

PŘÍPRAVA SČÍTÁNÍ LIDU, DOMŮ A BYTŮ V ROCE 2021

Josef Škrabal

Příprava sčítání lidu, domů a bytů (SLDB) v roce 2021 vychází z usnesení vlády ČR č. 5 ze dne 13. ledna 2016 ke způsobu realizace sčítání lidu, domů a bytů v roce 2021. Koncept připravovaného SLDB je založený na maximálním využití existujících administrativních zdrojů dat, které bude doplněno obsahově redukováným terénním šetřením. Tímto usnesením vlády bylo také uloženo vybraným rezortům předat Českému statistickému úřadu (ČSÚ) požadované údaje z administrativních zdrojů a poskytnout mu metodickou a technickou pomoc při jejich zpracování. Účelem předání dat bylo vypracování analýzy konzistence existujících administrativních zdrojů dat a možnosti jejich integrace. Tato analýza je nezbytným podkladem pro ověření navržené koncepce provedení sčítání a předpokladem pro zpracování návrhu věcného záměru zákona o sčítání v roce 2021.

ČSÚ proto musel nejdříve propojit záznamy ze vzorků administrativních zdrojů a převést data ze „surového“ stavu, v jakém je získal od rezortů, do podoby vhodné pro další zpracování. Původní data byla totiž mnohdy v textových formátech, s chybějícími hlavičkami tabulek apod. Následně byly harmonizovány tabulky v rámci jednotlivých zdrojů, byly odstraněny duplicity, nejednoznačné klíče a další „porušení“ pravidel relačních databází. Byly tak vytvořeny izolované datové modely pro jednotlivé zdroje. Poté byly zkontrolovány a opraveny přirozené identifikátory. Exaktní propojení pomocí přirozených identifikátorů bylo provedeno v dalším kroku, u většiny zdrojů s vysokou mírou úspěšnosti. Dalším úkolem bylo odvození ukazatelů (věk a pohlaví z rodného čísla, příznak platnosti záznamu k referenčnímu datu 1.1.2016 apod.), a to jednak pro usnadnění dalších prací s daty, ale především pro následné pravděpodobnostní párování záznamů nepropojených přes rodné číslo. Teprve poté bylo možno spustit výpočty podobnosti atributů vhodných pro párování. Podobnost jmen, příjmení a data naro-

zení v textové podobě byla pro porovnání vypočtena pomocí dvou různých metod, aby bylo možné posoudit jejich vhodnost.

Výsledky analýzy poskytují podkladové materiály pro nutná jednání s rezorty. Jednání se vedou o řešení nejasností a nekonzistencí mezi zdroji, ale také o formě předávaných dat. Postupně jsou rozpracovávány poznatky z pravděpodobnostního párování, aby bylo dosaženo maximální přesnosti, případně budou zkoumány i modifikace pro jednotlivé zdroje, resp. specifické subpopulace (zkušenosti s párováním budou po zobecnění velmi důležité zejména při propojování administrativních dat s formuláři, kde budou pravděpodobnostní techniky hrát mnohem významnější roli).

Na základě podrobných analýz možných způsobů provedení SLDB byl zpracován návrh věcného záměru zákona o sčítání lidu domů a bytů v roce 2021. Tento návrh vychází z výše uvedeného konceptu a je v souladu s evropskými právními předpisy ke sčítání lidu a s právním řádem České republiky. Snaží se zohlednit potřeby uživatelů dat z veřejnoprávní, akademické i komerční sféry, a to jak z hlediska obsahu dat, tak jejich kvality. Současně však byl kladen zásadní důraz na maximální snížení zátěže, které sčítání představuje pro obyvatele České republiky. K tomu jsou v návrhu využity dvě cesty.

První je výrazné omezení počtu otázek na sčítacích formulářích. Datová základna reprezentovaná dostupnými administrativními zdroji sice neumožňuje zcela upustit od plošného dotazníkového šetření, umožní však významně snížit počet zjišťovaných otázek. Spolu s vyřazením dnes již méně relevantních témat lze celkový počet otázek na sčítacích formulářích snížit ve srovnání s minulým cenzen zhruba na polovinu.

