

K metodickým otázkám vytváření fondů banky dat výzkumů veřejného mínění

PAVEL DVOŘÁK

Ústav pro výzkum veřejného
mínění při FSÚ, Praha

Nové požadavky kladené na výzkum veřejného mínění přináší s sebou také nové nároky na metodologickou a metodickou vybavenost výzkumu veřejného mínění. To se postupně promítá ve všech rovinách obecnosti — od teoreticko-metodologické až po technicko-organizační úroveň.

Jedním z možných a nutných směrů v rozvoji výzkumu veřejného mínění je aplikace moderních informačních technologií v uchovávání a soustavném využívání získávaných empirických dat o veřejném mínění, například prostřednictvím banky dat.

O výstavbě banky dat bylo rozhodnuto v souvislosti s projednáváním úkolů Ústavu pro výzkum veřejného mínění při FSÚ ve stranických orgánech. V roce 1984 byl zahájen zkušební provoz banky dat, v průběhu roku 1987 bylo politické praxi operativně předáno kolem 50 různých faktografických informací o veřejném mínění. Současné zkušenosti a požadavky na banku dat i poznatky o jejím využití a problémech potvrzují, že její výstavbu nelze chápat jako jednu ukončený proces. Potvrzuje se také, že zvolený přístup umožňuje postupně rozvíjet služby banky dat, zdokonalovat její vnitřní strukturu a využívat dokonalejší technické vybavení.(1)

Koncepce výstavby banky dat vychází z poznání, že na určitém stupni vývoje informačního systému dochází v důsledku jeho rozsahu a složitosti k objektivní potřebě vytvořit systematicky organizovaný subsystém o daném informačním systému.

Analýza budovaného informačního systému vedla k rozčlenění banky dat na tři části — technologickou bázi, кадровé zabezpečení činnosti a organizační uspořádání.

Základem je technologická báze. Je to souhrn fondů, soustavy jazykových prostředků, metodik a programového vybavení v určitém technickém prostředí.

Uspořádání fondů banky dat je podstatným způsobem závislé jednak na charakteru vlastních informací o veřejném mínění a jednak na jejich zapojování do různých procedur v rámci informačních procesů banky dat. Tyto momenty se odrážejí v celkové koncepci budování banky dat — tj. v aplikaci databankové technologie při zpracování údajů z výzkumů i při zpracování metainformací. Fondy banky dat zahrnují také ty části, které nejsou uloženy na počítačových médiích.

Konkrétní podoba fondů banky dat samozřejmě závisí na technických možnostech řešitele, resp. provozovatele banky dat výzkumů veřejného mínění, to také ovlivňuje konkrétní strukturu, rozsah, obsah a typy uchovávaných údajů.

Struktura fondů úzce souvisí s otázkou *zdroje údajů banky dat*. Vzhledem

(1) Tento text navazuje na stať „Ke koncepci banky dat výzkumů veřejného mínění“ uveřejněnou v č. 6/1986 Sociologického časopisu a věnuje se otázkám struktury informačních a metainformačních fondů banky dat. Vybrané otázky činnosti banky dat výzkumů veřejného mínění budou předmětem další práce.

k mnohorozměrnosti a složitosti problému tvorby banky dat sociologických výzkumů je zřejmě nutno alespoň v počátku podstatným způsobem omezit okruh uvažovaných zdrojů pro banku dat. Omezení zdrojů informací by zároveň mělo určitým způsobem reflektovat jak metodologické otázky srovnatelnosti uchovávaných údajů, tak organizačně právní aspekty využívání údajů produkovaných různými institucemi. Z hlediska optimálního „zaběhnutí“ banky dat bylo výhodné se omezit na metodologicky a metodicky srovnatelná data jedné či dvou institucí. V případě výzkumu veřejného mínění je situace v podstatě dána hlavními producenty výzkumů veřejného mínění – ÚVVM při FSÚ v Praze a ÚVVM při SSÚ v Bratislavě. S přihlédnutím k charakteru výzkumů ÚVVM při FSÚ byla také banka dat účelově budována.

V souvislosti se zdroji údajů banky dat výzkumů veřejného mínění je nutno se zmínit o dosud nerealizované možnosti zahrnutí údajů pocházejících i z jiných zdrojů, než jsou výzkumy veřejného mínění. Takovými relevantními exogenními údaji jsou bezesporu údaje sociálně demografického charakteru umožňující konstruovat a vyhodnocovat výběrové soubory a složení zkoumaných sociálních skupin. Jiným okruhem údajů by se mohly stát jednotlivé vybrané části sociálně ekonomických informací, produkovaných soustavou statistických orgánů, umožňující v určité míře konfrontovat názory občanů se stavem zobrazovaným soustavou státní statistiky [Illner 1981 : 95].

Fondy banky dat výzkumů veřejného mínění byly rozděleny na dvě základní části – informační a metainformační základnu.

K informační základně

Obsah informací o veřejném mínění je ve svém souhrnu determinován charakterem výzkumného procesu, který je produkuje a také předpokládanými informačními procesy a využitím banky dat. Konkrétní struktura informační základny musí reflektovat obecnější momenty spojené s tím, že představuje specifickou formu reprezentace poznatků o veřejném mínění [Řehák 1981 : : 273].

Tyto empirické poznatky mají různý stupeň obecnosti a je tak možno identifikovat škálu jejich různých typů. Na jedné straně stojí data, bezprostřední výsledek sběru dat, na straně druhé pak v některých případech i empirická zobecnění. Tato skutečnost by měla nalézt svůj adekvátní odraz ve struktuře fondů informační základny.

Funkce informační základny spočívají v uchovávání relevantních informací o veřejném mínění tak, aby mohly být systematicky využívány při realizaci funkcí banky dat.

Jako hlavní typy informací o veřejném mínění, které jsou součástí fondů banky dat, byly vymezeny:

- a) **Základní data** výzkumů uchovávaná ve formě matice uložené na počítačovém médiu (a také po určitou dobu ve formě jednotlivých vyplněných dotazníků). Pro každý výzkumný úkol je vytvořen jeden nebo více takových souborů dat. Údaje jsou zakódovanými výsledky zjišťování (prováděného zpravidla dotazováním) za jednotlivé jednotky (jednotlivce) výběrového souboru. Základní data výzkumů ÚVVM jsou ukládána do *báze dat PANDA*, která zahrnuje jednotlivé výzkumné soubory. Každý z výzkumů (ÚVVM) je realizován v rámci této báze dat jako svazek počítačových souborů řízených a zpracovávaných systémem Programů ANalýzy Dat (PANDA). Tyto soubory jsou umístěny na disku počítače CYBER 172.

