

ADMINISTRATIVNÍ ZDROJE DAT VE SČÍTÁNÍCH LIDU SE ZAMĚŘENÍM NA SČÍTÁNÍ 2011 V ČESKU

Robert Šanda¹⁾

THE USE OF ADMINISTRATIVE DATA SOURCES IN POPULATION CENSUSES WITH
A FOCUS ON THE CZECH 2011 CENSUS

Abstract

This article summarises the use of administrative data sources in population and housing censuses. It discusses the main advantages and drawbacks of combined and fully register-based censuses compared to traditional ones. The move from traditional to register-based censuses seems inevitable given the public's decreasing willingness to participate in the traditional form of enumeration. The article also describes the methods applied in the Czech 2011 census regarding record linkage and identifying overcoverage in the population register (ISEO) and highlights the most important impacts the use of administrative data has on census results.

Keywords: Population and housing census, Czechia, administrative data sources, overcoverage

Demografie, 2022, **64(2): 106–123**

DOI: <https://doi.org/10.54694/dem.0298>

1. ÚVOD

Na jaře 2021 se v Česku uskutečnilo zatím poslední sčítání lidu, domů a bytů, a v době přípravy předkládaného článku (leden 2022) probíhá zpracování jeho výsledků. Toto sčítání bylo od počátku příprav koncipováno jako tzv. kombinovaný census, využívající vedle dotazníkového šetření v maximální míře administrativní zdroje dat. Zpracování kombinovaného sčítání se od sčítání tradičního v některých krocích poměrně výrazně liší. Má potenciál přinést úplnější a kvalitnější výsledky, ale je komplikovanější a musí se vypořádat s některými problémy, které se v tradičním censu neobjevují, respektive mají okrajový význam.

Sčítání 2021 není první kombinované sčítání u nás. Tím bylo již sčítání 2011, které představovalo velmi významný milník v historii tzv. moderních censů z řady hledisek – bylo to první „povinné“ sčítání, které proběhlo podle evropských nařízení, poprvé v něm byly zavedeny elektronické sčítací formuláře, nově byl zaveden koncept obvykle bydlícího obyvatelstva (po 50 letech tak přestalo být základní určující charakteristikou místo trvalého, resp. evidovaného pobytu) a přineslo řadu dalších podstatných změn v organizaci a zpracování (viz např. *Habartová*, 2016; *ČSÚ*, 2013; *Bednářová*, 2011, 2012 nebo *Čtrnáct*, 2011). Posun od tradičního sčítání ke kombinovanému byl však změnou nejpodstatnější, i přesto, že čerpalo pouze z jednoho administrativního zdroje, informačního systému evidence obyvatel (ISEO).

1) Český statistický úřad, kontakt: robert.sanda@czso.cz.

Sčítání 2021 využívá více než desítku administrativních zdrojů. Nepředstavuje sice tak zásadní zlom jako sčítání předchozí, významně se však posunulo dále na cestě k plně administrativním sčítáním, při nichž se již neprovádí žádné terénní zjišťování a všechna data se získávají z administrativních zdrojů.

Přesun od tradičního sčítání k administrativnímu se v posledních desetiletích v Evropě silně prosazuje. O problematice využití administrativních dat v našich podmínkách však bylo dosud referováno spíše na mezinárodních workshopech úzce zaměřených na populační cenzu, v českých odborných kruzích byly prezentovány spíše jen stručné dílčí informace. Cílem článku je pokusit se dosavadní zkušenosti s administrativními záznamy shrnout, zakomponovat poslední vývoj sčítání u nás do evropského rámce a poskytnout základní údaje o jejich zpracování a vlivu na výsledky, přičemž se věnuje výsledkům sčítání 2011.

Zaměření části článku na sčítání 2011 má dva hlavní důvody. Z probíhajícího zpracování sčítání 2021 jsou zatím k dispozici jen dílčí poznatky, podrobněji bude možné o něm referovat až po získání definitivních výsledků. Druhým důvodem je snaha o zaplnění mezery v dosud poskytnutých informacích o minulém sčítání, což je podstatné i proto, že se způsob práce s administrativními zdroji stává jednou ze základních charakteristik jednotlivých sčítání (podobně jako třeba obsah formulářů nebo definice sledované populace). Nejde přitom jen o samotné sčítání, ale také o navazující průběžnou statistiku obyvatelstva, protože využití administrativních dat ve sčítání 2011 ovlivňuje ještě v současné době výsledky intercenzálních stavů obyvatelstva a demografických ukazatelů.

Na předkládaný článek by výhledově měla navázat další práce představující podrobněji postupy uplatněné ve zpracování sčítání 2021 a jejich vliv na výsledky. To bude možné po ukončení zpracování, resp. v jeho pokročilejší fázi. Od poznatků získaných při využití administrativních dat ve sčítání 2021 se budou odvíjet plány na reformu populační statistiky v příštím dese-

tileti a na zavedení každoročních tzv. malých cenzů podle připravovaného nového evropského nařízení o populační statistice. Těto problematice bude rovněž vhodné věnovat pozornost v navazujícím článku.

2. HISTORIE A HLAVNÍ DŮVODY VYUŽÍVÁNÍ REGISTRŮ VE SČÍTÁNÍ

Využívání administrativních zdrojů dat ve statistice obyvatelstva není novinkou, v evropském měřítku má již více než padesátiletou historii.²⁾ Její počátky jsou spjaty s vývojem v severovýchodních zemích. V Norsku byly výsledky sčítání lidu 1960 využity k založení centrálního populačního registru. Ten vznikl v roce 1964 a následně byl využit jako zdroj dat pro sčítání v roce 1970 (např. *Thorvaldsen – Solli*, 2012). Ve stejném roce byl populační registr poprvé využit i ve finském sčítání. O dekádu později, v roce 1981, proběhlo v Dánsku první sčítání lidu založené výhradně na administrativních zdrojích (*register-based census*). Vedle centrálního populačního registru těžilo toto sčítání z dat registru ekonomických subjektů, registru zaměstnání, registru příjmů, registru bytů a dalších zdrojů založených v průběhu 60. a 70. let (*UNECE*³⁾, 2007). Administrativní registry se v té době začaly prosazovat také ve sčítáních v některých dalších zemích, např. od roku 1975 ve Švédsku, od roku 1981 v Nizozemsku. V roce 1990 se Finsko stalo po Dánsku druhou zemí, která byla schopna zrealizovat ryze administrativní sčítání.

Rozmach přišel po roce 2000 (mj. *Valente*, 2010; *UNECE*, 2014). Ve sčítáních kolem roku 2000 byla sčítání využívající data z registrů ještě v menšině: pět zemí EU kombinovalo registry s dotazníkovými šetřeními, pouze v Dánsku a Finsku proběhl administrativní cenzus. V roce 2011 již registry do sčítání zapojila většina zemí EU včetně Česka, počet zemí s administrativním cenzem se rozrostl o Rakousko, Slovinsko a Švédsko. Téměř vždy vedla cesta od tra-

2) Myšlenka využití administrativních evidencí pro sčítání je ale starší. Pojí se již s působením „otce“ moderních sčítání A. Quetleta, který navrhoval využít sčítání v Belgii k založení registru obyvatel a z něho následně čerpat údaje o obyvatelstvu. Předpokládal, že sčítání 1846 bude v Belgii poslední, trvalo však ještě dalších 170 let, než tato země na administrativní sčítání přešla. (*Poullin – Herm*, 2013).

3) Evropská hospodářská komise OSN (United Nations Economic Commission for Europe)

dičního cenzu k administrativnímu přes alespoň jedno sčítání kombinované. Výjimkou bylo Rakousko, které jako jediné přešlo přímo z tradičního sčítání v roce 2000 na administrativní v roce 2011 (neznamená to ale, že se v Rakousku podařilo stihnout komplexní proces přechodu mezi tradičním a administrativními sčítáními během jednoho intercenzálního období, přípravy probíhaly již před rokem 2000).

V roce 2021 posun k registrům v rámci EU ještě zintenzivnil. Již pouze sedm současných členů EU organizuje tradiční sčítání (tab. 1).⁴⁾ Nejčastějším typem cenzu (10 zemí) je již čistě administrativní sčítání.⁵⁾ Tabulka 1 poskytuje přehled základních typů cenzů v EU (včetně Británie) v letech 2011 a 2021. Stejně jako ve všech předchozích obdobích platí, že posun je jednosměrný, od tradičního přes kombinovaný po čistě administrativní přístup. Žádná země se zatím nerozhodla vrátit k terénnímu šetření (ani výběrovému), pokud jej již jednou opustila. Někteří autoři uvádějí mírně odlišné údaje, než jaké prezentuje tabulka 1. Například *Valente* (2010) píše, že v roce 2011 pouze čtyři země EU provedou administrativní sčítání. Nepočítá Slovinsko, jehož sčítání 2011 však čistě administrativní bylo (zdroje tohoto sčítání uvádí např. *Josipovič*, 2015).⁶⁾ *Schulte Nordholt* (2014) naopak k administrativním přičítá ještě nizozemský a belgický cenzus. Tato sčítání však vedle registrů využila i data z výběrových šetření pracovních sil, tj. sice nebylo pro účely sčítání organizováno žádné ad-hoc šetření, bez informací od respondentů se však tato sčítání neobešla (*Schulte Nordholt*, 2014, 2018; *UNECE*, 2014). *UNECE* (2014) i *Schulte Nordholt* (2014) také chybně řadí české sčítání 2011 do skupiny tradičních cenzů, nikoliv kombinovaných. Mezi kombinované cenzy řadí ty, které čerpaly alespoň

jednu proměnnou výhradně z administrativních zdrojů, což české sčítání 2011 splňovalo (viz dále v textu).

