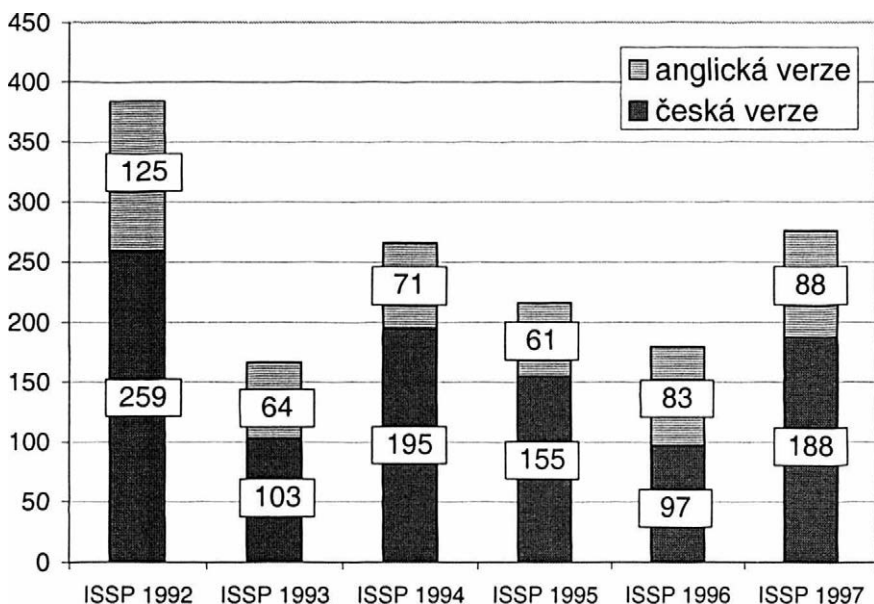
Vybrané datové soubory jsou dostupné přímo na internetu a je umožněno stáhnout je na vlastní počítač bez osobního kontaktu s archivem. Motivem pro založení této služby bylo zjednodušení výuky na vysokých školách. Data mohou být operativně využita učiteli při přípravě praktických seminářů nebo studenty pro vypracování seminárních prací. V grafu 1 je uvedena statistika používání této služby. V rámci mezinárodní spolupráce s Univerzitou v Kalifornii v Los Angeles (UCLA) archiv hostí internetové stránky výzkumného projektu Sociální stratifikace ve východní Evropě po roce 1989 (SSVE). Projekt SSVE nabízí možnost z internetu získat datové soubory z rozsáhlého šetření, které proběhlo v letech 1993-1994 v Bulharsku, ČR, Maďarsku, Polsku, Rusku a na Slovensku.

Datový katalog SDA vyšel též v tištěné podobě [Krejčí et al. 2000] v závěru roku 2000 jako interní publikace Sociologického ústavu AV ČR. Informace o datech jsou dále publikovány v bulletinu *SDA Info*, který archiv vydává, ve speciálních materiálech archivu [např. Krejčí 1999a] a v některých článcích v odborných periodikách [např. Krejčí 1999b].

Pro distribuci dat klientům a spolupracovníkům archivu byl vybudován speciální systém internetových stránek s přístupem chráněným pomocí hesla. Řadu souborů menšího rozsahu lze však prostě poslat elektronickou poštou. Tento systém je velmi operativní

a flexibilní, takže data se mohou ke svým uživatelům dostat velmi rychle. Data lze též vypálit na CD ROM a poslat poštou.

Graf 1. Využívání datových služeb Sociologického datového archivu „on-line“
Počet datových souborů v české a v anglické verzi, které byly staženy z internetových stránek SDA v období od 1. 1. 2001 do 30. 11. 2001. Registrováni jsou pouze externí uživatelé, přístupy ze sítě Sociologického ústavu (assoc.soc.cas.cz) registrovány nejsou.



Zdroj: Sociologický datový archiv.

Poznámka: Prostřednictvím služby jsou dostupné české soubory z výzkumů ISSP: • ISSP 1992 – Sociální nerovnosti, ISSP 1993 – Životní prostředí a lokální politika, ISSP 1994 – Rodina a role „gender“, ISSP 1995 – Národní identita, ISSP 1996 – Role vlády, ISSP 1997 – Pracovní orientace.