Základní data jednotlivých výzkumů představují jeden z klíčových informačních fondů banky dat. (2, 3)

b) *Agregované a shrnující údaje* zpravidla ve formě různých výstupů matematicko-statistického zpracování. Primární individuální informace jsou sumarizovány ve formě různých statistik — procentních četností, koeficientů apod. V závislosti na analytických cílech výzkumného týmu a na možnostech zpracování jsou k dispozici různé druhy, ať již z hlediska formy, tak z hlediska uspořádání a obsahu výstupních tabelací. Je nutno připomenout, že vzhledem k deskriptivnímu charakteru běžného výzkumného procesu jsou často využívány víceméně standardní sestavy obsahující nejzákladnější sumarizované údaje. V ÚVVM je takovýmto nástrojem tzv. stručné třídění, které obsahuje kontingenční tabulky všech zkoumaných proměnných tříděných podle standardních identifikací.

Vzhledem ke zkušenostem s tímto typem tabelací byl do fondů banky dat zahrnut *operativní fond dat*, základní standardní tabulkové sestavy (třídění meritorních proměnných se standardními identifikacemi) z jednotlivých výzkumů. Tyto tabelace jsou na *mikrofiších*. Fond operativních dat slouží k rychlému seznámení se základními údaji a k efektivnímu ukládání systemizovaných dat. (4)

(2) *Výhody* tohoto typu údajů jsou spojeny zejména s:

- možností replikace libovolné manipulace s daty, odvození složitějších a syntetičtějších datových struktur
- možností opakovat nebo iniciovat nové matematicko-statistické zpracování dat
- maximální podrobností uchovávané informace
- možností zpracovávat opakovaně údaje v různých sekundárních kontextech
- možností realizovat systematické metodologické zhodnocení kvality dat
- možností postupně vyhodnocovat přístupy ke standardizaci apod.

Nevýhody uchovávání základních dat spočívají zvláště v:

- nutnosti mít k dispozici vhodné programové vybavení, v kontextu banky dat pak musí jít o relativně univerzální systém analýzy sociologických dat
- potřebě velkokapacitních médií, nejlépe diskových pamětí, vyhovujících organizaci báze dat z řady výzkumů
- nutnosti vytvoření systému údržby celé datové báze, což znamená nároky jak na počítačové kapacity, tak na kádrové vybavení.

Další technické otázky jsou spojeny s volbou vhodného média uchovávání základních dat a samozřejmě s přístupem k jejich zpracovávání.

(3) V současné době je v bázi dat PANDA uchováváno kolem 75 výzkumných souborů z dosud provedených výzkumů. Data z těchto výzkumů je možno opakovaně zpracovávat prostřednictvím výše zmíněného systému. Časová dosažitelnost výsledků matematicko-statistického zpracování se pohybuje, v závislosti na provozních podmínkách počítačového systému CYBER 172, kolem 3 — 6 hodin od zadání požadavku.

(4) *Výhody* archivace tabulek (výstupů matematicko-statistického zpracování) lze spatřovat zejména v:

- efektivním využití jednou zpracovaných údajů, které jsou ve vhodné formě
- údaje tabelací nejsou zatíženy případným interpretačním kontextem (a možná i chybami) primárního zpracování.

Nevýhody uchovávání tabelací jsou:

- v tištěné formě jsou tabelace obtížně archivovatelné a katalogizovatelné (nároky na prostor, problémy s identifikací tabulek)
- obtížně zpřístupňovatelné uživateli
- v počítačové formě (na pásce či na disku) jsou velmi náročné na rozsah paměti, citlivé a komplikované na manipulaci
- závislé na profesionální připravenosti primárního zpracovatele (který provedl výběr metody zpracování a vstupních analyzovaných údajů) a kvalitě matematicko-statistického zpracování (správnost programů)
- vysoce variabilní typy sestav jak co do obsahu, tak co do uspořádání.

Údaje tohoto typu jsou totiž třeba téměř při každé analýze, zároveň samy o sobě mohou uspokojovat elementární informační požadavky. (5)

c) *Závěrečné zprávy výzkumů* obsahují interpretované, sumarizované a komentované výsledky výzkumných akcí. Reprezentují zpravidla „konečný“ výsledek celého výzkumného procesu, informaci pro zadavatele.

Archivace závěrečných zpráv výzkumů (a obdobných materiálů) je zřejmě nutná, nemůže však výrazněji ovlivnit hlubší sekundární analýzy. Výtah z těchto zpráv může ovšem uspokojit jednoduché operativní informační požadavky. (6)

Mimo uvedených tří hlavních typů informací bezprostředně spojených s výzkumným procesem je možno v budoucnosti uvažovat i o dalších typech informací:

- o vybraných publikacích v odborných časopisech (Sociologický časopis, Zprávy sociologické společnosti atp.) obsahujících údaje výzkumů
- o výzkumných zprávách různých institucí využívajících údaje o veřejném mínění apod.

V souvislosti s informačními procesy vzniká při fungování banky dat řada sekundárních materiálů. Jsou to jednak různé agregované údaje (obdobně jako v případě ad b)) a jednak další *informace o veřejném mínění*. Výsledky sekundárních analýz zpravidla obsahují souhrnné údaje z několika výzkumných akcí. Mají, obdobně jako závěrečné zprávy, určité výhody a nevýhody. Jejich zahrnutí do informační základny je však zdůvodněno potřebou systematického vyhodnocování vlastní činnosti banky dat, resp. činnosti realizované za pomoci banky dat.