Důvodů, proč se administrativní sčítání v Evropě tak výrazně prosazuje, je uváděno několik. Hlavní z nich jsou již vcelku všeobecně známé, jsou to možnost výrazně zkrátit periodu cenzů,⁷⁾ řádově nižší náklady ve srovnání s tradičním šetřením a zejména postupně se snižující ochota veřejnosti se sčítání zúčastnit. Negativní postoj veřejnosti k tradičnímu dotazníkovému šetření je z uvedených důvodů nejzávažnější. Pramení zejména ze snahy veřejnosti chránit své soukromí, ale podle *Skinnera* (2018) – a osobní zkušenosti s propagací sčítání to potvrzují – k tomu přispívají např. i klamné představy, že všechny potřebné informace jsou k dispozici, ale že stát je není schopen či ochoten využívat (např. přečtení informační hodnoty a využitelnosti „big dat“ mobilních operátorů).

Skinner (2018), resp. *UNECE* (2018) přidávají další, již spíše podružnější důvody pro přechod k administrativním cenzům. Jsou to komplikace při dotazníkových šetřeních v uzavřených či nebezpečných lokalitách, zlepšení spolupráce v rámci statistické instituce (vychází z představy, že administrativní cenzus vyžaduje zapojení více útvarů dané instituce), více času a zdrojů pro inovace a možnost flexibilněji reagovat na nové požadavky uživatelů.

Některé uváděné výhody jsou diskutabilní, například úspora nákladů je sice evidentní, ale uváděné konkrétní hodnoty nebývají srovnatelné, u administrativních cenzů bývají započítávány jen některé výdaje (navíc administrativní cenzy bývají sice levnější, ale obsahově chudší ve srovnání s tradičními). Zlepšení spolupráce uvnitř statistického úřadu rozhodně nemusí být pravidlem.

4) Tradiční sčítání proběhlo v roce 2021 také ve Spojeném království, které v Evropě patřilo mezi hlavní zastávce tradičních cenzů, nicméně i zde probíhají intenzivní přípravy na využívání registrů (situaci ovšem komplikuje absence centrálního populačního registru).

5) Mimo EU proběhlo administrativní sčítání ještě v Norsku. V ostatních státech světa stále dominuje tradiční cenzus (viz např. *Kukutai et al.*, 2015; *Eurostat*, 2012 nebo *Valente*, 2014 – sčítání 2011 v Evropě se proto občas nazývá světovou cenzovou laboratoří).

6) Později bylo slovinské sčítání přeřazeno do skupiny administrativních i v materiálech *UNECE*, viz např. *UNECE*, 2012 nebo <https://statswiki.unece.org/>.

7) V návaznosti na přirovnávání sčítání ke snímku společnosti se pro vystižení strategie přechodu na každoroční cenzu někdy používá obrat „from a snapshot to a movie“, např. *Eurostat* (2013).

Tab. 1: Základní typy sčítání lidu v zemích EU (EU-28) v letech 2011 a 2021

Basic types of census in EU countries (EU-28) in 2011 and 2021

		Typ sčítání 2011 Type of 2011 census			Počet zemí Number of countries
		tradiční traditional	kombinované combined	administrativní register-based	
Typ sčítání 2021 / Type of 2021 census	tradiční / traditional	Bulharsko/Bulgaria			8
		Chorvatsko/Croatia			
		Kypr/Cyprus			
		Francie/France			
		Řecko/Greece			
		Malta			
		Portugalsko/Portugal			
		Spojené království/UK			
	kombinované / combined	Maďarsko/Hungary	Česko/Czechia		10
		Irsko/Ireland	Estonsko/Estonia		
		Itálie/Italy	Německo/Germany		
		Lucembursko/Luxembourg	Polsko/Poland		
		Rumunsko/Romania			
		Slovensko/Slovakia			
	administrativní register-based		Belgie/Belgium	Rakousko/Austria	10
		Lotyšsko/Latvia	Dánsko/Denmark		
		Litva/Lithuania	Finsko/Finland		
		Nizozemsko/Netherlands	Slovinsko/Slovenia		
		Španělsko/Spain	Švédsko/Sweden		
Počet zemí Number of countries		14	9	5	28

Zdroj / Source: <https://statswiki.unece.org/>; UNECE, 2012, 2014; Valente, 2010.

Nejvíce zpochybnitelná je poslední uvedená výhoda, tj. větší flexibilita. Souhlasit lze v případě, že potřebu nových informací ze sčítání lze uspokojit údaji, které jsou v dostupných administrativních zdrojích k dispozici. Údaje, které žádný administrativní zdroj nevede, resp. ani z jejich povahy nemůže vést (subjektivní charakteristiky zjistitelné pouze deklarací respondenta), naopak v případě administrativních sčítání nejsou dosažitelné ani v delším časovém horizontu, zatímco do dotazníkového zjišťování je za určitých okolností zahrnout lze. Při přípravě zmíněného nového evropského nařízení se to opakovaně potvrzuje na jednáních pracovních skupin ke sčítání. Na nich prezentují hostující představitelé jednotlivých direktorátů Komise požadavky na roz-

šíření témat sčítání, přičemž často jsou to pouze země s tradičním či kombinovaným sčítáním, které jsou alespoň teoreticky schopny požadavky naplnit (témata jako je zdravotní omezení, bezdomovectví, pohlavní identita, sexuální orientace, rasa, některé otázky týkající se původu jako např. země narození rodičů atp.). Skutečnost, že požadovaná témata nejsou administrativními daty pokryta, je nejčastějším protiargumentem členských zemí. Omezená schopnost reagovat na měnící se poptávku po datech je tedy hendikepou administrativních sčítání. Do konfliktu se tak dostává rostoucí poptávka po „nových“ charakteristikách obyvatelstva s tlakem na využívání administrativních zdrojů a s tím spojenou vyšší frekvencí cenů.

3. ZÁKLADNÍ PRINCIPY POPULAČNÍCH CENZŮ V KONTEXTU VYUŽÍVÁNÍ ADMINISTRATIVNÍCH DAT

Podstatnou otázkou vztahující se k administrativním datům ve sčítání je, jak administrativní sčítání odpovídá základním atributům charakterizujícím populační cenzu obecně, čili zda je administrativní cenzus stále ještě cenzem. Základními charakteristikami populačních cenzů jsou: individuální sčítání, souběžnost, univerzálnost, velký detail a definovaná periodičita (např. *UNECE*, 2015).

Individuálním sčítáním (*individual enumeration*) je myšleno získávání informací o všech jednotlivých entitách (osobách, bytech...) podléhajících sčítání. Výsledkem jsou záznamy za každou osobu a z těchto mikrodat se získávají agregované výsledky. To umožňuje kombinování široké škály charakteristik. V tradičních sčítáních, při nichž se všechny údaje o každé sčítané osobě získávají pomocí osobního dotazníku, je tento princip zajištěn. U vícezdrojových cenzů, kam patří prakticky všechny administrativní cenzy, je pro naplnění požadavku na individuální sčítání nutné, aby administrativní záznamy byly mezi sebou propojitelné (*UNECE*, 2018). V českém prostředí tuto podmínku zajišťuje rodné číslo, ovšem zejména v posledním desetiletí je patrná snaha minimalizovat až eliminovat jeho využívání veřejnou správou. Systém základních a agendových identifikátorů, který má rodné číslo při propojování dat nahradit, však zatím funguje ve velmi omezeném rozsahu, není a ze své podstaty nikdy nebude využitelný bez napojení na systém základních registrů. Navíc je k dispozici pouze pro osoby vedené v základním registru obyvatel, který však neeviduje všechny obyvatele, protože občané EU žijící v Česku nemají povinnost se přihlásit k pobytu.

Souběžnost (*simultaneity*) znamená jednotné referenční období či přímo okamžik, k němuž se veškeré výsledky vztahují (v české zavedené terminologii *rozhodný okamžik*). V případě tradičního sčítání je toho dosahováno soustředěním dotazníkového šetření do co nejkratšího časového úseku. Při využívání administrativních zdrojů je pro splnění požadavku na jednotný referenční okamžik třeba vypořádat se s problémy, jako jsou opožděně zaznamenávané aktualizace (například narození, úmrtí či přestěhová-

ní zaznamenané s větším časovým odstupem od dané události), aktualizace neprováděné kontinuálně (např. u nás případ dat ze školských matrik), resp. zdroje nereferující o stavu v konkrétním okamžiku (u nás například údaje z přiznání k dani z příjmů).

Princip univerzálnosti (*universality*) je zajištěn, pokud sčítání k danému rozhodnému okamžiku pokrývá veškeré obyvatelstvo daného území a pokud se údaje o všech obyvatelích získávají stejným způsobem. *UNECE* (2018) zdůrazňuje problematiku překladů sčítacích formulářů do více jazyků – musí být zaručeno, že otázky a instrukce mají ve všech jazycích, do nichž je formulář překládán, stejný význam.

Univerzálnost je z dosud uvedených fundamentálních rysů sčítání nejproblematictější. *UNECE* (2018) to částečně připouští, protože je běžné, že řada administrativních zdrojů využívaných v cenzech pokrývá pouze část území, například v některých zemích neexistuje centrální evidence obyvatel, ale pouze obecní registry (např. Německo, viz *Scholz – Kreyenfeld*, 2016 či Itálie, viz např. *Casale et al.*, 2012, ucelený přehled vypracovali *Poullin – Herm*, 2013). Velmi často také administrativní zdroje pokrývají pouze specifické segmenty populace. Ze zdrojů pro SLDB 2021 je to většina zdrojů, pouze základní registr obyvatel a centrální registr pojištěnců by z principu měly pokrývat veškeré obyvatelstvo.

