

# Pro nás sčítání v dubnu neskočilo, říká Josef Škrabal, ředitel odboru statistiky obyvatelstva ČSÚ

***Jste ředitelem odboru statistiky obyvatelstva. Kdy jste nastoupil na statistický úřad a jaká byla vaše cesta na místo ředitele?***

Na statistickém úřadě jsem začal pracovat v roce 1979. Nastoupil jsem do oddělení statistiky obyvatelstva, kde jsem byl až do vojny. Po absolvování základní vojenské služby jsem nastoupil do oddělení sčítání a podílel se na zpracování výsledků sčítání v roce 1980. V druhé polovině 80. let jsem několik let dělal statistiku práce, ale jinak jsem až do roku 2005 pracoval v oddělení sčítání lidu. V roce 2005 jsem po ředitelce Jiřině Růžkové převzal odbor statistiky obyvatelstva, do kterého kromě sčítání patří demografická statistika, a od minulého roku také volební statistika.

***Aktivně se podílište už na čtvrtém sčítání. Sčítání je obrovská akce, ať už probíhá v roce 1980 nebo 2011, je s ní spojeno mnoho zážitků. Zkuste vybrat jednu konkrétní věc, na kterou vzpomínáte nejvíce.***

Nejvíc vzpomínám hned na to první sčítání v roce 1980, kdy jsem na základní vojenské službě dělal sčítacího komisaře a byl to opravdu velký zážitek. Byl jsem na vojně, v celém Československu se připravovalo sčítání lidu a armádě neušlo, že v civilu jsem pracovníkem statistického úřadu. Funkce sčítacího komisaře pak pro mě byla jasná. Dostal jsem 14 dní bez služeb a celou tuto dobu jsem měl na zabezpečení toho, aby všichni vojáci základní služby našeho malého útvaru opravdu vyplnili sčítací formuláře. To je jedna z těch věcí, na které se nezapomíná.

***30 let není krátká doba, kam se za tu dobu sčítání posunulo?***

Především na technice je znát, jak se sčítání vyvíjí, ale na druhou stranu je třeba říci, že už v roce 1980 jsme výsledky sčítání zpracovávali na poměrně výkonném počítači. Ovšem proti poslednímu sčítání těch změn zase tolik není. Stejně jako letos se sčítací formuláře skenovaly i v roce 2001, i když skenery byly méně výkonné. Mnozí také kritizují, že i v dnešní době se musí vyplňovat papírové formuláře jako za Marie Terezie. Proto také za největší pokrok letošního sčítání považuji možnost se sečíst elektronicky a musím říct, že tento způsob vyplňování se velmi osvědčil. Toto je cesta, kterou bychom se měli ubírat. Stejně tak by bylo potřebné přebírat více údajů z administrativních zdrojů a registrů, které tu v tuto chvíli ale neexistují nebo nejsou pro statistiku přístupné. Nějaké údaje z registrů jsme i při letošním sčítání samozřejmě využili, například základní údaje z evidence obyvatel, ale nebylo jich mnoho, protože více jich v evidenci není. Chybí nám například registry ekonomické aktivity nebo vzdělání, které by nám velmi pomohly. Pokud se v tomto státě nezmění přístup k vytváření a využívání registrů, tak se obávám, že příští sčítání za deset let bude z hlediska provedení velmi podobné tomu letošnímu. Doufám ale, že situace se změní a další sčítání bude tedy jiné.



***Základem pro změnu způsobu sčítání jsou tedy funkční registry, které v současnosti neexistují?***

Přesně tak, k tomu ale musí být politická vůle. Náš úřad si sám nedokáže prosadit zřízení registrů, to si zde všichni uvědomujeme. Sám jsem to velmi pocítil, když jsme připravovali zákon o sčítání. Projednávali jsme různé možnosti, jak případně využít výsledků sčítání 2011 pro založení registrů, které by bylo možné v budoucnu využívat, ale žádnou velkou podporu jsme nenašli. Funkční registry jsou naprosto zásadní, pokud se má snížit rozsah zjišťování u obyvatelstva a zejména na tomto rozhodnutí bude záležet, jak bude příští sčítání vypadat. Jsem přesvědčen, že nás čeká dlouhá a nelehká cesta.

***Určitě můžeme hledat inspiraci v jiných zemích, kde mají s registry zkušenosti a aktivně je používají. Zkuste zmínit nějaké konkrétní země.***

Určitě můžu uvést skandinávské země, které jsou v tuto chvíli opravdu jinde. Jejich registry existují více než 30 let a v základních věcech fungují bezchybně. Nejde jen o to, že mají registr obyvatel, ale mají fungující soustavy registrů, to je důležité. Pokud registry fungují, pak je možné informace z nich plně převzít nebo je využít alespoň částečně a pro zbylé

údaje provést výběrové šetření na menší části obyvatelstva. Tato cesta je možná, ale základ je ve fungujících registrech. Ke stejnému řešení jednou dospějeme i my, ať se to někomu líbí nebo ne. Situace je taková, že stále více zemí se snaží využívat administrativní zdroje. My jsme v letošním sčítání udělali takový první krok předvyplněním některých formulářů. Využili jsme registru obyvatel a registru sčítacích obvodů ČSÚ. Domovní listy jsme předvyplnili údaji z minulého sčítání, u sčítacích listů osob se zase jednalo o předtištění adresní části formulářů. A některé údaje o osobách jsme nezjišťovali, neboť je převzeme z evidence obyvatel.