Dalším metodologickým problémem je *zahrnutí exogenních údajů*, tj. údajů produkovaných různými zdroji dat mimo vlastní výzkumný proces. Jde zejména o sociálně demografická, zpravidla cenzová nebo mikrocenzová data, využívaná ke konstrukci a vyhodnocování výběrových souborů. V dalších etapách výstavby, po získání širších zkušeností s provozem banky dat, se strukturou a charakterem informačních požadavků, bude zváženo zahrnutí fondů sociálně ekonomických údajů. Lákavá je v této souvislosti myšlenka

(5) V operativním fondu dat, tj. na mikrofiších, je v současné době (říjen 1987) uloženo základní třídění údajů z téměř 50 výzkumů veřejného mínění. Pro tabelaci jsou využity základní identifikační proměnné – území (ČSR–SSR), pohlaví, věk, ekonomická aktivita, vzdělání a velikost místa bydliště dotazovaných. Operativní fond, dosažitelný prakticky ihned (v případě zobrazení dat na prohlížeče mikrofišů) či během několika minut (při použití papírové kopie mikrofiše), zahrnuje tedy v současné době kolem 40 tisíc tabulek druhého stupně. Ve spojení s metainformačními prostředky je tak možno předat uživateli jednoduchou informaci faktografického typu během několika minut i při prohlédávání celého fondu údajů.

(6) K *výhodám* uchovávání závěrečných zpráv výzkumů patří:

- obsahují již interpretované údaje
- informace jsou v uživatelsky přístupné podobě (texty, grafy)
- archivované zprávy vyžadují minimum údržby a lze je vhodně, běžnými katalogizačními prostředky uspořádat a využívat
- odpovídají svou formou a obsahem, tj. stupněm zobecnění údajů, většině uživatelů
- minimum nákladů na technické a kádrové zabezpečení archivace a zpřístupňování.

Nevýhody archivace závěrečných zpráv lze shledat zejména v:

- závislosti na interpretačním kontextu primárního zpracování, které nemusí odpovídat sekundárním požadavkům
- podmíněnosti kvality poznatků profesionální připraveností výzkumného týmu
- tom, že údaje nejsou vybrány tak, aby respektovaly nové informační požadavky (absentuje řada údajů)
- obtížnosti reprodukce závěrečných zpráv.

o propojení banky dat výzkumů veřejného mínění na jiné, obdobně organizované banky dat, například Federální banku dat či na speciální báze dat vytvořené na základě rozsáhlých datových základů sociologických výzkumů typu „mikrocenzus“.

K metainformační základně

Jedním z klíčových míst realizace automatizovaného informačního systému výzkumů veřejného mínění je vytvoření specifického informačního subsystému, jehož objektem je výzkumný proces a jeho produkty [Dvořák 1986 : 623].

Pro zabezpečení vysokého stupně operativnosti a efektivního fungování celého systému je potřebné, aby i procesy práce s metainformacemi byly automatizovány, nejlépe na úrovni využití databankové technologie. Datovou základnu tohoto subsystému tvoří v navrhovaném řešení metainformační základna.

Právě volba adekvátních úrovní a objektů metainformačního popisu je jedním z rozhodujících metodologických problémů výstavby banky dat. Je nutno zdůraznit, že jde zejména o problém z oblasti metodologie výzkumu veřejného mínění, resp. sociologie, nikoli z oblasti využití výpočetní techniky.

Jak již bylo při rozboru koncepce výstavby banky dat konstatováno, tvoří metainformace základ pro ukládání údajů různého druhu, pro jejich vyhledávání a systematizaci podle různých informačních potřeb. Struktura metainformační základny musí, mimo jiné, určitým způsobem respektovat uspořádání informační základny.

Z tohoto hlediska se ukazuje, že hlavní komponentou metainformačního subsystému jsou data o datech (metadata). Požadavek na tvorbu metadat je důsledkem celé řady faktorů. Mezi nejzávažnější patří zejména nárůst objemu uchovávaných dat. Zde nejde jen o počet výzkumů, ale zejména o polytematický záběr a široký rozsah ve výzkumech zjišťovaných sociologických znaků.

Při selekci proměnných odpovídajících určitému informačnímu požadavku je pak nutno prohledat celý tento objem nebo jeho podstatnou část. Efektivní automatizace vyhledávání relevantních znaků je tedy jednou z nejzákladnějších komponentů banky dat.

Náročnost a složitost vyhledávání a zprostředkování archivovaných údajů při sekundárním zpracování souhrnných informací je také odrazem značně variabilních informačních požadavků zadavatelů. I uspokojení potřeby soustavného a systematického metodologického hodnocení kvality získávaných údajů se neobejde bez vlastní údajové základny.

V souhlase s literaturou [Mezsároš 1979 : 473; Šoltéz 1984 : 251–262] věnovanou tomuto tématu je možno vyčlenit následující hlavní funkce metainformačních fondů:

1. *Informační (zobrazovací) funkce*, jejímž obsahem je informování různých subjektů, využívajících banku dat jako součást celku výzkumu veřejného mínění, o stavu datové základny, o prováděných výzkumech apod. Jde o určitou soustavu poznatků o systému výzkumu veřejného mínění.
2. *Integrační funkce*, spočívající v zajišťování propojitelnosti a vzájemné návaznosti jednotlivých automatizovaných i neautomatizovaných částí banky dat. Integrační funkce metainformací reaguje na rozmanité pohledy a požadavky. Nezanedbatelnou součástí integrační funkce je využití metainformací pro ochranu uchovávaných údajů.
3. *Komunikační funkce* je naplňována využitím určitých částí metainformačních údajů při vytváření a využívání vhodných způsobů komunikace mezi

uživatelé a celkem banky dat i způsobů komunikace mezi obsluhou banky dat a jejich fondy.

4. *Selektivní funkce* je spojena s rozsáhlým využitím metainformací při *vyhledávání relevantních údajů* v informačních fondech podle různých hledisek.
5. *Metodologická funkce* je úzce svázána s informační funkcí. Uchovávané metainformace se mohou stát vhodnými informacemi, fakty pro *uskutečňování metodologických rozborů a analýz nebo pro standardizaci přístupů*.
6. *Administrativně dokumentační funkce*, související s potřebou *správy datových fondů*, vyvolávající požadavky na řadu evidenčních, dokumentačních a administrativních údajů.

Důležitým momentem při koncipování metainformační základny je také *volba metod, forem a prostředků vytváření, ukládání a vyvolávání údajů*.