UNECE (2018) doporučuje v těchto případech hledat cestu, jak z jednotlivých zdrojů odvodit společné „cenzové“ charakteristiky. To však naráží na jeden z často zmiňovaných problémů administrativních sčítání – definice evidovaných proměnných vycházejí z účelu toho kterého administrativního zdroje, tj. nemusí odpovídat potřebám sčítání a mohou se zdroj od zdroje lišit. Možná ještě větším problémem než definice v metodikách jednotlivých registrů a databází je praxe jejich vedení. Ta se liší zdroj od zdroje, proměnná od proměnné (podstatné proměnné pro jeden zdroj jsou podružné pro jiný), editor od editora. Možnosti sladění informací z jednotlivých zdrojů jsou limitované. Princip univerzálnosti je tak více než ostatní principy spíše teoretickým konceptem, ke kterému je třeba v rámci možností směřovat, který ale nebývá naplněn.

Nutno však dodat, že to není pouze problém administrativních dat, ale i kombinovaných sčítání, která

zjišťují některé charakteristiky výběrově,⁸⁾ byť celosvětová doporučení (UN, 2017) výslovně uvádějí, že užití výběrových technik není s principem univerzálnosti v rozporu. Jestliže zajištění univerzálnosti je problematické na národní úrovni, na úrovni EU je i přes platná nařízení a soustavnou snahu Eurostatu o harmonizaci již jen ideálním, ale prakticky nedosažitelným cílem.

Dalším ze základních rysů definujících sčítání je velký detail, za nějž jsou dostupné výsledky (někdy se používá termín *granularita*). Tím je primárně míněn územní detail, k němuž se podle UNECE (2015, 2018) přidružuje i podchycení početně malých specifických subpopulací. Požadavek na územní detail vyústil v zahrnutí údajů na úrovni kilometrové čtvercové sítě do témat sčítání v Doporučeních (UN, 2017) a pro členské země EU v nový implementační předpis nařizující poskytnutí základních výsledků v detailu kilometrových čtverců. V našich podmínkách je problémem zejména nedostatečné pokrytí některých skupin obyvatelstva, například již zmíněných cizinců ze zemí EU neevidovaných v základním registru obyvatel nebo ekonomicky neaktivních osob s vlastním zdrojem obživy. Specifickou kategorií jsou lidé bez domova (primární a zejména sekundární bezdomovci), kteří sice mohou být v systémech veřejné správy evidováni, ale bez možnosti je identifikovat jako bezdomovce.

Stejně jako univerzálnost, i granularita je problematická u cenů využívajících některá data z výběrových šetření. Na tento problém (v souvislosti s využíváním tzv. krátkých a dlouhých dotazníků) poukazuje i Skinner (2018).

Periodicita je z pohledu administrativních dat asi nejnáze dosažitelný princip. Zatímco sčítání formou vyčerpávajícího šetření lze zajistit s desetiletou, v některých zemích s pětiletou periodou (Irsko, Austrálie, Kanada atd.), administrativní sčítání lze obvykle provádět častěji. Za zmínku stojí požadavek na periodicitu v českém kontextu. Je sice všeobecně předpokládáno, že sčítání se koná v desetiletých intervalech, ale národní právní úprava pravidelná sčítání nestanovuje. Zákon o státní statistické službě se

o sčítání nezmiňuje vůbec (s výjimkou reliktu v názvu § 12, který však nemá žádný význam) a zákony o sčítání jsou tradičně jednorázové, řešící pouze jednotlivá sčítání. Jediným platným předpisem stanovujícím povinnost provádět sčítání je evropské nařízení 763/2008.

4. HLAVNÍ METODICKÉ PROBLÉMY SPOJENÉ S ADMINISTRATIVNÍMI CENZY

S nástupem administrativních zdrojů a postupným upozadováním významu dotazníkového šetření se přirozeně proměnila i metodika zpracování dat. Problematika dotazníkových šetření, např. metodika tvorby dotazníků nebo organizace terénních prací v evropském prostředí ustupuje do pozadí. Do centra pozornosti se dostaly nové výzvy, jimž je třeba čelit, aby výsledky co nejlépe odpovídaly reálnému stavu sledované populace. Mezi nimi má zásadní význam téma propojování dat z jednotlivých zdrojů a problematika tzv. *overcoverage* (do češtiny přeložená v nařízení Komise 2017/881 jako *přesah pokrytí*, v české terminologii pro domácnostní šetření se vžil pojem *administrativní odpad* pro analogický problém s evidencí bytů ve statistickém registru sčítacích obvodů a budov, viz např. Zelený, 2006).

Propojování dat

Propojování dat je jádrem celého zpracování a má podstatný dopad na přesnost výsledků. V ideálním případě existuje pro propojování dat jednoznačný identifikátor, pomocí kterého lze danou entitu (např. osobu) nalézt ve všech využívaných zdrojích. U nás je v případě osob tímto identifikátorem rodné číslo, u domů je to nejčastěji adresa či kód adresního místa z registru územní identifikace, adres a nemovitostí. V případě bytů žádný všeobecně využitelný identifikátor neexistuje, pouze omezeně lze využít číslo jednotky z informačního systému katastru nemovitostí (ISKN). Přímé propojení záznamů přes jednoznačný identifikační údaj (*single-step deterministic linkage*)

8) Například nizozemský census v roce 2001 opustil celoplošné šetření, ale protože administrativní zdroje neposkytovaly dostatečně úplné informace o úrovni vzdělání a zaměstnání, byly tyto charakteristiky převzaty z výběrových šetření pracovních sil (blíže viz např. Schulte Nordholdt – Hartgers – Gircour, 2004). Výběrově byly v roce 2011 zjišťovány některé charakteristiky také např. v Itálii či Polsku (INIS, 2012; Dygaszewicz, 2020).

má jen omezené využití. V praxi se úloha propojování dat výrazně komplikuje z důvodů chybně zapsaných, neunikátních či neúplných identifikátorů.

Pokročilejším řešením je iterativní deterministické propojování, při kterém záznamy nepropojené podle jednoho identifikátoru postupují do dalších kroků, v nichž se propojují podle dalších, obvykle méně spolehlivých identifikátorů (viz např. *Dusetzina et al.*, 2014). V řadě případů není dostačující ani tento postup a je třeba aplikovat algoritmy pravděpodobnostního propojování (*probabilistic record linkage*). Při něm se záznamy z jednotlivých zdrojů propojí na základě podobnosti, nikoliv přesné shody vybraných identifikátorů. Základní koncept pravděpodobnostních technik formulovali *Fellegi a Sunter* (1969), od té doby se problematika pravděpodobnostního propojování poměrně dynamicky rozvíjí. Postupy pravděpodobnostního propojování jsou uplatnitelné i v dalších úlohách v rámci zpracování výsledků sčítání, především při kódování údajů z formulářů, které bývají respondenty zapsány velmi rozličně (viz např. *Šanda*, 2020).

Přesah pokrytí

Termín přesah pokrytí (*overcoverage*) pojmenovává problém osob (resp. bytů apod.) ve zdrojových resp. i výsledných datech, které by podle definice neměly patřit do sledované, v terminologii evropských nařízení *cílové populace* (*target population*). Je to téma úzce spjaté s administrativními zdroji, i když se nevyhýbá ani tradičním sčítáním, kde jsou příčinou přesahů hlavně neodhalené duplicity, případně fiktivní osoby na vyplněných dotaznících.

Při tradičních populačních cenzech je přesnost výsledných počtů obyvatel zásadním způsobem závislá na participaci obyvatel na dotazníkovém šetření, tj. na úplnosti pokrytí populace.⁹⁾ Hlavním rizikem z hlediska celkových počtů je tak podhodnocení

velikosti populace v důsledku neúplného pokrytí (*undercoverage*). Při využití administrativních zdrojů se riziko těchto nedopočtů neeliminuje, ale spíše než na výsledky za celkovou populaci mají dopad na kvalitu výsledků týkajících se specifických subpopulací, jejichž registrace není úplná.

Přesah pokrytí je všeobecným rysem populačních registrů a je třeba ho řešit při tvorbě veškerých statistik, které populační registry využívají. Tvoří ho osoby, které jsou registrovány jako obyvatelé žijící na daném území, kde se však fakticky nevyskytují. Příčinou je hlavně neúplná registrace vystěhování do zahraničí, v menší míře neregistrovaná úmrtí (za všechny viz např. *Bucher*, 2014, rozebírající dopady neregistrované emigrace na demografické charakteristiky německé populace).

S těmito komplikacemi se potýkají i země s obsahově rozsáhlou a kvalitní datovou základnou, jako jsou severské země, přestože ty mají velkou výhodu spočívající v zajištění vzájemného sdílení údajů o osobách.¹⁰⁾ Například ve Švédsku, kde je (na rozdíl od nás) povinné nahlášení odjezdu ze země na rok a déle, což ovšem často není dodržováno.¹¹⁾ Řešení navržené *Aradhya et al.* (2017) je založeno na vyřazení osob, které jsou sice evidovány v centrálním populačním registru, ale nemají evidovaný žádný příjem (žádný příjem z ekonomické činnosti, žádný důchod atd.). Tento přístup je však nepoužitelný pro děti a mladé lidi, resp. pro osoby ve vysokém věku. Nepočítá také s osobami sice v zemi fakticky bydlícími, ale bez oficiálních příjmů, naopak neumožňuje identifikovat vystěhovalé, kteří pobírají ve Švédsku příjem (např. starobní důchod). Druhý přístup, navržený *Monti et al.* (2019) pro subpopulaci migrantů, je založen na „trasování“ v registrech. Při něm se posuzují charakteristiky dané osoby v příslušném referenčním roce, jako například dokončení školy, evidované vnitřní přestěhování, evidované zaměstnání atd.