Podpora sekundární analýzy dat

Projekt datového archivu SDA směřuje k plnění cíle všeobecné podpory sekundární analýzy dat. To odpovídá zaměření i většiny ostatních evropských archivů. České specifikum je dáno dvěma okolnostmi: krátkou tradicí datových služeb a téměř neexistující ostatní infrastrukturou v oblasti sociálního výzkumu.

Krátká doba existence datového archivu a dvakrát přerušená tradice českého sociálního výzkumu způsobily, že problém rozvoje sekundární analýzy dat nestojí jen na straně nabídky datových služeb, ale i na straně poptávky po vhodných datech pro sekundární analýzu. Datové služby, ať již se jedná o služby SDA, jiných českých institucí nebo zahraniční služby, jsou méně používané, než by zasloužily, také proto, že výzkumníci málo znají existující možnosti. Dosud málo projektů obsahuje sekundární analýzu dat

jako podstatný prvek strategie zvolené k dosažení výzkumného záměru. Poměr investic do výběrových šetření a rozsahu publikovaných textů je také z tohoto důvodu často velmi nepříznivý. Pokud například projdeme již archivované datové soubory, zjistíme, že k některým se váží pouze dvě nebo tři analytické publikace a části některých výzkumů nebyly zpracovány vůbec.

SDA proto od počátku věnoval důraz na vybudování informační strategie, která zahrnuje publikování informací o archivu a dostupných službách v odborných periodikách, publikování vlastních informačních materiálů archivu, participaci ve výukových programech vysokých škol (např. je veden kurs Archivy sociálních dat na Fakultě sociálních věd UK) a organizování veřejných prezentací archivu a datových služeb. V tomto rámci SDA vydává též vlastní čtvrtletní informační bulletin SDA Info. Jeho obsah je zaměřen na přehledové informace o dostupné infrastruktuře sociálního výzkumu u nás i v zahraničí, informace o archivovaných datech a projektech, v jejichž rámci data vznikla, metodologii sociálního výzkumu a statistickou analýzu dat a na problematiku informačních technologií.

S důsledky problému neexistence některých dalších prvků infrastruktury sociálních věd se archiv setkal krátce po zpřístupnění internetových stránek v roce 1998. V té době se jednalo o jeden z mála českých internetových serverů v oblasti sociálních věd, který poskytoval též anglickou verzi svého obsahu. V důsledku toho se na archiv obracela řada lidí ze zahraničí, kteří hledali informace o české společnosti, politice nebo hospodářství i informace o České republice obecně. Zároveň byl archiv často kontaktován těmi, kteří hledali hotové výsledky výzkumů – výzkumné zprávy, odborné publikace, tabulky s daty atp.

Obě tyto skutečnosti se staly motivem k rozšíření služeb na internetu jednak o rozsáhlý adresář webovských zdrojů sociálních dat, českých a zahraničních center sociálního výzkumu a obecných informací o České republice, a jednak o nabídku analytických publikací z produkce Sociologického ústavu AV ČR. Podstatná část těchto publikací je dostupná ke stažení (*on-line*) v elektronické podobě. Rozšířené služby mají dvojí cíl. Poskytovat žádané informace z oblasti sociálního výzkumu, a zároveň přitáhnout pozornost ke službám datového archivu a k sekundární analýze dat obecně.

Technické zázemí pro speciální výzkumné projekty

Archiv SDA se podílí na tvorbě zázemí pro organizaci některých výzkumných projektů. Zejména se jedná o české výzkumy v rámci mezinárodního výzkumného programu ISSP (International Social Survey Programme). Pro projekt ISSP byl vytvořen systém čištění a transformace dat, sledování ukazatelů kvality a dokumentace šetření pro účely zpracování českých dat pro mezinárodní archiv. Členové týmu archivu soubory ISSP každoročně připravují pro jejich vložení do mezinárodního spojeného souboru. Datové soubory z těchto a některých dalších šetření jsou též standardním způsobem čištěny a upravovány pro účely jejich používání pracovníky Sociologického ústavu AV ČR. Transformacemi a dalšími úpravami dat tým archivu přispěl např. k českým výzkumům projektů International Social Justice Project (ISJP) a Second International Adult Literacy Survey (SIALS).