Využití databankové technologie samozřejmě předpokládá specifický přístup k vytváření metainformací. *Za základní metodu se považuje systémová analýza místa metainformačních údajů, navazující posléze na využití postupů formalizovaného popisu* nutných pro aplikaci databankové technologie.

V souvislosti s vytvářením automatizovaných informačních systémů jsou v literatuře [Veselý 1984 : 67–82] poměrně rozsáhle diskutovány i *organizační formy metainformačních údajů*.

Jako jeden ze základních prostředků metainformačního popisu, typický i pro řešení systémů řízení báze dat v rámci konkrétních databázových systémů, je uvažován *centrální adresář dat*, resp. *slovník a adresář dat*. (7)

V praktických aplikacích databankových systémů jsou tyto dvě složky odůvodněně spojovány v *slovník a adresář dat*, jenž je zpravidla podstatnou částí systému řízení bází dat. V případě banky dat výzkumů ÚVVM jde zejména o bázi dat PANDA.

Slovník dat je vymezován jako specifický seznam dat daného informačního systému, který obsahuje jejich jednoznačnou identifikaci, určitou obsahovou definici a případné dokumentační údaje. Z hlediska uživatele, pro něhož je zejména určen, reprezentuje určitou formu zobrazení *logické struktury dat*.

Adresář dat je možno charakterizovat jako mezičlánek, pojítko mezi metainformačními údaji a vlastními daty. Na rozdíl od slovníku obsahuje zejména údaje o místě, kde nachází vlastní informace, data. V případě uložení dat v bázi jsou pak v adresáři shrnuty údaje o fyzické struktuře konkrétního média, způsobu záznamu apod. a přístupu k jejich hodnotám. V širším pojetí adresáře, tak jak je uvažován v této práci, může adresář obsahovat i údaje o uložení určitých informací i v neautomatizovaných částech informační základny.

Z výše uvedených charakteristik slovníků a adresáře dat je zřejmé, že jsou primárně orientovány na datovou komponentu informačního systému. K dalším formám patří – v souladu s řešením statistického metainformačního systému – *katalogy a registry* (srv. Melišková–Fischer 1982), představující další organizační formy popisu i dalších objektů metainformační základny.

Registr je souborem relativně stabilních, systematicky doplňovaných a aktualizovaných údajů, popisujících existenci, vlastnosti a vztahy *jednotli-*

(7) V literatuře se také rozlišuje mezi vymezením adresáře dat (data directory) a slovníkem dat (data dictionary).

Slovník dat je charakterizován jako prostředek centralizované uživatelské dokumentace dat a nemusí být nutně umístěn na počítačových médiích a příslušně zpracováván.

Adresář dat je na rozdíl od slovníku dat úzce spojen s počítačovou (fyzickou) strukturou ukládaných údajů.

vých prvků určité třídy objektů. Například může jít o registr jednotlivých dokumentů, registr jednotlivých výzkumných pracovišť apod. Pro objekty popisované prostřednictvím registru je charakteristické, že jejich výskyt je *jedinečný*. O registru budeme v dalším hovořit ve významu počítačového souboru údajů, v případě, že údaje nebyly umístěny na počítačových médiích, bylo použito rejstříků.

Katalog je souborem relativně stálých nebo systematicky doplňovaných a aktualizovaných údajů popisujících existenci a vlastnosti určité třídy objektů, popřípadě vztahy mezi třídami. Zobrazuje tedy vlastnosti typů objektů na rozdíl od registru. V případě dané banky dat může jít například o katalog identifikačních proměnných nebo sociálních skupin (jak identifikační proměnná, tak sociální skupiny se ve výzkumech veřejného mínění vyskytují opakovaně). V případě uchovávání údajů o typech objektů mimo počítačová média bylo použito *kartoték*.

Celá soustava organizačních forem metainformačního popisu je doplňována *klasifikacemi a číselníky*, které vhodným způsobem zobrazují kódy (numerické, alfanumerické či alfabetycké) a jejich význam.

K celé soustavě organizačních forem metainformačních údajů je nutno poznamenat, že v bance dat výzkumů veřejného mínění se využívá vhodná, vzájemně propojená kombinace těchto forem.

Metainformační základna (i navazující způsoby jejího využívání) je řešena jako *otevřený systém*.

Ještě jednou považujeme za nutné připomenout, že *stanovení obsahu metainformační základny (a její naplňování) je primárně problémem metodologickým a teprve v realizační rovině je problémem zejména z oblasti výpočetní techniky*.

Metodologická analýza výzkumného procesu a jeho produktů umožnila stanovit různé objekty metainformačního popisu i vymezení konkrétních sledovaných metainformačních údajů.

Data z výzkumů jsou v ÚVVM organizována formou báze dat PANDA, kde jednu z určujících forem reprezentuje *matice dat*. Tato matice vzniká v procesu sběru dat a následného zpracování (příprava dat, záznam dat, transformace proměnných apod.). K ní se váže řada údajů metodologického a metodického charakteru, jejichž charakter i rozsah ukazuje, že *matice dat* by mohla být jako *určitý homogenní celek jedním ze základních typů objektů* metainformačního popisu. Pro tento objekt je charakteristická relativní shoda údajů z hlediska cíle získání údajů, času (tj. období získání konkrétních odpovědí dotazovaných) a zpravidla i *metody*, způsobu zjištění údajů. Přičemž z hlediska způsobu zjištění jde jak o zvolené výběrové postupy (kvótní výběry), tak o techniky sběru dat (standardizovaný rozhovor). Pro uvedenou matici dat je také charakteristické, že obsah zjišťovaných údajů je víceméně stejnorodý. Tato skutečnost je dána tím, že v praxi výzkumu veřejného mínění jsou prováděny většinou monotematické výzkumy.

Stěžejní datovou strukturou, užívanou ve výzkumech veřejného mínění (a v obdobných sociologických výzkumech tohoto typu) je *sociologická proměnná*, zkoumaný znak [Dvořák 1986 : 622].

V matici dat je možno jako specifické nositele informací rozlišit:

- a) *Meritorní proměnné*, jež reprezentují určité stránky veřejného mínění ve vztahu ke konkrétnímu objektu.
- b) *Identifikační proměnné*, jejichž obsahem je charakterizace dotazovaného, jeho zařazení do různých zkoumaných sociálních skupin.