9) K tomu se přidávají další aspekty, jako je kvalita odpovědí na otázku na místo pobytu, úroveň metodiky zpracování a flexibilita procesu zpracování, resp. jeho možnosti reagovat na zjištěné problémy.

10) Mezinárodní výměna dat na úrovni EU, je sice velmi obtížné představitelné opatření pro zvýšení kvality evidence obyvatel, ovšem nezdá se, že by existovala jiná efektivní cesta, jak v uspokojivé míře podchytit migraci mezi členskými zeměmi. Blíže o tomto problému viz např. *Holá* (2005).

11) Podle *Monti et al.* (2019) emigrující lidé nezdíka o této povinnosti ani nevědí, případně se obávají „přetržení“ formální vazby na Švédsko.

K nim se přibírají (s rozdílnými vahami) informace z okolních let, které indikují přítomnost osoby v referenčním roce, nebo naopak její nepřítomnost.

Trasování je de facto metoda obecně označovaná jako metoda „známek života“ (*signs-of-life analysis*), spočívající ve vyhodnocení informací, které o osobě evidované v centrálním populačním registru poskytují další administrativní zdroje. Často zmiňovaným příkladem tohoto přístupu je vypočítávání tzv. rezidenčního indexu pro vymezení obyvatelstva v Eston-sku (*Maasing – Tliit – Vähi*, 2016). Jiné varianty metody hledání známek života byly aplikovány například ve Slovinsku (*Dolenč*, 2012) nebo Rakousku (*Lenk*, 2008). Signs-of-life analýza může efektivně řešit problém s overcoverage (dokonce se uvažovalo o jejím zmínění přímo v definici populace v připravovaném novém evropském nařízení k populační statistice). Podmínkou je však existence dostatečného množství kvalitních informací ve více administrativních zdrojích, jejich dostupnost pro statistické účely a možnost jejich spolehlivého propojení.

V českém sčítání 2011, jak bude dále popsáno, měly oba zmíněné metodické problémy (propojování dat i přesah pokrytí) vliv na výsledný počet obyvatel v řádu statisíců, přičemž vliv u některých skupin obyvatel byl velmi výrazný.

5. VYUŽITÍ ADMINISTRATIVNÍCH ZDROJŮ DAT V ČESKÉM SČÍTÁNÍ LIDU V ROCE 2011

V roce 2011 byl jediným pro sčítání dostupným administrativním zdrojem o osobách informační systém evidence obyvatel (ISEO). Ten v té době evidoval české občany i cizince hlášené u nás k pobytu. Údaje z ISEO byly ve sčítání využity následujícími způsoby:

a) Předvyplnění sčítacích formulářů

Z hlediska způsobu využití se jednalo spíše o podporu terénních prací, nikoliv o využití jako zdroje dat pro výsledky. Dopad byl však významný a nelze se o předvyplnění nezmínit. V říjnu 2010 byly podle údajů ISEO předvyplněny sčítací formuláře jménem, příjmením a adresou místa sečtení (podle adresy evidovaného pobytu). V průběhu distribuce formulářů komisaři na jednotlivých adresách zjišťovali, zda na dané adrese žijí osoby, jejichž jména byla na předtištěných

formulářích. Pokud si formulář s určitým jménem na adrese nikdo nepřevzal, zůstal nevyužit. Respondent zastížený komisařem na adrese, na níž neměl hlášený pobyt, dostal bílý formulář. Toto řešení pomohlo zefektivnit terénní práce jen částečně. Řada obyvatel žila na jiných než evidovaných adresách, v ISEO byly evidovány osoby vůbec nežijící na území ČR (overcoverage). Mnoho předvyplněných formulářů tak nakonec nebylo použito.

b) Využití ISEO jako alternativního zdroje

Přístup ČSÚ k datům ISEO byl při sčítání 2011 poměrně zdrženlivý. Sčítací list osoby obsahoval otázky na téměř všechny údaje, které byly pro sčítání potřebné, včetně těch, které byly v ISEO k dispozici. Vedle identifikačních údajů nutných k propojení formulářů s evidencí (příjmení, jméno, rodné číslo, datum narození) to byly informace o pohlaví, státním občanství, rodinném stavu, resp. registrovaném partnerství. V případě rozporu měl (s výjimkou údaje o registrovaném partnerství) vyšší prioritu údaj z formuláře. Údaj z ISEO se použil v případech, že příslušná otázka na formuláři nebyla zodpovězena, resp. nebyla zodpovězena smysluplně.

c) ISEO jako zdroj údajů nezjišťovaných na formuláři

Pro některé ukazatele byly záznamy z evidence obyvatel jediným zdrojem. Jednalo se o místo evidovaného pobytu, datum sňatku a rok příchodu obyvatele do země. Poslední zmíněný údaj byl jedinou charakteristikou povinně vyžadovanou evropským nařízením, který nebyl zjišťován na formulářích. Datum sňatku z evidence obyvatel se při zpracování ukázalo jako nepoužitelné (často chybělo, nebylo u manželů totožné apod.), doba trvání sňatku tak byla z výsledků sčítání 2011 zcela vyřazena.

d) Další využití, které mělo ze všech uvedených největší dopad na výsledky sčítání 2011, spočívalo v rozšíření počtu osob o osoby v ISEO, které se nepodařilo sečíst v dotazníkovém šetření. Poprvé v historii sčítání lidu na našem území tak počet obyvatel nevycházel pouze z počtu obyvatel sečtených při terénním šetření.

Vymezení obyvatelstva ve sčítání 2011

V dotazníkovém šetření v rámci sčítání 2011 byly shromážděny vyplněné formuláře obsahující

10 148,8 tisíc záznamů osob (tj. před deduplikací a dalšími procesy vedoucími k vymezení trvale a obvykle bydlicího obyvatelstva). Podle intercenzální bilance obyvatelstva mělo Česko k 31. 12. 2010 necelých 10 533 tisíc obyvatel. Při předcházejících cenzech bylo při sčítání pravidelně zjišťováno méně obyvatel, než kolik udávala bilance obyvatel vycházející z deset let starého základu (minulého sčítání), viz tab. 2. Rozdíl se však obvykle pohyboval v řádu desítek tisíc osob, ačkoliv v roce 2001 tomu napomohla změna definice obyvatelstva (podle závěrečné zprávy pro vládu by v případě pouze trvale bydlicího obyvatelstva pokles činil 100 tisíc obyvatel).¹²⁾ Propad o zhruba půl milionu trvale bydlicích obyvatel v roce 2011 by byl bezprecedentní a výsledný počet obyvatel zcela nevěrohodný.

Bylo zjevné, že nezanedbatelná část obyvatelstva se v roce 2011 nesečetla. Doplnění osob z ISEO tak bylo klíčové pro zpřesnění výsledků sčítání, i když za tyto osoby nebylo možné získat charakteristiky nevidované v ISEO, jako je například úroveň vzdělání nebo ekonomická aktivita.

Podle záznamů z ISEO poskytnutých Ministerstvem vnitra pro účely sčítání bylo k rozhodnému okamžiku u nás evidováno 10 715,1 tisíc osob. Jestliže počet

osob sečtených prostřednictvím formulářů byl zjevně příliš nízký, počet osob v registru byl příliš vysoký. To nebylo překvapivé, neboť souhrnné počty evidovaných osob v ISEO byly známé, ale poprvé bylo třeba ve zpracování sčítání řešit problematiku propojování dat a přesahu pokrytí.

K propojení záznamů bylo v roce 2011 primárně využito kombinace (iterativního) deterministického přístupu a pravděpodobnostního propojování. Jako základní identifikátor bylo přirozeně použito rodné číslo, sekundárně byly využity údaje o jménu a příjmení, datu narození a adrese pobytu. Takto se podařilo s ISEO propojit 96 % vyplněných formulářů. Zbývající 4 % formulářů vstoupila do kroku pravděpodobnostního propojování. Pro tento krok byly identifikovány nejpodobnější záznamy. Pokud míra podobnosti dosáhla určité hranice, empiricky stanovené týmem zajišťujícím zpracování, byly záznamy propojeny. Určitou slabinou celého postupu propojování bylo přílišné spolehnutí se na rodné číslo, aniž by bylo dále řešeno, zda nedošlo k záměně (dosavadní poznatky ze zpracování dat 2021 ukazují, že k takovým záměnám dochází).

Po propojení zbylo 127 tisíc formulářů, které se nepropojily s žádným administrativním záznamem.¹³⁾ Na straně ISEO zůstalo 693 tisíc záznamů, na které se

Tab. 2 : Obyvatelstvo ČR podle základních věkových skupin ve vybraných letech (stavy před a po sčítání lidu)

Czech population by basic age groups in selected years (before and after the population and housing censuses)

Věk Age	Stav k / Population as of									
	31.12.1969	1.1.1970	31.12.1979	1.1.1980	31.12.1990	1.1.1991	31.12.2000	1.1.2001	31.12.2010	1.1.2011
0–14	2 115 048	2 089 919	2 407 530	2 405 708	2 193 682	2 175 638	1 664 434	1 662 040	1 518 142	1 521 765
15–64	6 608 415	6 530 232	6 504 728	6 474 884	6 867 991	6 826 911	7 179 109	7 159 345	7 378 802	7 327 997
65 +	1 183 011	1 169 328	1 403 411	1 392 005	1 302 451	1 302 058	1 423 003	1 410 642	1 635 826	1 636 969
Celkem Total	9 906 474	9 789 479	10 315 669	10 272 597	10 364 124	10 304 607	10 266 546	10 232 027	10 532 770	10 486 731

Pozn.: Dvojice koncový stav (31. 12.) a počáteční stav (1. 1.) se vztahují ke stejnému časovému okamžiku. Koncové stavy vycházejí ze „starého“ sčítání, počáteční stavy ze sčítání konaného v daném roce.

