

2. Sčítání lidu, domů a bytů 2021

Každé sčítání lidu je něčím specifické. Předchozí sčítání, konané v roce 2011, bylo jedním z nejvýznamnějších milníků v historii tzv. moderních cenzů, protože zavedlo celou řadu zásadních inovací, byť je někdy aplikovalo spíše okrajově. Sčítání v roce 2021 tolik principiálních změn nepřineslo, ale bylo charakteristické především tím, že novinky z roku 2011 plně rozvinulo a přetvořilo je v základní pilíře celého procesu.

Platí to zejména o využití administrativních zdrojů dat. Sčítání 2011 bylo v Česku prvním tzv. kombinovaným cenzem, který vedle dotazníkového šetření využíval i administrativní data. Byl však tehdy k dispozici jen jeden administrativní zdroj dat o osobách, který sloužil téměř výhradně jako náhradní zdroj základních charakteristik v případě nevyplnění sčítacího formuláře.² Sčítání 2021 bylo již plnohodnotným kombinovaným cenzem, který přebíral údaje o osobách zhruba z deseti administrativních zdrojů. Řada stěžejních charakteristik (mj. státní občanství, rodinný stav...) se již v dotazníkovém šetření nezjišťovala a administrativní záznamy byly pro ně výhradním zdrojem. Sčítací formuláře v roce 2021 tak byly z hlediska počtu otázek historicky zdaleka nejušpornější. Administrativní zdroje také pomohly výrazně zmírnit dopad non-response.

Dalším příkladem jsou elektronické sčítací formuláře. Ty byly také poprvé zavedeny ve sčítání 2011 – vyplňovaly se v off-line v PDF formátu a jejich prostřednictvím se sečetla zhruba čtvrtina respondentů. Sčítání 2021 nabídlo komfortní, dynamický a velmi snadno dostupný (i na mezinárodní poměry) online formulář, který využilo zhruba 85 % sečtených obyvatel. Listinný formulář byl sice stále k dispozici (na rozdíl od řady jiných zemí včetně např. Polska, Slovenska či Maďarska), uplatnil se však již jen doplňkově.

Terénní práce, zejména distribuci a sběr formulářů, zajišťovala až na specifické případy Česká pošta. I tím Sčítání 2021 navázalo na novinku z předchozího cenzu, přičemž spolupráce mezi Českou poštou a ČSÚ byla zdokonalena, zejména modernizací sdílení logistických informací. Charakteristické bylo také prohloubení mezinárodní spolupráce, Sčítání 2021 lze již označit za primárně mezinárodní statistickou akci. V oblasti metodiky a zpracování výsledků se výrazně rozvinuly některé postupy, které byly v předchozím cenzu aplikovány jen velmi omezeně. Je to například pravděpodobnostní propojování dat z jednotlivých zdrojů nebo tzv. signs-of-life analýza, která podstatně ovlivnila celkové výsledky sčítání a „rozhodla“ o početně nejvýraznější korekci intercenzálních odhadů počtu obyvatel Česka v historii moderních cenzů. O většině výše zmíněných témat blíže pojednávají následující kapitoly.

Jeden aspekt Sčítání 2021 byl však zcela jedinečný, a sice podmínky, za jakých se uskutečnilo. V březnu až květnu 2021, kdy probíhaly terénní práce, život v Česku i jinde ve světě intenzivně ovlivňovala pandemie nemoci Covid-19. V některých zemích tak muselo být sčítání odloženo, někdy až na rok 2022 (z EU Irsko, Maďarsko, Německo). O posun termínu sčítání se intenzivně debatovalo i u nás. Kombinace snadno přístupných online formulářů a (v zahraničí již zastaralé) možnosti vyplnit listinný formulář samotným respondentem³ však nakonec umožnila provést sčítání v původním termínu při dodržení přísných protiepidemických opatření. Původní plán průběhu terénních prací však musel být na poslední chvíli znatelně upraven, aby byl splněn požadavek na minimální osobní kontakt sčítacích komisařů s respondenty. Nejvýraznější změny spočívaly v zákazu vstupu komisařů do vnitřních prostor budov a odevzdávání vyplněných formulářů prakticky výhradně poštou a řešení všech z toho vyplývajících konsekvencí. Přejeme si, abychom nabyté zkušenosti z této situace nepotřebovali při příštích sčítáních uplatnit.

² V některých mezinárodních srovnáních je proto české sčítání 2011 řazeno stále do kategorie tzv. tradičních cenzů.