Je zřejmé, že pro zachycení obsahu dat je klíčovým místem metainformační

popis meritorních proměnných prostřednictvím vhodně koncipovaného *registru*. Tento registr meritorních proměnných popisuje jak obsah proměnných, tak jejich další (dále uvedené) charakteristiky.

Na druhé straně identifikační údaje se vyskytují víceméně *standardně a opakovaně*. Pro popis tohoto typu objektů byl zvolen jako vhodná forma *katalog (kartotéka) identifikačních proměnných*.

Vymezení základních objektů metainformačního popisu (metadat) ukazuje na poměrně složitou strukturu jak metainformační základny, tak i koncepcí vlastní informace o veřejném mínění. Výše uvedené objekty by bylo samozřejmě možno doplnit dalšími. Z možných typů objektů se například nabízí *soustava ukazatelů veřejného mínění*.

Za určitých dodatečných podmínek, stojících víceméně mimo banku dat (resp. mimo vlastní řešení banky dat), spojených se zkvalitněním výzkumného procesu, rozvojem sekundárních analýz a teorie veřejného mínění, je možno soustavu ukazatelů veřejného mínění chápat nejméně ve dvou úrovních jako:

- a) Určitou soustavu relativně standardizovaných proměnných (dotazů) zjišťujících zpravidla opakovaně podstatné názorové struktury populace ČSSR. Takováto soustava metodologicky zhodnocených a ověřených dotazů (i konstrukcí a indexů) by měla velký efekt pro zkvalitnění celého poznávacího procesu, pro tvorbu informací o veřejném mínění. Zvláštní význam by pak spočíval v možnosti sledovat různé typy změn mínění v čase, v různých sociálně politických situacích apod.
- b) Specifickou soustavu ukazatelů operacionalizujících různé strukturální vlastnosti veřejného mínění jako sociálního jevu. Takovýto přístup by umožnil překročit horizont atomizovaného, mozaikového pohledu na veřejné mínění směrem k zobecňujícímu poznávání jeho relevantních vlastností. Tvorba této soustavy ukazatelů je také podmíněna (a ještě výrazněji než v případě ad a)) teoretickým rozvojem sociologie veřejného mínění.

Dalším typem objektů metainformační základny, souvisejících s vlastními daty, jsou tabelace operativního fondu dat a *primární i sekundární dokumenty*, které obsahují sociologicko-statistické fakty (a někdy i vyšší typ zobecnění) získané z uvedených dat. Jejich obsah je vždy ohraničen obsahem proměnných dané matice dat (resp. 1–2 matic dat). Registr těchto dokumentů je propojen vhodným způsobem s registrem proměnných.

V dalším období fungování a rozvoje banky dat výzkumu veřejného mínění bude nutno mezi objekty metainformačního popisu zařadit také *sekundární dokumenty* vznikající v procesu využití banky dat.

Jiný typ potenciálních objektů metainformačního popisu je úzce spojen s převládající praxí výzkumu veřejného mínění, kdy sběr dat je prováděn prostřednictvím *tazatelské sítě*. Je zřejmé, že komplexní přístup k využití automatizace ve výzkumu veřejného mínění zahrne i propojení na existující *bázi dat o tazatelích*.

Dalšími potenciálními objekty metainformačního popisu, s nimiž se uvažuje v dalším rozvoji banky dat, jsou například externí výzkumy, výzkumné zprávy jiných institucí, články či statistické údaje apod.

Informační a metainformační základna je propojena prostřednictvím různých vazeb i prostřednictvím jednotlivých *informačních procesů* různého charakteru. Struktura a založené informační vazby zabezpečují ve svém souhrnu možnost realizace procesů, jejichž konečným efektem jsou služby banky dat.

Metainformační popis zahrnuje poměrně široké spektrum typů objektů. Pro provoz banky dat však zaujímají klíčové místo metainformace, popisující da-

tovou složku informační základny. Na základě dosavadního rozboru i praktické zkušenosti s realizací banky dat je možno říci, že *hlavním článkem metainformací jsou popisy výzkumných souborů a meritorních proměnných*. Právě tyto soubory údajů — metadata — totiž umožňují za určitých podmínek zprostředkovat operativní údaje o veřejném mínění a uspokojovat tak převážnou část informačních požadavků. (8)

V bance dat výzkumů veřejného mínění bylo pro první etapu fungování zvoleno pro popis výzkumných souborů více než 60 nejrozličnějších položek metainformačního charakteru (údaje o výběru, metodě, nákladech apod.).

Při koncipování obsahu metainformačního popisu meritorních proměnných bylo nutno v první řadě zabezpečit propojení na vlastní bázi dat. Řešení tohoto problému se zakládá zejména na použití shodného pojmenování jednotlivých proměnných v rámci metainformací i báze dat. Například v bance dat výzkumů veřejného mínění byla použita standardní jména popisu, skládající se ze jména výzkumného souboru (jedinečného v rámci konkrétního souboru dat. Na identifikační údaje navazují další metodické charakteristiky, údaje o obsahu proměnné i data o vazbě proměnné na ostatní. *Klíčové a nejobtížnější místo metainformačního popisu proměnných spočívá ve vhodné volbě popisu obsahu proměnné*, kde je nutno uplatnit specifické informatické poznatky o tvorbě klasifikačních a selekčních systémů.

V dosavadní výzkumné praxi ÚVVM bylo již tradicí zařazovat jednotlivé zjišťované otázky podle určitého klasifikačního systému polytematického charakteru. Každým rokem byl jako nástroj popisu získaných dat vydáván *Rejstřík otázek*, který zařazoval každou ze zkoumaných otázek do některé třídy klasifikace. Takovými třídami klasifikace byly například: „světový názor obecně“, „názory na socialismus a kapitalismus“, „efektivnost a kvalita“, „vztah k práci a povolání“ apod. Tematická struktura, stanovené třídy — tématata, v podstatě vycházely ze snahy pokrýt všechny zkoumané oblasti. Jak však ukázaly předprojektové analýzy i údaje v informatické literatuře, není takováto klasifikace pro použití v rámci automatizovaného systému nejvhodnější.