Note: The data on total population as of 12/31 (end of the given year) and 1/1 (beginning of the following year) refer to the same point in time. However, the population at 12/31 is based on the "old" census, while the population at 1/1 is based on the "new" one.

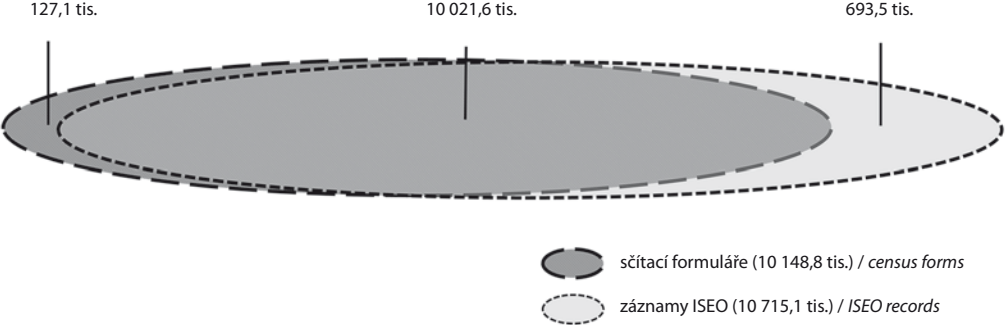
Zdroj: ČSÚ.

Source: CZSO.

12) Do roku 2001 bylo sledováno pouze trvale bydlicí obyvatelstvo, počínaje sčítáním 2001 i obyvatelstvo s dlouhodobým pobytem a uděleným azylem – tyto osoby zmírnily propad počtu obyvatel, který představovalo nasazení na stav vycházející ze sčítání 2001.

13) Řada z nich byla vyřazena v dalších fázích zpracování, do výsledného obvykle bydlicího obyvatelstva se z těchto 127 tisíc osob dostalo jen 54 tisíc, tato skupina představovala obvykle bydlicí obyvatele bez trvalého pobytu (viz Šanda, 2015).

Obr. 1: Vstupní záznamy osob ve zpracování SLDB 2011 podle zdroje dat
Input data from personal records analysed from the 2011 Population and Housing Census by data source



nenapojil žádný formulář (obr. 1). Mezi těmito záznamy bylo dále třeba identifikovat záznamy představující přesah pokrytí. Výše popsané postupy aplikované v jiných zemích nebylo možné efektivně použít, protože jiné administrativní zdroje než ISEO nebyly k dispozici. Byly stanoveny následující podmínky (popsané

již v Šanda, 2015), z nichž alespoň jedna musela být splněna, aby byl záznam ISEO zahrnut do zpracování výsledků:

- Evidovaná osoba (resp. někdo z domácnosti) v místě trvalého bydliště převzala předvyplněný sčítací formulář, který však neodevdala.

Tab. 3 : Vybrané alternativní podmínky pro určení přesahu pokrytí ISEO; sčítání lidu 2011
Selected alternative conditions for defining overcoverage in ISEO records; 2011 census

Alternativy výběru záznamů osob pro další kroky zpracování sčítání <i>Alternatives to the use of personal records for further census processing</i>	Počet záznamů zařazených do sčítání <i>Number of records included in further processing</i>	Vyřazené záznamy ISEO (přesah pokrytí) <i>Excluded ISEO records (overcoverage)</i>	
Pouze osoby na formulářích <i>Only persons enumerated using census forms</i>	10 148 755	693 499	
Navíc osoby v ISEO, jejichž předvyplněný formulář byl distribuován, ale nevrátil se vyplněný <i>Plus persons recorded in ISEO who were sent a pre-filled census form but did not return it completed</i>	10 247 059	595 195	
Navíc záznamy v ISEO, u nichž byla evidována změna <i>Plus ISEO records updated within:</i>	v posledním roce <i>the last year</i>	10 348 050	494 204
	v posledních 2 letech <i>the last 2 years</i>	10 419 380	422 874
	v posledních 3 letech <i>the last 3 years</i>	10 480 258	361 996
	v posledních 4 letech <i>the last 4 years</i>	10 527 452	314 802
	v posledních 5 letech (zvolená varianta) <i>last 5 years (chosen alternative)</i>	10 568 191	274 063
	kdykoliv (všechny osoby s evidovaným pobytem) <i>whenever (all persons with a registered residence)</i>	10 842 254	–

Zdroj: Zpracovatelská databáze SLDB 2011.
Source: Processing database of 2011 Census.

– V posledních pěti letech před rozhodným okamžikem došlo u dané osoby v záznamu ISEO ke změně (např. se osobě narodilo dítě, změnilo se místo pobytu, rodinný stav apod.)

Vedle toho musela být splněna podmínka vyžadující, aby adresa místa evidovaného pobytu osoby v populačním registru byla dohledatelná ve statistickém registru sčítacích obvodů a budov.

V principu se jednalo o přístup analogický signs-of-life analýze, i když při ní jsou využívány výrazně větší počty relevantních administrativních zdrojů. První podmínku (úspěšná distribuce předvyplněného formuláře) splnilo 98,3 tisíc evidovaných osob, podmínku na změnu záznamu v ISEO v posledních pěti letech splnilo dalších 321,1 tisíc evidovaných osob. Celkově tak bylo do souboru pro další zpracování doplněno 419,4 tisíc osob, což by, pokud byl přístup správný, měli být obyvatelé, kteří se nezúčastnili sčítání lidu, přestože fakticky na našem území žili.

Podmínka na úspěšnou distribuci předvyplněného formuláře byla založena na předpokladu, že je nepravděpodobné, aby pro osoby dlouhodobě mimo naše území někdo od komisaře přebíral předvyplněný formulář. Volba horizontu pěti let pro poslední změnu v ISEO byla do velké míry subjektivní, přičemž svoji

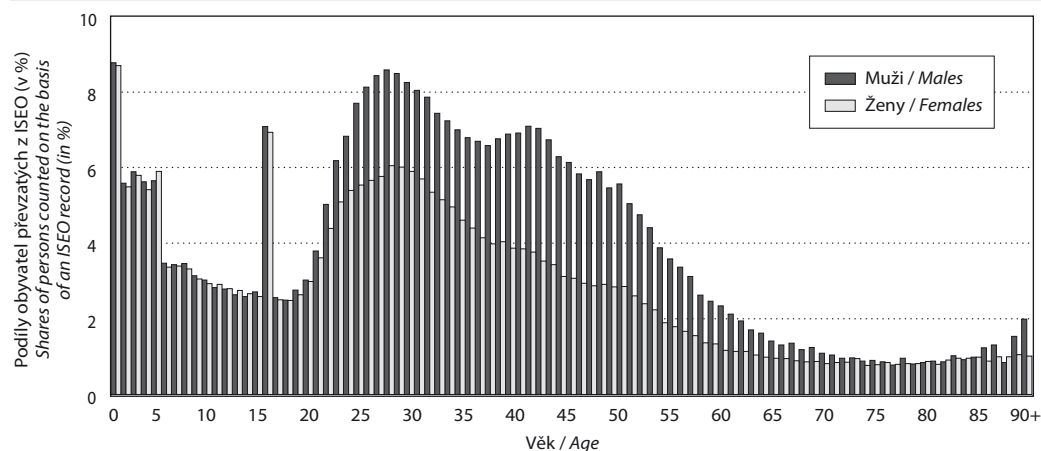
roli bezesporu sehrála skutečnost, že počet osob v souboru (ještě ne výsledný počet obyvatel) se tak nejvíce blížil stavu obyvatelstva podle intercenzální bilance. Alternativní časové intervaly a jejich dopad na vymezení základního souboru obyvatel představuje tab. 3. Z ní je patrné, že volba délky období, v němž muselo dojít ke změně v ISEO, měla nezanedbatelný vliv na výsledný počet obyvatel.

Popsaný poněkud „hrubý“ přístup k rozhodnutí o zařazení/vyřazení osob v ISEO do sčítání byl zapříčiněn omezeným množstvím dostupných informací a tedy nemožností provést plnohodnotnou signs-of-life analýzu. I proto byl poté pro sčítání 2021 kladen velký důraz na zpřístupnění všech existujících informačních systémů veřejné správy.

Sjednocením množin obyvatel sečtených na formulářích a osob doplněných z ISEO vznikl základní soubor osob vstupující do dalšího zpracování. Ten obsahoval 10 568,2 tisíce záznamů. Z nich se zpracovávaly výsledky sčítání. Všechny osoby doplněné z ISEO byly zařazeny do trvale bydlicího obyvatelstva, a to na adresu svého evidovaného pobytu. Do obvykle bydlicího obyvatelstva nebylo zahrnuto všech 419 tisíc doplněných osob, byly vyřazeny ty, které měly místo evidovaného pobytu na ohlašovně pobytu – těchto osob bylo 104 tisíc.

Obr. 2: Podíly obyvatel převzatých z ISEO (nesečtených na formulářích) na celkových výsledných počtech obyvatel podle věku a pohlaví (SLDB 2011)

Shares of persons counted on the basis of an ISEO record (not counted using a census form) out of the total number of inhabitants by age and sex (2011 Population and Housing Census)



Zdroj: Zpracovatelská databáze SLDB 2011.