³ Země EU, které v roce 2021 ještě jako součást sčítání prováděly celoplošné dotazníkové šetření, využívaly spíše kombinaci online sčítání a tzv. CAPI šetření (rozhovor respondenta se sčítacím komisařem, který odpovědi zaznamenával do elektronického formuláře).



2.1 Právní úprava sčítání

Stejně jako předešlá sčítání se také Sčítání 2021 řídilo samostatným zákonem, který vyšel ve sbírce zákonů v roce 2020. Jde o *zákon 332/2020 Sb. ze dne 22. července 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*. Zákon ustanovuje především

- přesný termín sčítání (půlnoc z pátku 26. března 2021 na sobotu 27. března 2021),
- rozsah sčítání (kdo (co) přesně podléhá sčítání),
- přesný výčet otázek na formuláři za jednotlivé entity,
- údaje, které má Český statistický úřad uchovávat v CIS (Cenzovém informačním systému),
- údaje přebírané z informačních systémů veřejné správy,
- povinnosti zúčastněných stran (fyzických či právnických osob, ČSÚ, ministerstev, obcí a jiných subjektů),
- okolnosti provedení sčítání (tj. sčítací formuláře, online sčítání, terénní došetřování, vymezení sčítacích obvodů, sčítací komisaře, jejich povinnosti, průkaz sčítacího komisaře).

Dále zákon zmocňuje Český statistický úřad k vydání prováděcí vyhlášky, jež má obsahovat vzor sčítacího formuláře, formu přístupu ke sčítacímu formuláři v elektronické podobě a vzor průkazu sčítacího komisaře. *Vyhláška ze dne 23. listopadu 2020*, kterou se provádějí některá ustanovení zákona č. 332/2020 Sb., 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 byla vydaná ve Sbírce zákonů v prosinci 2020.

Dotčen byl také zákon o státní statistické službě, který byl doplněn o paragraf 12a, který stanoví Českému statistickému úřadu povinnost zřídit Cenzový informační systém. Jedná se o informační systém veřejné zprávy vzniklý za účelem naplnění povinností dané Evropskou unií v oblasti migrace a mezinárodní ochrany, evropské demografické statistiky, sčítání lidu, domů a bytů.

Také v evropském měřítku došlo k aktualizaci legislativních předpisů vztahujících se ke sčítání. Byla vydána čtyři nová nařízení Evropské komise, která se vztahují ke stále platnému a klíčovému *Nařízení Evropského parlamentu a Rady (ES) č. 763/2008 ze dne 9. července 2008 o sčítání lidu, domů a bytů*, které koordinovalo již předešlé sčítání v roce 2011.

Jako první bylo vydáno *Prováděcí nařízení Komise (EU) 2017/543⁴* ze dne 22. března 2017, které se věnuje technickým specifikacím témat a jejich rozdělení. Dokument předepisuje přesnou strukturu a úroveň detailu témat povinně poskytovaných Eurostatu. Určuje také některé metodické a technické požadavky tak, aby výstupy jednotlivých členských států byly v co možná nejvyšší míře srovnatelné. Druhým právním ustanovením je *Nařízení Komise (EU) 2017/712⁵* ze dne 20. dubna 2017, kterým se zavádí referenční rok a program statistických údajů a metadat. Definuje základní pojmy, za referenční rok stavuje rok 2021 a především obsahuje předpis tabelací dat tak, jak mají být poskytována Eurostatu. Stanoví také rozsah metadat. *Prováděcí nařízení Komise (EU)⁶ 2017/881* ze dne 23. května 2017 stanoví technický formát předávání dat, rozsah a obsah zprávy o kvalitě. Nově je evropskou legislativou popsáno téma posledního *Prováděcího nařízení Komise (EU)⁷ 2018/1799* ze dne 21. listopadu 2018. Věnuje se diseminaci vybraných témat ze sčítání lidu v souřadnicové síti s buňkou o ploše 1 km².

⁴ PROVÁDĚCÍ NAŘÍZENÍ KOMISE (EU) 2017/543 ze dne 22. března 2017, kterým se 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. Úřední věstník Evropské Unie L 78/13

⁵ NAŘÍZENÍ KOMISE (EU) 2017/712 ze dne 20. dubna 2017, 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. Úřední věstník Evropské Unie L 105/1

⁶ 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. Úřední věstník Evropské Unie L 135/6

⁷ PROVÁDĚCÍ NAŘÍZENÍ KOMISE (EU) 2018/1799 ze dne 21. listopadu 2018 o zavedení přechodné přímé statistické akce pro šíření vybraných témat sčítání lidu, domů a bytů v roce 2021 geokódovaných v souřadnicové síti s buňkou o ploše 1 km². Úřední věstník Evropské Unie L 296/19

Na celosvětové úrovni jsou cenzy koordinovány Statistickou divizí OSN. Pro populační cenzy konající se okolo roku 2020 byl zrevidován dokument *Principles and Recommendations for Population and Housing Censuses*, jehož třetí revizi schválila Ekonomická a sociální rada OSN v roce 2015. Rozsáhlý dokument obsahuje doporučení jak z oblasti organizace cenzu, tak z oblasti věcné náplně cenzu. Stanoví také období mezi lety 2014-2024, kdy každá členská země musí alespoň jednou provést populační census.

