Tisková zpráva

5. května 2021

**Život sčítacího formuláře v deseti krocích: od tisku až po skartaci**

**Od chvíle, kdy listinné sčítací formuláře opouštějí zdi tiskárny, je čeká dlouhá a náročná cesta, na jejímž konci stojí kompletní znehodnocení v podobě skartace. Se získanými daty se pracuje výhradně v digitální podobě a naprostá většina úkonů probíhá automatizovaně bez zásahu člověka. Pro správné zpracování je ale třeba dát si pozor na pravdivost a čitelnost uvedených údajů.**

1. **Tisk**

*„Pro letošní sčítání bylo vytisknuto celkem 3,8 milionu listinných formulářů. Naprostou většinu z nich tvořily formuláře pro domácnost, zbytek pak dodatečné formuláře pro početné domácnosti a samostatné formuláře pro osobu. Tisk, ořez a sešití do výsledné podoby sešitu zabraly téměř 6 měsíců,“* říká Robert Šanda, ředitel odboru statistiky obyvatelstva. Barevnost, grafické prvky, podoba a rozmístění polí pro zápis údajů, druh papíru, použité barvy a technologie tisku jsou všechny voleny tak, aby bylo možné formulář bez problémů vyplnit (např. bez rozpíjení a rozmazávání zápisů) a později naskenovat.

1. **Roznos formulářů**

Deset tisíc sčítacích komisařů formuláře následně ve dvou kolech rozneslo do domácností po celé České republice. Vzhledem k protiepidemickým opatřením komisaři formuláře předávají před domy a do vnitřních prostor budov tak vůbec nevstupují.

1. **Vyplnění formulářů**

Respondenti mají na vyplnění formulářů čas až do 11. května. Pro následné skenování formuláře je třeba údaje vyplnit černou nebo modrou propiskou. Pokud dojde při vyplňování listinného sčítacího formuláře k chybě, je nutné příslušná pole začernit a správnou odpověď následně zaznamenat do zbývajících bílých polí pro příslušnou odpověď. V případě poškození formuláře je třeba vyplnit formulář nový. K dnešnímu dni Český statistický úřad obdržel 310 tisíc vyplněných listinných formulářů. Obyvatelé tak dávají zcela jednoznačně najevo, že preferují online sečtení.

1. **Sběr a digitalizace formulářů**

Vyplněné formuláře už komisaři nepřebírají, respondenti je totiž zasílají v odpovědní obálce poštou nebo osobně odevzdávají na kontaktních místech sčítání. Formuláře jsou následně převáženy na sběrné místo, odkud co nejdříve putují rovnou na pracoviště digitalizace. Na digitalizačním pracovišti už na formuláře čekají velkokapacitní skenery. Ze skenovaných obrazů se rozpoznávají údaje, které obyvatelé do formulářů zapsali. Ne vždy jsou ale všechny znaky jasně čitelné. Na řadu tak přicházejí pracovníci digitalizace, kteří se postarají, aby byly všechny údaje přepsány do databáze tak, jak je respondenti vyplnili.

1. **Kódování**

Dalším krokem je kódování, při kterém se identifikuje význam odpovědí a různě formulované odpovědi se převedou do standardizované podoby. Příkladem může být otevřená otázka na národnost – ať už respondent odpoví např. „moravská“, „Moravan“, nebo „moravan“. Všechny tyto varianty budou zahrnuty pod jednu odpověď. *„Digitalizace, kódování i oprava drobných chyb, jako jsou například překlepy v názvu adresy, probíhají převážně automatizovaně. Ne všechny chyby lze však ošetřit algoritmy, a tak je občas třeba manuální opravy, při které ale naši zaměstnanci nemají přístup k celému formuláři, jen k jeho výřezu. Pouze bezchybně a čitelně vyplněný listinný formulář projde procesem automaticky, tedy bez nutnosti lidského zásahu,“* vysvětluje Robert Šanda.

1. **Konsolidace**

Následně dochází ke konsolidaci záznamů v rámci formuláře. Pokud se ve formuláři například neshoduje pořadí členů domácnosti s pořadím údajů o jednotlivých osobách, je třeba je sjednotit. Vyplněné údaje jsou poté propojeny se záznamy v registru obyvatel a dalších administrativních zdrojích.

1. **Deduplikace a prioritizace**

Z propojení s registry v předchozím kroku mimo jiné vyplyne, které osoby na formulářích chybí (ty se z registrů doplní), a naopak které osoby či domácnosti jsou sečteny vícekrát. Aby byly výsledky přesné, musí se duplicitní záznamy odstranit. Přitom je nutné nastavit soubor pravidel, podle kterých se rozhodne, který z duplicitních záznamů zachovat a který odstranit – tomuto kroku se říká prioritizace.

1. **Vytvoření domácností**

Po doplnění osob z registrů a odstranění duplicit je třeba „sestavit“ posloupnost
*osoba-domácnost-byt* a určit typ domácnosti (například manželský pár s rodičem jednoho z manželů, osamělý rodič s dítětem, třígenerační domácnost atd.). To je vcelku snadný úkol v případě, že je celá domácnost jednoznačně sečtena na jednom formuláři a má správně vyplněné vztahy mezi osobami. Složitější je to v případech, kdy bylo třeba do formuláře zasahovat při deduplikaci, a velmi náročné tam, kde se doplňovaly nesečtené osoby z registru obyvatel. Musí se také napravovat nelogické vztahy, jako třeba rodiče mladší než děti, k čemuž může dojít například kvůli nepozornosti respondentů při vyplňování vztahů v domácnosti.

1. **Anonymizace a statistické zpracování**

V rámci anonymizace se údaje přenesou do jiné databáze, avšak už bez rodných čísel, jmen, příjmení, typů a čísel občanských průkazů nebo ID datových schránek. Z důvodu ochrany osobních údajů se všechny vazby mezi záznamy od tohoto kroku udržují pouze ve formě relací za pomoci bezvýznamových klíčů, a už je tak nebude možné spojit s konkrétní osobou. Následně budou data z původní databáze s výjimkou údajů stanovených zákonem o sčítání kompletně vymazána. Dále se již pracuje pouze s anonymními záznamy. Provádějí se kontroly souladu jednotlivých odpovědí, aby se údaje očistily od nelogičností, které vznikly nepozorností respondentů a znehodnocovaly by výsledky (jako třeba třísetmetrové garsonky, děti na rodičovské dovolené, nezaměstnané osoby dojíždějící do práce apod.). Kombinováním zjištěných údajů se následně odvozují desítky dalších ukazatelů. Na závěr se ze záznamů za jednotlivé osoby, domácnosti a byty vytváří souhrnné údaje za sídla, obce, kraje či celou republiku. Tím vrcholí statistická práce, jejíž výsledky budou následujících deset let hrát zásadní roli při plánování dopravy, zdravotní péče nebo integrovaného záchranného systému.

1. **Skartace**

Poslední etapou v životě sčítacího formuláře je skartace, při které jsou formuláře kontrolovaně znehodnoceny. Ze vzniklé papírové drti už není možné získat jakékoli údaje, poslouží však po recyklaci při výrobě nového papíru.

Kontakt:

Jolana Voldánová

tisková mluvčí Sčítání 2021

+420 704 659 357

jolana.voldanova@scitani.cz