Source: Processing database of 2011 Census.

Ve výsledném souboru osob doplněných z ISEO dominovali muži, tvořili 60 % doplněných. Výrazně byly ve srovnání s celou populací zastoupeny věkové kategorie od 25 do 50 let, resp. u žen od 25 do 35 let. Vysoké byly také relativní počty dětí, hlavně dětí do pěti let. To bylo primárně způsobeno výše uvedenou podmínkou na změnu záznamu v ISEO v posledních pěti letech, tj. všechny evidované děti mladší pěti let byly do sčítání zahrnuty, pokud jejich adresa evidovaného pobytu byla nalezena v registru sčítacích obvodů a budov. Výraznější zastoupení dětské složky v doplněných osobách zřejmě souviselo také s tím, že se ve větší míře neúčastnili sčítání jejich rodiče (zmiňované kategorie osob ve věku 25–40 let), ti pravděpodobně nesečetli ani své děti.

Ve věkové struktuře osob doplněných z ISEO na obr. 2 je nepřehlédnutelný exces u patnáctiletých osob, jejichž počet několikanásobně přesahuje počty v okolních věcích. Je to důsledek zvoleného řešení jedné z nejvýraznějších komplikací spojených s doplňováním osob z evidence. Aby byly výsledky sčítání

konzistentní, je třeba zajistit vazby mezi sledovanými entitami – osoby (s výjimkou osob bez domova a hromadně ubytovaných jednotlivců) patří do hospodařících domácností, hospodařící domácnosti do bytových domácností, bytové (většinou) do bytů, byty do domů. Přidat osobu z registru do sčítané populace tedy znamená zařadit ji do konkrétní domácnosti, případně pro ni vytvořit novou domácnost v konkrétním bytě. Přidává-li se do sčítání osoba z evidence, jediným vodítkem pro její lokalizaci je adresa evidovaného pobytu. Nevyhnutelně tak nastávají situace, kdy na dané adrese neexistuje žádná „vhodná“ domácnost, do které lze osobu přiřadit. Je-li doplňovaná osobou dítě, nelze pro něho vytvořit samostatnou domácnost. V těchto případech byl ve sčítání 2011 přijat předpoklad, že taková osoba nemůže být dítětem, tj. že je chybná informace o datu narození – osobě tak byl věk změněn na 15 let a poté mohla být vytvořena i nová domácnost jednotlivce. To zapříčinilo výsledně vysoké počty 15letých osob doplněných z ISEO.

Tab. 4 : Obyvatelstvo podle zdroje dat (sčítací formulář, resp. ISEO) a vybraných státních občanství (SLDB 2011)

Population by data source (census form or ISEO record) and by selected state citizenships (2011 census)

Občanství Citizenship	Obvykle bydlící podle zdroje dat Usually resident population by data source			Trvale bydlící podle zdroje dat Permanent resident population by data source		
	Formulář Census form	ISEO	ISEO (%)	Formulář Census form	ISEO	ISEO (%)
Obyvatelstvo celkem Total population	10 082 249	354 311	3,4	10 071 173	419 436	4
Občané ČR Czech citizens	9 770 055	182 519	1,8	9 803 455	244 929	2,4
Cizinci – celkem Foreign nationals – total	251 625	170 651	40,4	204 227	173 359	45,9
Ukrajina / Ukraine	57 942	58 197	50,1	54 075	58 990	52,2
Slovensko / Slovakia	62 730	19 521	23,7	44 050	19 572	30,8
Vietnam	33 397	19 215	36,5	28 828	19 551	40,4
Rusko / Russia	17 693	13 852	43,9	15 730	14 011	47,1
Německo / Germany	5 505	9 402	63,1	3 722	9 513	71,9
USA	3 116	4 316	58,1	2 214	4 438	66,7
Spojené království UK	2 938	2 244	43,3	1 634	2 248	57,9

Zdroj: Zpracovatelská databáze SLDB 2011.
Source: Processing database of 2011 Census.

Vedle specifíků v pohlavně věkové struktuře se komentovaná skupina obyvatel vyznačovala také vysokým podílem cizinců, ti představovali 42 % všech doplněných osob. Třetinu těchto cizinců tvořili občané Ukrajiny. Cizinci zvyšovali u doplněných obyvatel zastoupení mužů a osob ve věku 30–40 let, ovšem i mezi českými občany tyto skupiny tvořily výrazný podíl.

Z uvedeného plyne, že obyvatelé doplnění z ISEO ovlivnili výslednou velikost různých skupin obyvatelstva nestejnou měrou. S tím jsou spojeny i počty nezjištěných údajů, které tak byly všeobecně vyšší než v předchozích sčítáních. O doplněných osobách nebyly známy žádné informace, které se zjišťovaly výhradně prostřednictvím formulářů, tj. například ekonomická aktivita, místo pracoviště/školy, úroveň vzdělání, mateřský jazyk, národnost atd. Konečně i ztotožnění místa evidovaného pobytu s místem obvyklého pobytu bylo jistě v určité míře zkreslující.

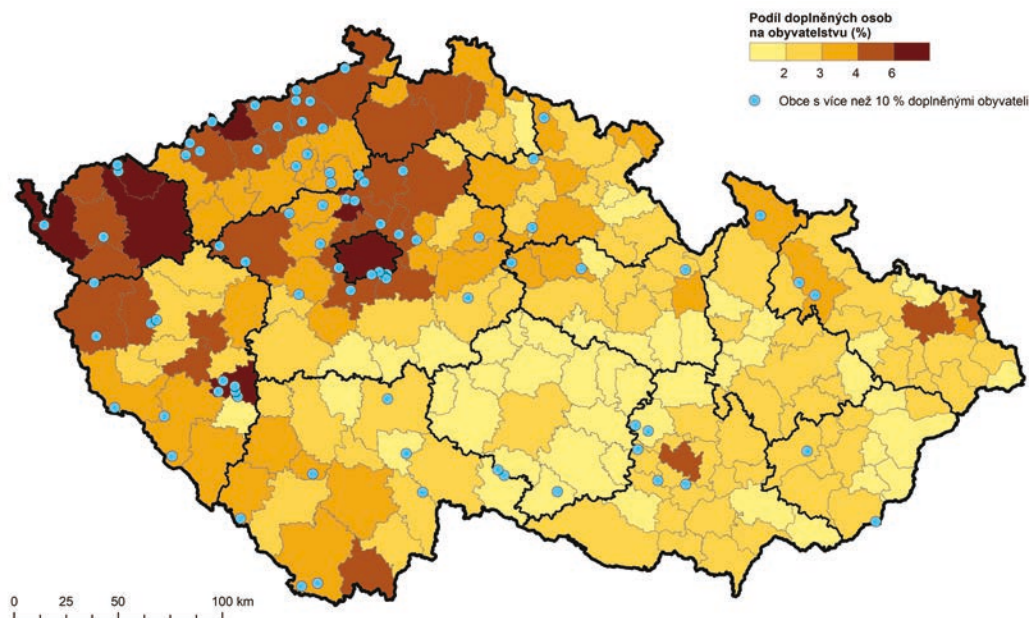
Vliv osob doplněných z ISEO na výsledné počty obyvatel byl diferencovaný i prostorově (obr. 3). Nejvý-

razněji se záznamy ISEO podílely na výsledném počtu obyvatel Prahy, kde představovaly 9 % trvale bydlících obyvatelstva. Relativně vysoké byly podíly doplněných osob v západní polovině území Česka, resp. zejména na severozápadě (Karlovarský kraj a příhraniční oblasti sousedící se Saskem). V porovnání se svým okolím byl podíl nápadně vysoký v ORP Nepomuk (7 % trvale bydlících bylo doplněno z ISEO). Nejméně doplněných (a tedy nejujplnější pokrytí dotazníkovým šetřením) bylo na Vysočině. Extrémní podíly doplněných osob vykazovaly některé obce. V Chlumech v okrese Plzeň-jih to byla rovná polovina výsledného počtu 196 obyvatel, v obci Selmice na Pardubicku 39 % (96 z 248 obyvatel), ve Výsluní na Chomutovsku 36 % (126 z 351 obyvatel). V Chlumech a Výsluní byly doplněnými osobami až na drobné výjimky občané Německa, v Semicích téměř výhradně Ukrajinci hlášení na adrese místní ubytovny.

Nejvýraznější vliv měly záznamy ISEO na počty cizinců. Ze všech cizinců zahrnutých do výsledků

Obr. 3: Podíly obyvatel doplněných z ISEO na trvale bydlícím obyvatelstvu správních obvodů ORP (sčítání lidu 2011)

Shares of inhabitants enumerated [counted] on the basis of ISEO in the permanent resident population at the microregional level (2011 Population and Housing Census)



Zdroj: Zpracovatelská databáze SLDB 2011.
Source: Processing database of 2011 Census.

sčítání bylo v případě obvykle bydlících sečteno doplněním z ISEO 36 % (171 792), mezi trvale bydlícími cizinci to bylo 40 % (174 507). Tab. 4 uvádí podíly doplněných osob na populacích cizinců vybraných občanství. Nejvyšší byl tento podíl mezi obyvateli s německým občanstvím, téměř dvě třetiny (obvykle bydlících), resp. tři čtvrtiny (trvale bydlících) Němců bylo do české populace doplněno z ISEO. Následovali cizinci s občanstvím USA, více než polovinou se ISEO podílelo i na početně nejvýznamnější skupině občanů Ukrajiny (v absolutních počtech to bylo 58, resp. 59 tisíc Ukrajinců). To je hlavní důvod, proč měla ve výsledcích více než polovina Ukrajinců, resp. kolem 70 % Němců nezjištěnou ekonomickou aktivitu či vzdělání (viz např. tabulka 159 ve výsledcích sčítání 2011).