Při volbě konkrétního způsobu popisu obsahu meritorních proměnných bylo nutno vzít do úvahy i širší důsledky zvoleného řešení. Způsob zachycení obsahu proměnných je zároveň důležitým *komunikačním prostředkem* pro klasifikaci a posléze i selekce údajů. Tento *selekční jazyk* musí být dostatečně *flexibilní*, reagující na různé cíle, resp. zkoumané problémy, které se mohou v průběhu doby měnit. Musí být uzpůsoben hlavním typům uživatelů — zejména obsluze banky dat a i dalším okruhům uživatelů tak, aby se ho mohli naučit a efektivně využívat. Zároveň musí pokrýt široký, polytematický záběr faktografického i dokumentografického fondu.

Jedním z možných řešení je využití makroklasifikačního třídícího systému s využitím dvou selekčních jazyků na různých hierarchických úrovních. Prvá hierarchická rovina odpovídá RUBRIKÁTORU — tematickému rejstříku, navazujícímu na dříve používané klasifikace. Vznikl jak převzetím určitých rubrik nebo oblastí, tak frekvenčním zpracováním používaných pojmů. Rubrikátor umožňuje zařadit každou proměnnou do jedné nebo několika rubrik — témat. Podrobnější vymezení obsahu proměnné spočívá na systematickém využití TEZAURU. (9)

(8) V současné době jsou na discích počítače uloženy metainformační údaje o cca 60 výzkumech veřejného mínění, provedených ÚVVM při FSÚ. Jsou zde uloženy údaje jak o výzkumech jako celku, tak o jednotlivých zkoumaných proměnných, znacích. Ve fondech je popsáno kolem 8,5 tisíce proměnných. Metainformační údaje jsou neustále přístupné prostřednictvím terminálu ÚVVM.

Při výstavbě banky dat z výzkumů veřejného mínění vytvořil výzkumný tým aposteriorní tezaurus na základě frekvenční analýzy pojmů vyskytujících se v dotaznících nebo indikovaných položkami dotazníku. Šlo tedy o statistický způsob stanovení deskriptorů (klíčových slov) tezauru, zabrazujících tematickou strukturu daného faktografického a dokumentografického fondu.

Funkce takto koncipovaného rubrikátoru a tezauru jsou spojeny jak s vytvořením *sémantické definice proměnných* (tj. přiřazení rubrik a deskriptorů konkrétní proměnné), tak s následným *vyhledáváním relevantních údajů*.

Sémantická definice proměnných je nejdůležitějším momentem metainformačního popisu, proto se ji také pokusíme podrobněji přiblížit. Formální struktura *sémantické definice* spočívá ve *výčtu deskriptorů (klíčových slov)* a zařazení do určitého *tématu*. Tento proces, označovaný shodně s praxí informatiky jako *indexování*, je samozřejmě náročný na intelektuální činnost. Řešitelský kolektiv zvažoval různé verze automatizace této činnosti, došel však k názoru, že za daných zkušeností s využíváním banky dat a tohoto postupu není možno výrazněji indexování automatizovat. Je však možné jej určitým způsobem objektivizovat, aby nebyl závislý jen na zkušenostech jednotlivce. (10)

V souvislosti s popisem obsahu proměnné je nutno upozornit na vážný metodologický problém. Konkrétní obsah zkoumaných proměnných výzkumů veřejného mínění má z různých důvodů (styl operacionalizace, realita fungování veřejného mínění a způsob získávání údajů apod.) nejméně následující úrovně obsahu:

- a) *Bezprostřední manifestovaný obsah*, zpravidla vyjádřený použitými výrazy běžného jazyka. Tyto výrazy odpovídají určitému pojetí sociální reality, resp. chápání termínů a sociální zkušenosti populace. Jsou přiřaditelné k určitým obecnějším termínům. Tento bezprostřední manifestovaný obsah je zpravidla spojen s vlastním textem otázek dotazníků (včetně textu odpovědi).
- b) *Zprostředkovaný indikovaný obsah*, který lze vyjádřit prostřednictvím obecných termínů jazyka sociologie. Tento zprostředkovaný indikovaný obsah má nejméně tyto aspekty:
 - *co jsme chtěli zjistit* — tedy obsah, který jsme v procesu operacionalizace projektovali
 - *co z toho jsme skutečně zjistili* — jaká data jsme vzhledem k situaci dotazování skutečně získali
 - *jak jsme údaje použili při interpretaci*
 - *o čem ještě proměnná, resp. data vypovídají*, jaké má daná proměnná další indikační schopnosti.

(9) Konkrétní návrh tezauru banky dat výzkumů veřejného mínění připravili ss. JUDr. V. Švestka, RNDr. M. Königová, CSc. a PhDr. E. Krátká. V závislosti na obsahu fondů je každoročně inovován.

(10) Dosud využíval selekční jazyk banky dat výzkumů veřejného mínění 47 témat. Mezi nejčastější témata patří: „IF01 — údaje o dotázaných“, „HS00 — hromadné sdělovací prostředky“, „OS00 — Obchod, služby“, „RS01 — řízení socialistického hospodářství, plánování a organizace práce“, „ZZ01 — socialistický způsob života“, „SN01 — světový názor“, „PZ03 — poměr k práci, pracovní iniciativa, vztahy na pracovišti“. Každé téma je označeno příslušným kódovým označením. Uvedená témata byla použita ve více než 300 proměnných.

Tezaurus, používaný při indexování proměnných, zahrnuje kolem 850 deskriptorů. Mezi nejvíce užívaná klíčová slova patří např.: faktor, práce, informace, život, lidé, členství, socialismus, význam, kvalita, vývoj, vztah, úroveň, společnost, hodnota, důvod apod. Každý z uvedených deskriptorů byl použit více než 200krát.

K *sémantickému* popisu proměnné zpravidla postačují 1–2 témata a 5–7 klíčových slov.