Mezinárodní spolupráce

Efektivním nástrojem práce datových archivů je mezinárodní spolupráce a vytváření mezinárodních sítí datových služeb. Jak již bylo uvedeno, archiv SDA je od letošního roku členem evropské rady archivů CESSDA. Díky ujednání o bezplatné výměně dat mají

nyní lokální uživatelé datových služeb SDA snadnější přístup k datům uloženým ve všech ostatních členských organizacích CESSDA, a to bez zvláštních poplatků.² V členských archivech je uložena např. řada mezinárodních souborů z projektů s českou účastí, jejichž data přímo v SDA dostupná nejsou.

Velký význam pro budoucí rozvoj SDA má ta skutečnost, že mezi archivy panuje dělba práce a princip sdílení výsledků společných projektů. To je mimo jiné důležité v oblasti rozvoje nových informačních technologií. Archiv díky členství v CESSDA získá možnost distribuce českých dat a informace o nich pomocí systému NESSTAR (viz Reference na internetu: NESSTAR), který propojuje datové služby do skutečné sítě tak, že materiály z participujících archivů jsou dostupné na jednom místě. Členství v CESSDA je výhodné i z hlediska financování rozvoje archivu. Vzhledem k tomu, že nastavení grantových programů Evropské komise předpokládá mezinárodní spolupráci, mají v nich koordinovaně postupující datové organizace větší šanci uspět.

Další důležité mezinárodní organizace jsou Mezinárodní federace datových organizací IFDO (International Federation for Data Organizations, viz Reference na internetu: IFDO) a zmíněné konsorcium ICPSR. Archiv SDA, ani žádná z institucí u nás zatím není jejich členem. IFDO je zaměřena na podporu sekundární analýzy dat, podporu mezinárodní integrace datových služeb a podporu rozvoje nových anebo nerozvinutých datových organizací ve světě. ICPSR pracuje při Univerzitě v Michiganu v USA. Je to sdružení velkého množství výzkumných a datových organizací, které poskytuje vlastní datové služby, a jedná se proto zároveň o největší datový archiv na světě. Služby ICPSR jsou organizovány na členské bázi. Přístup k nim je diferencován podle existence či neexistence členství a podle jeho druhu. Pro nečleny jsou služby poskytovány na komerčním základě. Členem může být buď instituce, nebo celý stát zastoupený konkrétní institucí. Podle toho mají přístup ke službám ICPSR buď členové konkrétní instituce, nebo všichni zájemci z dané země. Podle toho (tj. podle počtu potenciálních klientů, objemu, v jakém využívají služby ICPSR, a dalším ukazatelům) se ovšem odvíjí i výše členských příspěvků a podíl na organizaci činnosti ICPSR. Jestliže se v budoucnu v ČR vytvoří dostatečný zájem o využívání služeb ICPSR, je možné začít vyjednávat o našem členství.

Technologie XML a dokumentační standard DDI

Jako u většiny současných datových archivů je těžiště komunikace mezi SDA a uživateli jeho služeb umístěno na internetu. Webovský katalog poskytuje podobné služby a v podobném uspořádání jako katalogy řady dalších podobných serverů. Viditelné nedostatky ve srovnání s katalogy velkých archivů spočívají hlavně v menší systematickosti a někdy v neúplnosti informací. Plnohodnotnému zařazení materiálů z českého archivu do mezinárodních sítí ovšem brání problém, který je na první pohled skrytý. Formát informací nevyhovuje dokumentačnímu standardu a v pozadí služeb serveru je starší technologie, která neodpovídá trendu plánovaného vývoje.

Mezinárodní spolupráce archivů v posledních letech směřuje ke stanovení mezinárodního dokumentačního standardu pro popis dat. Definice obsahu a uspořádání dokumentace je založena na protokolu XML (eXtensible Markup Language). Jedná se o nový formát pro uchování elektronické informace. XML má podobný princip jako HTML, na

²) Bezplatný je transfer dat, tj. zájemce neplatí za datové služby zahraničního archivu. Datový soubor a dokumentace však může být např. dodána v podobě publikace na CD-ROM, která má stanovenou cenu, některé archivy též vyžadují úhradu za dokumentaci, kterou k datům zpracovávají.

jehož základě jsou dnes vytvářeny internetové stránky. Zatímco HTML vychází z popisu rozložení dokumentu (velikost fontu, barva atp.), XML popisuje strukturu a obsah dokumentu. Nový dokumentační standard je nazván DDI (Data Documentation Initiative) podle projektu, v jehož rámci byl vytvořen. Po stránce obsahu je dokumentace vytvořená v DDI rozdělena do pěti hlavních částí, které se dále větví až k úrovni jednotlivých konkrétních údajů dokumentace. Pomocí dalšího software, např. pomocí jednoduchého programu napsaného ve Visual Basicu nebo v jazyce XSL,³ lze tuto strukturu transformovat požadovaným, téměř jakýmkoliv způsobem – do naformátovaného dokumentu pro Word nebo do podoby stránky HTML apod.