Druhým, pro respondenty ještě významnějším krokem, je navržený způsob přístupu k dotazníkům, jejich vyplnění a odevzdání. Předložený návrh nabízí respondentům možnost (nikoliv povinnost) sečíst se online, a na rozdíl od minulého sčítání zcela bez nutnosti osobního či telefonického kontaktu se sčíta-

cím komisařem. Sčítací komisaři následně provedou klasické dotazníkové šetření pouze v domácnostech, které možnost sečíst se online nevyužijí.

Příprava SLDB pokračuje i na celosvětové a evropské úrovni. K provedení sčítání v roce 2021 byly připraveny tři prováděcí předpisy, navazující na obecné nařízení ke sčítání lidu (Nařízení Evropského parlamentu a Rady č. 763/2008 ze dne 9. července 2008 o sčítání lidu, domů a bytů). Všechny již prošly celým schvalovacím procesem a byly zveřejněny v Úředním věstníku EU.

Prováděcí nařízení Komise (EU) 2017/543 ze dne 22. března 2017 stanoví prováděcí pravidla k nařízení Evropského parlamentu a Rady (ES) č. 763/2008 o sčítání lidu, domů a bytů, pokud jde o technické specifikace témat a jejich rozdělení. Vedle definic jednotlivých témat jsou obsahem nařízení i jejich požadovaná rozdělení (číselníky). Konkrétně je zde uvedeno 14 témat za osoby (např. země/místo narození, rok přistěhování do země, místo obvyklého pobytu rok před sčítáním, typ ekonomické aktivity, zaměstnání, postavení v zaměstnání, místo pracoviště), 6 témat za domácnosti (např. typ a velikost soukromé domácnosti, postavení v rodině/domácnosti) a 14 témat za domy a byty (např. způsob bydlení, typ vlastnictví bytů, charakteristiky zalidnění bytu, vybavení bytu nebo domu).

Dne 20. dubna 2017 bylo přijato nařízení Komise (EU) 2017/712, kterým se zavádí referenční rok a program statistických údajů a metadat pro sčítání

lidu, domů a bytů stanovený nařízením Evropského parlamentu a Rady (ES) č. 763/2008. Jako referenční rok pro sčítání lidu, domů a bytů se určuje rok 2021. Každý členský stát musí určit pro údaje ze SLDB referenční datum, které oznámí Komisi (Eurostatu) do konce roku 2019. V nařízení je také definován program statistických údajů, jež mají být předány Komisi (Eurostatu) za referenční rok 2021. Tento program je tvořen tzv. hyperkrychlemi, což jsou vícerozměrná křížová srovnání jednotlivých rozdělení témat. Seznam hyperkrychlí je uveden v příloze tohoto nařízení. Celkový počet požadovaných hyperkrychlí je 41, v tom je 32 hyperkrychlí za osoby, 4 za domácnosti a 5 za domy a byty.

Třetím prováděcím předpisem je Prováděcí nařízení Komise (EU) 2017/881 ze dne 23. května 2017, kterým se provádí nařízení Evropského parlamentu a Rady (ES) č. 763/2008 o sčítání lidu, domů a bytů, pokud jde o způsob zpracování a strukturu zpráv o kvalitě a technický formát pro předávání údajů, a kterým se mění nařízení (EU) č. 1151/2010. Nařízením mj. ukládá povinnost členským státům předložit Komisi (Eurostatu) do 31. března 2024 podkladové informace a údaje a metadata související s kvalitou (obsah a struktura zprávy o kvalitě je v příloze nařízení). Nařízením rovněž definuje technický formát, který se použije pro předání údajů a metadat za referenční rok 2021 (formát pro výměnu statistických údajů a metadat – Statistical Data and Metadata Exchange – SDMX).