### **Zpátky jsme se dostali k letošnímu sčítání. Jaká je tedy vaše role v této celonárodní akci?**

Mám dvě pozice. Jednak jsem věcným gestorem projektu sčítání. To znamená, že jsem tím, kdo definuje věcné požadavky na projekt, zodpovídá za věcnou správnost provádění sčítání, ať již jde o logistické či metodické věci, ale také za správnost věcné stránky objednávek a veřejných zakázek. Toto sčítání je postaveno na tom, že rozhodující část prací je řešena veřejnými zakázkami. Druhou rolí je ředitel odboru statistiky obyvatelstva a opravdu nevím, která z nich mě v současné chvíli více zatěžuje. Pod sebou mám hodně přes 400 lidí a s tím je spojeno mnoho povinností v personální a mzdové oblasti. V některých obdobích mi to zabírá spoustu času. Například ke konci června má z našeho odboru skončit více než 300 zaměstnanců a s tím bude spojeno velké množství administrativní práce.

### **Určitě vás čekají i další činnosti spojené se sčítáním. V dubnu skončila fáze distribuce a sběru sčítacích formulářů, ale sčítání pokračuje dál.**

Sčítání rozhodně neskončilo. Teď skončila pouze nejviditelnější část, která se dotýkala každého občana. V současné chvíli nastává období činností, které už nejsou tak vidět, ale jsou stejně důležité. Teď běží první fáze zpracování výsledků, do které patří skenování sčítacích formulářů a proces validace neboli rozpoznávání údajů na sčítacích formulářích.

### **Kolik se má celkově naskenovat formulářů?**

V tuto chvíli je sebráno více než 12,5 milionů sčítacích formulářů, které musí být naskenovány. Formuláře nám ale stále ještě chodí, protože v druhé polovině května proběhlo došetřování. To se týká domácností, které sčítací formuláře převzaly, ale neodevzdaly. Desítky tisíc, možná až 100 tis. formulářů nám tedy ještě přijde.

### **Bude se došetřování týkat všech domácností, které z nějakého důvodu nevrátily sčítací formuláře?**

Došetřování se bude týkat podstatného podílu těchto domácností. Pokusíme se je navštívit a formuláře od nich získat. Rozpočet sčítání není neomezený, i když to podle některých novinových titulků může tak vypadat, a nemáme takové kapacity, abychom mohli navštívit každou domácnost, ze které nám nepřišel nějaký formulář. Je to kompromis, kdy na jedné straně máme domácnosti, o kterých víme, že mají neodevzdané formuláře, a na druhé straně máme mzdové prostředky, které můžeme použít na tuto akci.

### **Do konce června budou naskenovány všechny papírové formuláře, včetně těch získaných v rámci došetření. Pak už bude možné přistoupit ke zpracování výsledků?**

Ještě ne. Současné se skenováním probíhá validace. Skenery udělají obrazy formuláře a většinu znaků na něm bezpečně rozpoznají. Máme

ale velké množství případů, kdy skener úplně jasně nerozpozná, jaká je to číslice nebo písmeno. V těchto případech je třeba tyto znaky posoudit, a to je práce validátorů. Ti si po naskenování zobrazují formuláře, kde mají označeno, který znak skener nepřečetl nebo si u něj není jistý. Validace v současné chvíli probíhá a ukazuje se, že je o něco pomalejší, než jsme předpokládali. Termín ukončení této etapy je do konce srpna a stále věřím, že se stihne.

### **Jak je to s validací u elektronických formulářů? S těmi by problémy být neměly.**

U nich tyto problémy skutečně nejsou, a to je velká výhoda. Už teď můžeme říci, že zápisy v elektronických formulářích jsou přesnější a kvalitnější než v papírových formulářích.

### **Jak vlastně vy hodnotíte vyplnění formulářů přes internet? Kolik lidí je vyplnilo?**

Pro mě je to určitě úspěch, já jsem patřil spíš k pesimistům. I když srovnání s obdobím před deseti lety není úplně ideální, u většiny států, které elektronické vyplnění sčítacích formulářů využily při sčítání v roce 2001, se výsledky pohybovaly pouze v jednotkách procent. Za deset let se situace významně změnila, pro mě je těch 26 % formulářů sebraných přes internet potěšitelné číslo.

### **Přesuňme se na začátek září. Je po validaci, jaký další krok sčítání čeká?**

Ještě musím uvést jednu důležitou etapu, která musí proběhnout, a tou je kódování. Na rozdíl od odstraňování duplicit nebo anonymizace není tato etapa vázána na konec nějaké předcházející, ale probíhá z velké části souběžně s validací. Elektronické i validované papírové formuláře obsahují slovní zápisy, které musíme zakódovat. To proběhne ve většině případů automatizovaně. Ovšem například u adres, národností nebo zaměstnání je rozmanitost zápisů taková, že i když používáme slovníky a číselníky, tak leckteré zápisy se v nich prostě nenajdou a pak nastává ruční práce. Pro další zpracování se databáze údajů z elektronických a papírových formulářů sloučí do jedné. Za každý formulář budeme mít v databázi jednu větu.