Dokumentace archivovaná v SDA i katalog na serveru SDA je založen na prostém textovém dokumentu. Bylo stanoveno několik jednoduchých pravidel, podle kterých je dokumentace upravena. Speciální software převede takto upravený textový dokument do podoby internetové stránky ve formátu HTML a umístí ji na patřičné místo na serveru.

Nový, výše zmíněný nástroj NESSTAR pracuje na bázi standardu DDI. V době založení archivu byly oba projekty DDI i NESSTAR teprve v počáteční fázi. Nový český archiv nebyl součástí mezinárodních struktur a neměl možnost se projektů účastnit. Nyní před archivem SDA tedy stojí úkol zapojit se do navazujících projektů, převzít dokumentační standard DDI, transformovat stávající dokumentaci uložených dat, doplnit ji a publikovat ji pomocí systému NESSTAR. Jakmile se to podaří, SDA se plnohodnotně zapojí do mezinárodní struktury datových organizací. Útěchou nám může být, že tento úkol dosud leží i před některými renomovanými a velkými archivy. Paradoxem nového standardu totiž je, že jeho převzetí je nejnáročnější pro etablované dlouho fungující instituce, které jsou nuceny transformovat obrovské množství informací nashromážděných za řadu let jejich existence.

Přínosem nového systému není jen mezinárodní propojení služeb, ale i jejich podstatné rozšíření a zvýšený komfort při jejich používání [viz Ryssevik 1999]. Systém NESSTAR nabízí několik různých systémů vyhledávání a přístup ke kvalitní dokumentaci datových souborů. Vedle toho ovšem umožňuje přímo na internetu analytickou práci s datovým souborem na úrovni deskriptivních a regresních analýz a časových řad a možnosti kvalitního grafického vyjádření výsledků. Důležitým příspěvkem je zjednodušení dosud obtížného a zdoluhavého systému kontroly přístupu k materiálům a autorizace. V systému NESSTAR jsou v současnosti dostupné datové knihovny britského, dánského, finského a norského archivu.

Archivace kvalitativních dat

Cílem činnosti archivů kvalitativních dat je systematizace dostupných empirických materiálů v oblasti kvalitativního sociálního výzkumu, jejich digitalizace a zachování pro další výzkum. Práce kvalitativního archivu je ovšem velmi komplikovaná. Dodatečné náklady na archivaci dat jsou zpravidla vyšší než u kvantitativních výzkumů. Přímým impulsem pro rozvoj datových služeb v oblasti kvalitativních dat ostatně bylo až zlepšení technických podmínek pro převod materiálů do elektronické podoby a markantní zlevnění počítačové paměti v průběhu 90. let.

³) Oba tyto programovací jazyky jsou snadno dostupné. Visual Basic je součástí množství aplikací, např. MS Wordu, v XSL se dá pracovat např. na síti prostřednictvím aplikace Telnet.

Prvním problémem datových služeb v oblasti kvalitativního výzkumu je zajištění souladu legálních a etických pravidel ochrany osobních údajů a možností nakládání s dostupnými daty. V materiálech z kvalitativního výzkumu lze často buď přímo, nebo nepřímo identifikovat konkrétní osobu. Data z výzkumu potom mohou mít charakter osobních údajů, a vzhledem k obsahu se navíc většinou jedná o tzv. „citlivé údaje“. Podle současné legislativy platné u nás [Zákon č. 101/2000 Sb.] i v zemích Evropské unie [Směrnice 95/46/EC Evropského parlamentu a Rady] nelze data tohoto druhu používat bez písemného souhlasu „subjektu údajů“ a i v tom případě je lze používat jen pro účel, pro nějž byla původně pořízena.⁴ Starší kvalitativní materiály obvykle písemný souhlas neobsahují a zpětně ho získat bývá nereálné. V tom případě je třeba data anonymizovat nebo zlikvidovat. Pokud se podaří zajistit legální podmínky pro archivaci a práci s daty, datové archivy stanoví přísný režim pro jejich poskytování a zpracování. Kvalitativní data bývají archivem poskytována pouze na základě osobního souhlasu depozitora dat anebo jsou archivovány pouze informace o datech a kontakt na autorský tým.