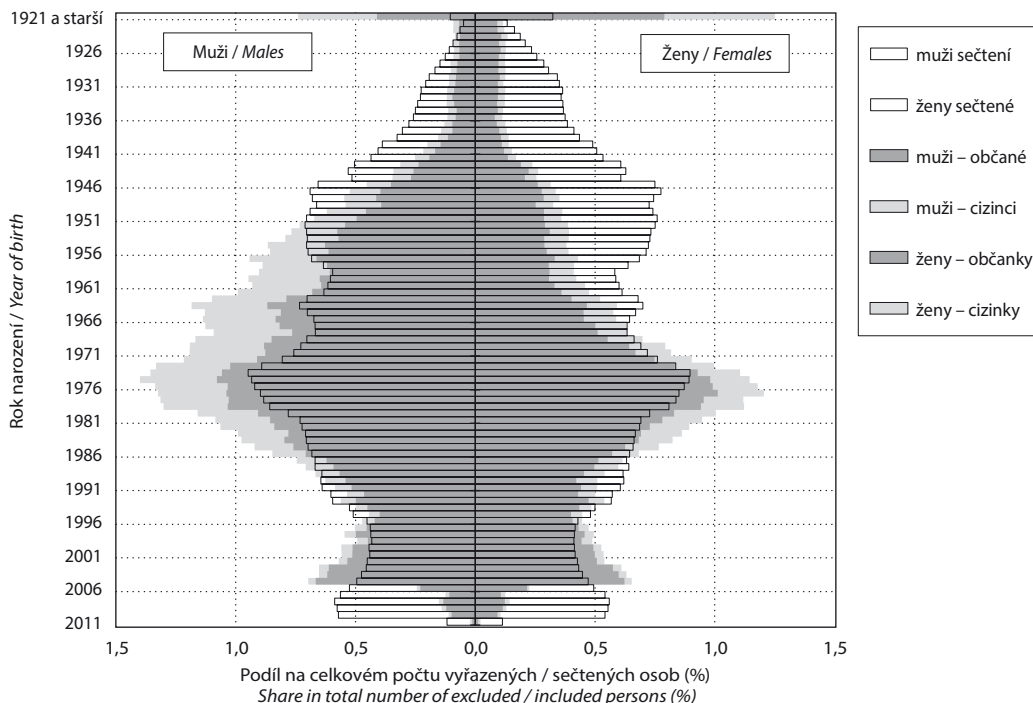
Záznamy ISEO, které výše popsané podmínky nespĺnily, představovaly výsledný přesah pokrytí. Bylo to 274 tisíc záznamů, tedy 2,6 % všech záznamů osob s platným evidovaným pobytem na území ČR. Na roz-

díl například od zmiňované situace ve Švédsku nelze tento přesah chápat čistě jako chybu evidence, protože u nás není stanovena povinnost oficiálního ukončení pobytu při odchodu do zahraničí. Záznam o platném pobytu osoby, která fakticky na našem území dlouhodobě nežije, proto není jednoznačně nesprávný. To je však spíše nedostatek platné právní úpravy, protože negativně ovlivňuje využitelnost klíčového zdroje dat o obyvatelstvu. V každém případě pro účely populační statistiky se jedná o přesah pokrytí, který je pro věrohodné výsledky nezbytně ošetřit.

Na obr. 4 je znázorněna struktura souboru 274 tisíc osob představujících přesah pokrytí podle věku a pohlaví. Ve srovnání se sečtenou populací se vyznačovala především relativně vysokým zastoupením mužů, osob ve věku kolem 25–40 let, resp. u mužů až zhruba do 55 let. Velmi výrazné bylo relativní zastoupení nejvyšších věkových kategorií (90 a více let). Charakteristické dále bylo relativně vysoké zastoupení cizinců

Obr. 4: Osoby evidované v ISEO vyřazené ze sčítání 2011 (přesah pokrytí) – struktura podle věku, pohlaví a státního občanství

Persons registered in ISEO excluded from the 2011 census (overcoverage) by age, sex, and citizenship



Zdroj: Zpracovatelská databáze SLDB 2011.
 Source: Processing database of 2011 Census.

a také vysoký podíl osob hlášených k pobytu v největších městech (téměř čtvrtina těchto osob z nich měla evidovaný pobyt v Praze, třetina ve třech největších městech). Všechny tyto charakteristiky patří k typickým rysům pozorovaným u osob představujících přesah pokrytí v populačních registrech v zahraničí.

Stojí za povšimnutí, že obě skupiny osob nesečtených v terénu (doplnění do sčítání i vyřazení) se v některých rysech podobají. Z toho je patrné, že cizinci, muži a osoby kolem 30 až 40 let věku představují skupiny, jejichž charakteristiky jsou zatíženy poměrně velkou nejistotou, jak v případě údajů z dotazníkového šetření, tak u administrativních dat.

6. ZÁVĚR

Přestože se ve sčítání lidu 2011 využily administrativní záznamy jen v omezené míře, na výsledky měly nezanedbatelný vliv. Nebýt údajů ISEO, měl by pokles sečtenosti obyvatel v terénu ve srovnání s předchozími cenzy za následek výrazně (v měřítkách požadavků na sčítání) neúplné výsledky, které by ve vývoji obyvatelstva vytvořily nevěrohodný „zub“ a na deset let negativně ovlivnily intercenzální populační statistiku. Pokles účasti na sčítání nebyl jednorázovým problémem sčítání 2011, opakoval se i ve sčítání 2021, jehož zpracování v současné době probíhá.

Sčítání lidu 2021 využívá výrazně více administrativních zdrojů než sčítání 2011. Z hlediska informací relevantních pro sčítání jsou to vesměs zdroje poměrně chudé, jejich nesporným přispěním je však rozšíření možností *signs-of-life* analýzy a tím přesnějšího vymezení výsledné populace. Stěžejními zdroji zůstávají jako v roce 2011 sčítací formuláře a evidence obyvatel (dnes již ne ISEO, ale soustava systémů *základní registr obyvatel – agendový informační systém evidence obyvatel – agendový informační systém cizinců*, z hlediska věcných informací však tato soustava odpovídá tehdejšímu ISEO).

K využití dat z registrů ve sčítání a zejména k čistě administrativním cenzům lze mít oprávněné výhrady. Administrativní zdroje vzdalují sčítání od některých jeho základních atributů, omezují flexibilitu z hlediska obsahu, nutí částečně přijímat definice proměnných vyhovujících nikoliv statistice, ale účelům, pro které jsou jednotlivé zdroje vedeny, v jistém smyslu berou cenzům možnost zkoumat „reálný“ stav populace

v terénu atd. Praxe sčítání 2011 i sčítání 2021 v souladu s poznatky ze zahraničí ovšem ukazuje, že i přes tyto výhrady jsou administrativní zdroje v současné době pro sčítání nejen přínosné, ale dokonce nutné. Bez administrativních zdrojů zřejmě již prakticky nelze dosáhnout požadované úplnosti výsledků a nelze vyhovět stále rostoucí poptávce po rychleji dostupných a častěji aktualizovaných informacích.

V reakci na tuto poptávku je v připravovaném novém evropském nařízení o populační statistice stanovena povinnost každoročního poskytování dat cenového typu v podrobném územním detailu. Ačkoliv bude nařízení opět výstupově orientované (definuje požadavky na výsledky, ale ponechává členským státům volnost ve způsobu jejich získání), prakticky si vynutí intenzivní využívání administrativních dat, protože bez nich nebude možné požadavky naplnit.

Dlouhodobým cílem je proto propracovat se přes kombinované sčítání ke sčítání čistě administrativnímu, tak jako k němu dospívá čím dál více evropských zemí. Velkou překážkou v dosažení tohoto cíle je však chudá datová základna. Informační systémy u nás nepokrývají některá klíčová témata, jako je úroveň vzdělání, dojíždka, charakteristiky bytů a domácností. Největším, a v dohledné době asi zřejmě neřešitelným úskalím je absence registru bytů. Informační systém katastru nemovitostí pokrývá jen část bytového fondu, navíc nevedou o bytech prakticky žádné pro sčítání relevantní údaje (evidovanou entitou ani nejsou byty, ale jednotky, které ne vždy s byty korespondují). Základní registr územní identifikace, adres a nemovitostí neeviduje byty vůbec. S tím souvisí i nejspolehlivější překážka na cestě k administrativnímu cenzu – adresy u nás končí na úrovni domu, nikoliv bytu, tj. i evidovaný pobyt obyvatel končí na úrovni domu. Z registru obyvatel tak nelze zjistit, kdo bydlí v kterém bytě, což vylučuje konstrukci jakýchkoliv údajů o domácnostech. Extrémně byrokraticky náročný proces hlášení změny pobytu, který u nás způsobuje velké nesoulady mezi místem evidovaného a místem obvyklého (faktického) pobytu, je ve srovnání s absencí evidence bytů druhořadý problém, přesto je významný. Mezi další komplikace patří neexistence registru vzdělání či absence údajů o místě pracoviště v administrativních zdrojích. To je výčet hlavních důvodů, proč se ani příští sčítání v roce 2031 s největší pravděpodobností neobejde bez dotazníkového šetření.