Vytvoření sémantického popisu proměnné musí určitým způsobem respektovat uvedené úrovně obsahu proměnné zejména v souvislosti s různým kontextem použití. Je také nutno upozornit na to, že z matice dat je nutno vyřadit ty proměnné, jež nenesou informaci, kterou jsme chtěli získat, resp. nenesou informaci v té kvalitě, v jaké to požadujeme. Nesmí také překvapit, že při podrobnější analýze sekundárního charakteru můžeme odhalit další obsahy, které jsou s danou proměnnou spojeny. Taková analýza pochopitelně také může ukázat scestnost předcházejícího použití proměnné.

Na sémantickou definici proměnných navazují i další metainformační údaje:

- typ proměnné (prvotní – odvozená)
- typ dotazu (otevřená otázka, uzavřená otázka, prvek baterie apod.)
- typ transformace, vytvoření odvozené proměnné (rekategorizace, shrnutí údajů několika proměnných apod.)
- předmět dotazu (zda se jedná o výpověď o chování, vlastnictví nebo o názor apod.)
- časová dimenze dotazu (zda jde o dotaz na minulost, přítomnost, budoucnost)
- použití kategorie „nevím“ při dotazu
- použití pomůcek při dotazu
- kvalita proměnné
- počet a seznam filtrujících (filtrovaných) otázek
- zařazení dotazu do baterie – seznam prvků baterie
- počet a seznam shodně formulovaných či vytvářených proměnných
- využití proměnné v závěrečné zprávě, v sekundárních dokumentech apod.

Jak je vidět ze stručného a zjednodušujícího přehledu, představuje každý metainformační údaj sám o sobě specifický metodologický problém.

Řešitelský kolektiv banky dat z výzkumů veřejného mínění podrobně analyzoval jednotlivé metainformační údaje o proměnných a zvolil strukturu zahrnující více než 30 položek a různých seznamů. Přístup k datům je tak možný například i v kontextu metodologické analýzy. Ukazuje se, že volba širokého rozsahu metainformačních údajů umožňuje mnohorozměrný přístup k datům, i když zároveň komplikuje celý systém vytvoření i uložení metainformačních údajů. Je zřejmé, že i pro strukturu metainformačních údajů o meritorních proměnných platí nutnost jejího ověření v několikaletém provozu.

Z výše uvedeného je zřejmé, že *metainformační základna banky dat tvoří zároveň specifický zdroj údajů o vlastním výzkumném procesu a jeho výsledcích, resp. souvislostech. Spolu s informační základnou tak umožňuje poskytovat široké spektrum služeb různým kategoriím uživatelů.*

V současné době (říjen 1987) zahrnují fondy banky dat cca 60 výzkumů veřejného mínění provedených ÚVVM při FSÚ. Jde o cca 60 % dosud realizovaných výzkumů tohoto ústavu.

K technologickým fondům

Na rozdíl od informační a metainformační základny, které jsou svými funkcemi, a tedy i obsahem orientovány zejména na poskytování informací o veřejném mínění, resp. jejich zprostředkování, je obsah a struktura technologických fondů podřízena konkrétnímu funkčnímu uspořádání banky dat a bezprostředně navazuje na použité programové, organizační a metodické zabezpečení banky dat.

Technologické fondy zahrnují dvě velké skupiny údajů. Jde jednak o **pro-**

covní fondy a soubory, které jsou vytvářeny v průběhu realizace informačních procesů a mají zpravidla určitý dočasný charakter, a jednak o *provozní soubory*. Ty slouží spíše jako nástroj technologie práce s daty a informacemi.

Pracovní fondy a soubory jsou vytvářeny a používány pro realizaci jednotlivých vstupních, zprostředkujících nebo výstupních operací s daty, informacemi a metainformacemi.

Všechny pracovní fondy a soubory jsou popsány v rámci metodik, postupů, v nichž jsou využívány.

Provozní soubory jsou úzce spojeny s využitím databankové technologie zpracování informací a metainformací a s konkrétním využitím technických i programových prostředků, dostupným řešitelskému kolektivu. Při budování banky dat v jiných technických podmínkách, než které byly k dispozici pro budování banky dat z výzkumů veřejného mínění ÚVVM, je pochopitelné možné, že mohou mít jinou strukturu a rozsah.

Je třeba připomenout, že při projektování automatizovaných částí banky dat je nutno vzít v úvahu i příslušné kapacitní nároky všech druhů souborů umístěných na prostředcích výpočetní techniky.

Shrnutí

Fondy banky dat výzkumů veřejného mínění se skládají z informační a metainformační základny a technologických fondů.

Informační základna je organizovaný a systematicky uspořádaný souhrn údajů a dat z výzkumů veřejného mínění.

Obsahem informační základny jsou trvale uchovávané a doplňované informace (data, fakty) o veřejném mínění občanů ČSSR. Její rozhodující část — datové báze — je umístěna na disku počítače.

Metainformační základna je organizovaným a systematicky uspořádaným souhrnem relevantních údajů o informacích, o prvcích a souborech informační základny a o vybraných článcích výzkumného a informačního procesu. Její obsah a uspořádání vyplývá z potřeby zabezpečit potenciálně rozmanité požadavky na přístupy k informacím o veřejném mínění. Metainformační údaje popisují jak výsledek výzkumného procesu, tak některé jeho základní prvky.

Obsahem metainformační základny jsou trvale uchovávané, doplňované a aktualizované informace o datech a výzkumu veřejného mínění. Její všechny části jsou umístěny na trvale přístupném disku počítače CYBER 172.

Technologické fondy jsou tvořeny pracovními a provozními soubory.

Jejich obsah, forma a umístění vyplývá z jejich funkce v informačních procesech banky dat.

Literatura

- Afanasjev, V. G.: *Sociální informace a řízení společnosti*. Praha, Svoboda 1978.
Andrejenkov, V. G. — Čeredničenko, V. A.: *K voprosu o sozdaniji banka sociologičeskoj informacii*. Sociologičeskiye issledovanija 1982, č. 1.
Bisco, R. L.: *Data bases, Computers and the Social Sciences*. New York, Wiley-Interscience 1970.
Cejpek, J. — Delikát, I.: *Automatizovaný systém právních informací. Některé zkušenosti z projektování a pokusné realizace*. Čs. informatika 1981, č. 1.
Cejpek, J. — Holina, V. — Kohlman, J.: *Perspektiva automatizovaného systému novinářských informací*. Bratislava, SÚTI 1982.
Compstat: *Proceedings in computational statistics — 6. symposium*.
Čížek, F.: *Teorie a empirie*. Praha, Svoboda 1974.
Databázový systém AŠIS (všeobecná část technického projektu). Praha, FSÚ-VVS 1983.