Druhým problémem je zajištění dostatečné dokumentace kontextu kvalitativního výzkumu tak, aby bylo možné archivované materiály použít pro jiný projekt.

Starší z obou českých kvalitativních archivů, brněnský Český archiv kvalitativních dat a dokumentů, má větší ambice. Jeho cílem je zachovat kvalitativní data ze sociologických a sociálně psychologických výzkumů a informace o jejich dostupnosti v České republice [Katrňák 1999]. Archiv pracuje již několik let a shromáždil určité kvantum materiálů a informací o nich. Podrobnějších informací o jeho fondu je však pomálu. Aktivnější se v současné době zdá Digitální archiv měkkých dat MEDARD. Je propojen s prací autorů kvalitativních výzkumů spolupracujících s pražským Virtuálním institutem. Služby a částečně i obsah archivu MEDARD jsou prezentovány na internetu. Důraz je kladen též na oblast digitalizace dokumentů z kvalitativních šetření. Oba archivy mají kontakty na zahraniční aktivity. I v oblasti archivace kvalitativních dat se buduje mezinárodní síť datových služeb.

Datové služby ČSÚ

Datové služby Českého statistického úřadu (viz Reference na internetu: ČSÚ) se omezují na výzkumy statistického úřadu nebo případně na data získaná ze spolupráce ČSÚ s jinými institucemi. ČSÚ organizuje následující pravidelná výběrová šetření:

- *Mikrocensus*, což je pravidelné šetření opakované každých 3 až 5 let na 1 až 2 % reprezentativně vybraném vzorku domácností v republice. Šetření je zaměřeno na příjmové statistiky.
- *Rodinné účty* jsou pravidelné šetření založené na kvótním výběru cca 0,1 % domácností. Respondenti zaznamenávají své denní příjmy a výdaje.
- *Výběrové šetření pracovní síly* je čtvrtletně opakovaný výzkum cca 0,7 % reprezentativně vybraných domácností. Design šetření odpovídá mezinárodnímu standardu výzkumů typu Labour Force Survey, který je předepsán Eurostatem.

Na základě individuální dohody je možné vedle výsledků získat též přístup k datovým souborům z šetření. Lepší podmínky pro vyjednání přístupu k datům mají státní a spolupracující instituce. Dokumentace dat odpovídá spíše internímu používání dat v ČSÚ. Data

⁴) Podle zákona se ovšem o osobní údaj nejedná, „pokud je třeba ke zjištění identity subjektu údajů nepřiměřené množství času, úsilí či materiálních prostředků“ [Zákon č. 101/2000 Sb., § 4].

z ČSÚ jsou ovšem dostupná též prostřednictvím mezinárodních projektů, které jsou zaměřeny na srovnání dat z oficiálních statistik. K těmto projektům patří aktivity lucemburského institutu CEPS/INSTEAD (Centre d'Études de Populations, de Pauvreté et de Politiques Socio-Économiques/International Networks for Studies in Technology, Environment, Alternatives, Development, viz Reference na internetu: CEPS/INSTEAD), tedy LIS (Luxembourg Income Study) a LES (Luxembourg Employment Study), a dále pak některé projekty organizací EUROSTAT a CESTAT.

Praxe zpřístupňování datových souborů ČSÚ zhruba koresponduje s podmínkami, které nabízí také řada zahraničních statistických úřadů. Principy přístupu k oficiálním statistikám se od situace v akademickém výzkumu, kde je získání data pro sekundární analýzu relativně snadné, obvykle liší.⁵

Speciální projekty

Některé mezinárodní projekty z českou účastí v rámci svého výzkumného záměru poskytují též datové služby. Česká data, srovnatelná s ostatními národními daty mezinárodního šetření, jsou potom dostupná prostřednictvím těchto projektů. V jiném kontextu již byly zmíněny služby projektu Sociální stratifikace ve východní Evropě po roce 1989 a projekty LIS a LES, které také spadají do této oblasti. Nebudeme se k nim tedy již vracet. Zmíníme ovšem některé další významné aktivity.