### **A co kdyby někdo vyplnil jak papírový, tak elektronický formulář?**

Právě proto je dalším krokem zpracování porovnání těchto vět a nalezení duplicitních záznamů. Je opravdu velmi pravděpodobné, že se někdo sečetl nebo byl sečen dvakrát.

Například vysokoškolský student vyplnil sčítací formulář na koleji, kde bydlí, ale jeho rodiče za něj vyplní formulář i doma, kde je přihlášen k trvalému pobytu. Při nalezení duplicitních záznamů je třeba rozlišit, který z nich je kvalitněji vyplněn.

Počítáme i s možností spojení těchto duplicitních záznamů, abychom měli co nejúplnější údaje. V této fázi budeme ještě využívat evidenci obyvatel, ze které se budou k existujícím záznamům dotahovat některé údaje, které jsme nezjišťovali, protože jsou v evidenci k dispozici. Podobné to bude s údaji za domy a byty, u kterých také využijeme existující administrativní zdroje. U nich proběhnou rovněž kontroly na údaje z městské a obecní statistiky. Tento proces je velmi náročný a jsme sami zvědaví, nakolik bude mít automatizovaný průběh úspěch a kolik nám zůstane ruční práce. Pak následuje ještě etapa logických kontrol nebo kontrol přípustnosti, zda se tam vyskytují hodnoty, které jsou přípustné a reálné.





**Po odstranění duplicit budou už záznamy připraveny na odstranění osobních údajů?**

V podstatě ano. Před tím ale bude ještě jedna etapa, o které bych se rád zmínil. Na rozdíl od sčítání v roce 2001 bude potřeba přiřadit jednotlivé osoby do příslušných bytů a byty do domů.

Minule nám sčítací komisaři vraceli vyplněné formuláře pěkně pohromadě. Teď mohl každý člen domácnosti vyplnit svůj formulář jiným způsobem. Až všechny zařadíme, kam patří, pak opravdu nastává fáze anonymizace. Při té se odstraní datum narození, rodné číslo, jméno a příjmení. V dalším zpracování se budou používat pouze anonymizované údaje, kde už nebude možné konkrétní osobu dohledat. To bude závěr první etapy zpracování, která probíhá na režimovém pracovišti za zprísněných bezpečnostních podmínek. Tato etapa by měla být ukončena ještě v letošním roce. Výstupy budou poté zpracovávány z anonymizovaných údajů a tam už bezpečnost našťastí nemusí být tak vysoká. Druhá etapa už bude probíhat přímo v ČSÚ.

**Kdy se můžeme těšit na první výsledky ze sčítání 2011?**

Slibili jsme, že předběžné výsledky zveřejníme do konce letošního roku. Proti minulosti budou předběžné výsledky trochu později. Tentokrát je totiž jiný způsob odevzdávání formulářů, než byl minule, kdy to bylo jednodušší. Sumarizací přes sčítací komisaře jsme se dostali k předběžným výsledkům za republiku poměrně rychle. To by teď nemělo smysl, neboť komisaři sebrali tentokrát velmi málo formulářů.

Předběžné výsledky můžeme zpracovat až v okamžiku, kdy máme všechno naskenováno, rozpoznáno, bez duplicit. Po provedení základních kontrol pak teprve můžeme dát dohromady pár základních údajů.

**Co budou předběžné výsledky obsahovat za údaje?**

Nebudou to žádné detailní tabulky, to určitě ne. Za osoby zveřejníme přehledy obyvatelstva podle pohlaví, základních věkových skupin, stupně vzdělání, několik údajů z ekonomické aktivity.

Určitě hodně zajímavé budou údaje o národnosti a náboženském vyznání. U bytů bude k dispozici právní důvod užívání, vybavení koupelnou, WC, způsob vytápění, vybavenost počítačem a internetem. Za domy samozřejmě jejich celkový počet, z toho domy neobydlené, a údaje o vlastníkově domu.

Údaje budou k dispozici od úrovně správních obvodů obcí s rozšířenou působností výše. Ještě je potřeba říci, že předběžné výsledky budou za trvale bydlící obyvatelstvo. Je to dáno tím, že se vytvářejí v určité etapě zpracování výsledků. Definitivní výsledky však budou za obvykle bydlící obyvatelstvo.

**Kdy můžeme očekávat zveřejnění definitivních výsledků SLDB?**

Předpokládáme, že do konce roku 2012 bude uveřejněna rozhodující část definitivních výsledků. Budou to standardní výstupy za obyvatelstvo, domy, byty nebo domácnosti. Na konec se zpracovávají údaje o dojíždě do zaměstnání. Toto zpracování je svým způsobem mnohem náročnější, protože se zjišťují například dojížděkové proudy. Tato část by měla být dokončena začátkem roku 2013.

Děkuji za rozhovor.