- Dvořák, P.: *Vybrané metodologické problémy výstavby banky dat výzkumů veřejného mínění*. Kandidátská disertační práce 1985.
- Dvořák, P.: *Ke koncepci banky dat výzkumů veřejného mínění*. Sociologický časopis 1986, č. 6.
- Charvát, F.: *Sociální vztahy a sociologické proměnné, metodologická úvaha nad koncepty stratifikace a mobility*. Sociologický časopis 1974, č. 6.
- Chvalovský, V.: *Banky dat*. Praha, SNTL 1984.
- Illner, M.: *K možnosti vytvoření archivu dat z výzkumů veřejného mínění (interní materiál)*. Praha, ÚVVM 1977.
- Illner, M. — Foret, M.: *Sociální ukazatele*. Praha, Svoboda 1980.
- Illner, M.: *Faktografické informace v sociologii*. In: Informatika v sociologii. Pardubice, Dům techniky ČSVTS 1982.
- Kovář, B.: *Věcné pořádání informací a selekční jazyky*. Praha, UVTEI 1981–1982.
- Krjučkova, L. I.: *Opyt sozdanija i funkcionirovanija bankov sociologičeskich dan-nych*. In: Problemy pererabotki i avtomatizirovannogo poiska sociologičeskoj informacii. Moskva, ISI AN SSSR 1977.
- Melišková, J. — Fischer, J.: *K problematice budování statistické banky dat*. Statistika 1981, č. 12.
- Melišková, J. — Fischer, J.: *Přístupy k budování metainformačního systému v československé statistice*. Statistika 1982, č. 2.
- Mezšároš, L.: *K problematice integrácie dát*. Statistika 1979, č. 11.
- Molčanov, V. I.: *Systemnyj analiz sociologičeskoj informacii*. Moskva, Nauka 1980.
- Řehák, J.: *Centrální evidence a archivace výzkumů veřejného mínění (interní materiál)*. Praha, ÚVVM 1977.
- Řehák, J.: *Data, statistika a sociologie*. Sociologický časopis 1981, č. 3.
- Řehák, J.: *Uživatelský pohled na programy pro zpracování sociologických dat*. In: Informatika v sociologii. Pardubice, Dům techniky ČSVTS 1982.
- Smetáček, V.: *Trendy ve vývoji selekčních jazyků*. Praha, UVTEI 1982.
- Sundgren, B.: *Metainformácie v štatistických úradoch*. Informačné systémy 1978, č. 3.
- Šoltéz, D.: *Informačný systém o štatistickom informačnom systéme — základné vymedzenie a funkcie*. Statistika 1979, č. 6.
- Šoltéz, D.: *Bázy metadat — katalogy, slovníky a adresáře dat*. Informačné systémy 1980, č. 6.
- Šoltéz, D.: *Metainformačný systém a navrhovanie dátovej zložky štatistického informačného systému*. Informačné systémy 1984, č. 3.
- Veselý, J.: *Informační metasystém banky dat a jeho význam*. Organizace a řízení 1984, č. 4.
- Zich, F.: *K problematice charakteru empirických dat v sociologickém výzkumu*. Sociologický časopis 1977, č. 2.
- Zich, F.: *Poznámky k problematice tvorby informačního systému o veřejném mínění pro potřeby politické praxe (interní materiál)*. Praha, ÚVVM 1979.

Резюме

П. Дворжак: К методическим вопросам создания фондов банка данных по обследованиям общественного мнения

В статье рассматривается проблематика создания фондов банка данных по обследованиям общественного мнения. Во вводной части автор характеризует решение, приведшее к разделению банка данных на три основные части — технологический базис, кадровое обеспечение деятельности и организационный порядок. Фонды банка данных являются составной частью технологического базиса и их можно разделить на информационный базис, метainформационный базис и технологические фонды. Информационный базис является организованной и упорядоченной совокупностью данных по обследованиям общественного мнения, его решающей частью являются данные отдельных обследований, оперативный фонд агрегированных данных и архив заключительных отчетов. В статье также рассматриваются вопросы источников данных и включения экзогенных данных в информационный базис.

Метainформационный базис является совокупностью данных об элементах и массивах информационного базиса. Его содержание и упорядочение вытекает из потребности в обеспечении разнообразных запросов по подходу к данным. В статье приводятся функции метainформации и ее примененные организационные формы. Далее здесь описываются вопросы создания метаданных, т. е. информации о сохраненных данных, создающих основу системы. Особое внимание уделяется описанию содержания сохраняемых переменных, несущих информацию о мнениях граждан. Приводится способ формирования семантического описания переменных с использованием информационно-поискового языка.

В заключительной части кратко характеризуются технологические фонды банка данных, вытекающие из примененных технических, программных и методических средств для создания банка данных.

Summary

P. Dvořák: On Methodical Problems of Data-bank Formation in the Area of Public-opinion Research

The paper deals with problems connected with the formation of data-bank funds in public-opinion research. Its introductory part characterizes the solution leading to the break-up of the data bank into three basic parts — technological basis, staff of employees to secure the activities involved, organizational arrangement. Data-bank funds are a part of the technological basis and can be divided into the information basis, the metainformation basis, and technological funds. The information basis is an organized and ordered complex of data obtained from public-opinion research; its decisive component are primary research data the operative fund of aggregated data, and archives of final reports. The paper also discusses issues of the sources of data and the incorporation of exogenous data in the information basis.

The metainformation basis is a complex of data on elements and sets included in the information basis. Its content and arrangement are given by the necessity to satisfy the manifold requirements for access to data. The paper points out the functions of metainformation and their employed organizational forms. It goes on to describe issues of the formation of metadata, i. e. information on stored data forming the basis of the system. Particular attention is devoted to the description of the content of stored variables bearing information about the citizens' views. The way of forming the semantic description of variables with the use of a selective language is referred to.

The concluding part brings a brief characterization of technological data-bank funds ensuing from the applied technical, programmatic and methodical means for the data-bank formation.