Speciální server, který umožňuje analýzu dat prostřednictvím internetu, uvedl do provozu projekt New Democracies Barometer (viz Reference na internetu: CSPP). Zatím jsou v tomto systému k dispozici data z šetření projektu NDB realizovaných v 10 postkomunistických zemích střední a východní Evropy, včetně České republiky, a v Rakousku v letech 1990 až 1995. Brzy mají přibýt data z roku 1998. Data z mezinárodního šetření Family and Fertility Survey, které bylo zaměřené na rodinné a reprodukční chování a kterého se účastnila Česká republika v roce 1997, distribuuje populační komise OSN (viz Reference na internetu: PAU). Data je možné objednat přes internet. Česká republika přispěla též do projektu CSES (Comparative Study of Electoral Systems, viz Reference na internetu: CSES). Data z výzkumu volebního chování jsou k dispozici na internetu. Data z množství dalších mezinárodních šetření jsou dostupná standardním způsobem prostřednictvím služeb datových archivů.

Závěr

Výčet možností, jak se dnes dostat k datům relevantním pro českou sociologii, zde bohužel není úplný. Nebyly podrobněji popsány služby zahraničních datových archivů, které jsou dostupné i pro české badatele, a nebyla uvedena řada českých institucí, jež se sice nezabývají přímo poskytováním datových služeb, ale jsou ochotny umožnit přístup k některým svým datům. Lze též předpokládat, že některé datové služby se brzy zlepší a rozšíří. Cílem tohoto textu však bylo upozornit na rozvíjející se oblast datových služeb, seznámit odbornou veřejnost s jejími možnostmi a principy činnosti, a přesvědčit čtenáře, že přínos investic do budování infrastruktury sociálního výzkumu je pro českou sociologii důležitý. Fungující výzkumná infrastruktura má potenciál přispět k lepšímu umístění českého výzkumu do evropského i světového vědeckého prostoru. V národním měřítku může infrastruktura přinést větší systematičnost a více komunikace mezi výzkumníky,

⁵) Přehled datových služeb statistických úřadů v západní Evropě zpracoval projekt EuReporting: http://www.gesis.org/en/social_monitoring/social_indicators/EU_Reporting/index.htm.

jednotlivými vědními obory a mezi vědou a společností. Toto jsou vše cíle, které patří k prioritám, jež si definovala česká akademická veřejnost.

JINDŘICH KREJČÍ absolvoval v roce 1996 studium sociologie na Fakultě sociálních věd UK. Od roku 1998 pracuje v Sociologickém ústavu AV ČR, kde je vedoucím Sociologického datového archivu. Podílí se na mezinárodních výzkumných projektech ISSP (International Social Survey Programme) a ESS (European Social Survey) a zabývá se politickými postoji a kvalitou sociologických dat. Na Fakultě sociálních věd UK vede magisterský kurs o zdrojích sociálních dat.

Reference na internetu

- CESSDA – Council of European Social Science Data Archives: <http://www.nsd.uib.no/cessda/>
CSES – Comparative Study of Electoral Systems: <http://www.umich.edu/~nes/cses/>
CSPP – Centre for the Study of Public Policy: <http://www.cspp.strath.ac.uk/>
Český archiv kvalitativních dat a dokumentů: <http://www.fss.muni.cz/qarchiv/>
ČSÚ – Český statistický úřad: <http://www.czso.cz/>
DDI – Data Documentation Initiative: <http://www.icpsr.umich.edu/DDI/>
GESIS Branch Office Berlin, Studies from Eastern Europe: http://www.gesis.org/en/data_service/eastern_europe/index.htm
IASSIST – International Association for Social Science Information Systems and Technology: <http://datalib.library.ualberta.ca/iassist/>
IFDO – International Federation for Data Organizations: <http://www.ifdo.org>
ICPSR – Inter-university Consortium for Political and Social Research: <http://www.icpsr.umich.edu/>
MEDARD – Digitální archiv měkkých dat: <http://www.soc.cas.cz/trends/>
NESSTAR: <http://www.nesstar.org>
PAU – Population Activities Unit of the United Nations Economic Commission for Europe: <http://www.unece.org/ead/pau/>
SDA – Sociologický datový archiv SoÚ AV ČR: <http://archiv.soc.cas.cz/>
Sociální trendy: <http://www.soc.cas.cz/trends/>
SoÚ – Sociologický ústav AV ČR: <http://www.soc.cas.cz/>