Literatura

- Aradhya, S. – Scott, K. – Smith, C. 2017. Repeat immigration: a previously unobserved source of heterogeneity? *Scandinavian Journal of Public Health*, Vol 45 (Suppl. 17), s. 25–29. <https://doi.org/10.1177/1403494817702334>.
- Bednářová, H. 2011. Sčítání s využitím poštovních služeb. *Demografie*, 53(1), s. 52–55.
- Bednářová, H. 2012. Vyplnění sčítacích formulářů prostřednictvím internetu. *Demografie*, 54 (3), 280–289.
- Bucher, H. 2014. Der Zensus 2011 und seine Auswirkungen auf demographische Eckwerte. *Europa Regional*, 20.2012(4), s. 147–159.
- Casale, D. et al. 2012. *The Use of Population Registers in the 15th Italian Census: Challenges and Preliminary Evidence*. UNECE-Eurostat Expert Group Meeting on Censuses Using Registers, Geneva, 22. 5.–23. 5. 2012.
- Česko 1995. Zákon č. 89 ze dne 29. dubna 1995 o státní statistické službě. In: *Sbírka zákonů České republiky*. 1995, částka 19.
- Česko 2020. Zákon č. 332 ze dne 22. 7. 2020 o sčítání lidu, domů a bytů v roce 2021 a o změně zákona č. 89/1995 Sb., o státní statistické službě, ve znění pozdějších předpisů. 2020, část 134.
- ČSÚ 2013. *Sčítání lidu, domů a bytů - Pramenné dílo*. Praha: Český statistický úřad.
- Čtrnáct, P. 2011. Ukončení etapy terénních prací sčítání lidu 2011. *Demografie*, 53(3), s. 238–240.
- Dolenč, D. 2012. Register-Based Census 2011 in Slovenia – novelties and lessons learned. Working paper, Neighbouring Countries Cooperation Initiative workshop on Novelties of 2011 Censuses and Challenges for the Future, 24. 4. 2012, Ljubljana.
- Dusetzina et al. 2014. *Linking Data for Health Services Research: A Framework and Instructional Guide*. Chapel Hill: AHRQ.
- Dygaszewicz, J. 2020. Transition from traditional census to combined and registers based census. *Statistical Journal of the IAOS*, 36(1), s. 165–175. <https://doi.org/10.3233/SJI-190566>.
- Eurostat. 2012. *A census is a census is a census?* Paris: UNECE Conference of European Statisticians, 60th plenary session. Dostupné z: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2012/25__A_census_is_a_census_is_a_census.pdf.
- Eurostat 2013. *Issues paper for the DSS Board and DSS discussions*. Meeting of the European directors of social statistics, Luxembourg, 18. 11. – 19. 11. 2013. <https://doi.org/10.7202/1013885ar>.
- Evropská unie. 2008. Nařízení Evropského parlamentu a Rady (ES) č. 763/2008 ze dne 9. 7. 2008 o sčítání lidu, domů a bytů. In: Úřední věstník Evropské unie. 2008.
- Evropská unie. 2017. 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. In: Úřední věstník Evropské unie. 2017.
- Fellegi, I. P. – Sunter, A. B. 1969. A Theory for Record Linkage. *Journal of the American Statistical Association*, 40, s. 1183–1210. <https://doi.org/10.1080/01621459.1969.10501049>.
- Habartová, P. 2016. *Rodiny a domácnosti ve sčítání lidu se zaměřením na metodologické aspekty dat*. Disertační práce, Praha: Univerzita Karlova
- Holá, B. 2005. Srovnatelnost statistiky zahraniční migrace. *Demografie*, 47(3), s. 177–187.
- INIS 2012. *Lessons learned from use of registers and geocoded databases in population and housing census*. UNECE Conference of European Statisticians, Sixtieth plenary session, Paris, 6. 6. – 8. 6. 2012.
- Josipovič, D. 2015. Slovenia and the Census: From the 20. Century Yugoslav Counts to the Register-based Census of 2011. *Contemporary Southeastern Europe*, 2(2), s. 159–175.
- Kukutai, T. – Thompson, V. – McMillan, R. 2015. Whither the census? Continuity and change in census methodologies worldwide, 1985–2014. *Journal of Population Research* Vol. 32, No 1, p. 3–22. <https://doi.org/10.1007/s12546-014-9139-z>.
- Lenk, M. 2008. Methods of Register-based Census in Austria. Statistik Austria. Dostupné z: https://international.ipums.org/international/resources/enum_materials_pdf/source_doc_at2011a.pdf.
- Maasing, E. – Tliit, E. M. – Vähi, M. 2016. Residency index – a tool for measuring the population size. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 21(1), s. 129–139. <https://doi.org/10.12697/ACUTM.2017.21.09>.
- Monti, A. – Drefahl, S. – Mussino, E. – Härkönen, J. 2019. Over-coverage in population registers leads to bias in demographic estimates. *Population Studies*, 74(3), s. 451–469. <https://doi.org/10.1080/00324728.2019.1683219>.
- Poullin, M. – Herm, A. 2013. Central Population Registers as a Source of Demographic Statistics in Europe. *Population*, 68(2), s. 183–212. <https://doi.org/10.3917/pope.1302.0183>.

- Scholz, R. – Kreyenfeld, M. 2016. The Register-based Census in Germany: Historical Context and Relevance for Population Research. *Comparative Population Studies*, 41(2), s. 175–204. <https://doi.org/10.12765/CPoS-2016-08>.
- Schulte Nordholt, E. 2014. Introduction to the Dutch Census 2011. In: Schulte Nordholt, E. – Van Zeijl, J. – a Hoeksma, L. eds. *Dutch Census 2011 – Analysis and Methodology*. The Hague: Statistics Netherlands.
- Schulte Nordholt, E. 2018. The usability of administrative data for register-based censuses. *Statistical Journal of the IAOS*, 34(4), s. 487–498. <https://doi.org/10.3233/SJI-180425>.
- Schulte Nordholt E. – Hartgers, M. – Gircour, R. 2004. The Dutch Virtual Census of 2001 – analysis and Methodology. Hague: Statistics Netherlands.
- Skinner, C. 2018. Issues and challenges in census taking. *Annual Review of Statistics and its Application*, 5, s. 49–63. <https://doi.org/10.1146/annurev-statistics-041715-033713>.
- Šanda, R. 2015. Obvyklý a trvalý pobyt ve sčítání lidu 2011. *Demografie*, 57(2), 113–126.
- Šanda, R. 2020. Vyšetřenosť údajů o dojíždě ve sčítání lidu v roce 2011 a jejich rekonstrukce pomocí metody Jaro-Winkler. *Demografie*, 62(1), s. 27–42.
- Thorvaldsen, G. – Solli, A. 2012. Norway: From Colonial to Computerized Censuses. *Revista de Demografia Histórica*, 1, s. 107–136.
- UN 2017. *Principles and Recommendations for Population and Housing Censuses*. New York: United Nations.
- UNECE 2007. *Register-based statistics in the Nordic countries*. New York and Geneva: United Nations.
- UNECE 2012. Overview of the 2010 round of population and housing censuses in the UNECE region. Group of Experts on Population and Housing Censuses, Geneva, 24. 5. – 25. 5. 2012.
- UNECE 2014. *Measuring population and housing – Practices of UNECE countries in the 2010 round of censuses*. New York and Geneva, United Nations.
- UNECE 2015. *Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing*. New York and Geneva: United Nations.
- UNECE 2018. *Guidelines on the use of registers and administrative data for population and housing censuses*. Geneva: United Nations Economic Commission for Europe.
- Valente, P. 2010. Census taking in Europe: How are populations counted in 2010? *Population and Societies*, 467, s. 1–4.
- Valente, P. 2014. Innovative Approaches to Census-Taking: Overview of the 2011 Census Round in Europe. In Crescenzi, F. – Mignani S. eds. *Statistical Methods and Applications from a Historical Perspective*, s. 187–200, Springer International Publishing. https://doi.org/10.1007/978-3-319-05552-7_17.
- Zelený, M. 2006. *Míra ohrožení chudobou jako indikátor chudoby v Evropě – analýza vlastností a vypovídací schopnosti*. Disertační práce, Praha: Vysoká škola ekonomická v Praze.

ROBERT ŠANDA

Je absolventem Geografického ústavu Přírodovědecké fakulty Masarykovy univerzity, kde v současnosti pokračuje v doktorském studiu programu sociální geografie a regionální rozvoj. Působí na Českém statistickém úřadě v odboru statistiky obyvatelstva. Podííl se na přípravě sčítání lidu, domů a bytů v roce 2021.

SUMMARY

The move from traditional to register-based censuses is the most significant trend in recent developments in European population censuses. While in the 2000 census round only a few censuses in Europe employed administrative data, in 2021 more than two-thirds of EU members conducted a combined or fully register-based census.

The substantial cost reduction and the possibility of reducing the length of the census period are two significant advantages often assigned to register-based censuses. The most important reason for moving from traditional field enumeration is the public's decreasing willingness to participate in the census. However, there also serious drawbacks

to the use of a registerbased census, such as a dependency on register holders, different definitions of variables etc. These problems are often mentioned in literature. Very limited flexibility in terms of content is another severe (but generally overlooked) disadvantage. Only topics covered by existing administrative sources can be included as content in registerbased censuses; this means that it is impossible to include especially declaratory characteristics (health conditions, religion, ethnic groups etc.). Register-based censuses are also not fully consistent with the essential features of censuses in general.

Despite legitimate reservations, the use of administrative sources in censuses is becoming inevitable, because of growing public aversion to multiple providing personal data. Therefore, the Czech Statistical Office is following the Europe-wide trend and aims to move from a traditional to a fully register-based census. The 2011 census was based on a combination of field enumeration and administrative records from the central population register. For the 2021 census the number of registers used increased substantially. However, a persistent lack of essential information on households, dwellings, education, and commuting means it is still impossible to conduct a fully registerbased census.