Literatura

- Alan, Josef, Tomáš Bitrich, Zdeněk Konopásek 2000. „Digitální archiv měkkých dat MEDARD“. *SDA Info* 3/2000. ISSN 1212-995X.
Hausstein, Brigitte (ed.) 2001. *Social Science Data Archives in Eastern Europe*. Köln: ZA. URL: http://www.gesis.org/en/data_service/eastern_europe/news/NAF2001.pdf
Illner, Michal 1968. „Archív společensko vědních informací – nový pomocník empirické sociologie.“ *Sociologický časopis* 4: 507-510.
Katrňák, Tomáš 1999. „Český archiv kvalitativních dat a dokumentů.“ *SDA Info* 3/1999. ISSN 1212-995X.
Krejčí, Jindřich 1998. „Nový zdroj sociologických dat.“ *Sociologický časopis* 34: 384. ISSN 0038-0288.
Krejčí, Jindřich 1999a. *SDA – Sociologický datový archiv. Knihovna datových souborů ze sociologických výzkumů*. Praha: Interní publikace Sociologického ústavu AV ČR.
Krejčí, Jindřich 1999b. „Sociological Data Archive in Prague“. *ZA Information* 45: 142-152. ISSN 0723-5607.

- Krejčí, Jindřich, Lumír Gatnar, Miroslav Kříž, Petr Soukup, Vladislav Tytko 2000. *Katalog dat Sociologického datového archivu*. Praha: Interní publikace Sociologického ústavu AV ČR.
- Krejčí, Jindřich. 2001. „The Czech Sociological Data Archive“. In: *Social Science Data Archives in Eastern Europe*. Ed. by: Brigitte Hausstein. Köln: ZA.
- Lane, Jan-Erik 1990. „Data Archives as an Instrument for Comparative Research.“ In *Comparative Methodology. Theory and Practice in International Social Research*, ed. by Else Oyen. London, Newbury Park, New Delhi: SAGE Publications. ISBN 0-8039-8325-5.
- Matějovský, Antonín, J. Hudeček, H. Jeřábek, J. Ježek, V. Stupka, M. Tuček, A. Vodáková, J. Voženílek 1979. *Metodika Výzkumu třídní a sociální struktury ČSSR 1978*. Díl I. Praha: Interní publikace Ústavu pro filozofii a sociologii ČSAV.
- Mochmann, Ekkehart 1998. *European Cooperation in Social Science Data Dissemination*. IFDO Net: http://www.ifdo.org/archiving_distribution/index_bfr.htm
- Mochmann, Ekkehart 1999. „European Infrastructure Needs for Comparative Socio-economic Research.“ Paper presented at the European Socio-economic Research Conference in Brussels, 28-30 April 1999.
- Mochmann, Ekkehart, Paul de Guchteneire 1998. *The Social Science Data Archive Step by Step*. IFDO Net: http://www.ifdo.org/archiving_distribution/index_bfr.htm
- Ryssevik, Jostein 1999. „Providing Global Access to Distributed Data Through Metadata Standardisation – The Parallel Stories of NESSTAR and the DDI.“ Paper given at the UN/ECE Work Session on Statistical Metadata, Geneva, September 1999. URL: <http://www.nesstar.org/papers/>
- Směrnice 95/46/EC Evropského parlamentu a Rady z 24. října 1995, o ochraně jednotlivců v souvislosti se zpracováním osobních údajů a s volným pohybem těchto údajů*. URL: <http://www.uoou.cz/legisl.php3>
- Tuček, Milan 2000. „Přehled výzkumů sociální struktury a mobility v ČR.“ *SDA Info* 4/2000. ISSN 1212-995X.
- Warden, Campbell (ed.) 2001. *Reflections on the Role of Research Infrastructures in the European Research Area*. Luxembourg: European Commission, Office for Official Publications of the European Communities. ISBN 92-894-1070-1.
- Zákon č. 101/2000 Sb., o ochraně osobních údajů a o změně některých zákonů ze dne 4. dubna 2000*. URL: <http://www.uoou.cz/legisl.php3>