2.2 Obsah sčítání

Údaje zjišťované ve Sčítání 2021 lze rozdělit do čtyř hlavních entit: obyvatelstvo, domy, byty a domácnosti. Snahou bylo minimalizovat zátěž kladenou na respondenty, a tak došlo oproti SLDB 2011 k vyřazení některých údajů, které byly z hlediska účelu sčítání vyhodnoceny jako nadbytečné a málo relevantní. Některé údaje uvedené v této kapitole nebyly součástí sčítacích formulářů, neboť jsou plně dostupné v administrativních zdrojích dat, což koresponduje se snahou o maximální využití dostupných registrů.

V rámci entity obyvatelstvo bylo zjišťováno celkem 22 proměnných. Jednalo se o:

- sociodemografické charakteristiky (věk, pohlaví, rodinný stav (včetně registrovaného partnerství), státní občanství, národnost (nepovinný údaj), náboženská víra (nepovinný údaj), nejvyšší dosažené vzdělání (pouze u osob 15letých a starších), počet živě narozených dětí (pouze u žen 15letých a starších), mateřský jazyk)
- ekonomické charakteristiky (ekonomická aktivita, zaměstnání (pouze u zaměstnaných), postavení v zaměstnání (pouze u zaměstnaných), odvětví ekonomické činnosti (pouze u zaměstnaných))
- lokalizace (místo obvyklého pobytu, místo registrovaného pobytu, místo bydliště jeden rok před sčítáním, místo bydliště v době narození, rok přistěhování do země)
- způsob bydlení
- údaje o dojíždě (místo pracoviště nebo školy (zjišťovalo se pouze u zaměstnaných osob nebo žáků a studentů), frekvence dojíždě (zjišťovalo se pouze u vyjíždějících) a dopravní prostředek použitý pro cestu do práce nebo školy (zjišťovalo se pouze u vyjíždějících, bylo možné uvést hlavní dopravní prostředek použitý k nejdelší části cesty a až 2 další dopravní prostředky použité ke zbytku cesty)

Oproti sčítání 2011 se nezjišťoval obor vzdělání a počet živě narozených dětí v současném nebo posledním manželství (zjišťoval se pouze celkový počet živě narozených dětí). Významná metodická změna byla realizována u otázky na ekonomickou aktivitu, kde bylo pozměněno vymezení jednotlivých kategorií. Nově se rozlišovaly osoby na rodičovské dovolené jako samostatná kategorie. Naopak se už nerozlišují nezaměstnané osoby hledající první zaměstnání. V tématech týkajících se dojíždě bylo nejvýznamnější metodickou změnou zjišťování místa pracoviště u pracujících studentů a žáků. Ve sčítání 2011 se u pracujících studentů a žáků zjišťovalo místo školy. Na rozdíl od sčítání 2011 se také nezjišťovala časová délka vyjíždě. Frekvence dojíždě a dopravní prostředek byly k dispozici pouze ze sčítacích formulářů. Údaje o místu pracoviště nebo školy byly ve zpracování doplněny údaji z registrů.

Některé proměnné za osoby nebyly součástí sčítacího formuláře, neboť byly plně k dispozici v administrativních zdrojích dat. Jedná se o: postavení v zaměstnání, rodinný stav (včetně registrovaného partnerství), rok přistěhování do země, místo registrovaného pobytu a státní občanství.

Za domy bylo zjišťováno celkem 10 proměnných: obydlenost domu, počet bytů v domě, druh domu, druh vlastníka, období výstavby nebo rekonstrukce, materiál nosných zdí, počet nadzemních podlaží, připojení na odpad, způsob vytápění a výtah. Oproti sčítání 2011 nedošlo k vyřazení žádné proměnné, ale změnil se zdroj dat, neboť všechny údaje za domy jsou plně k dispozici z administrativních zdrojů dat a nebyly ve Sčítání 2021 součástí sčítacího formuláře. Metodická změna byla realizována u způsobu



vytápění, kde se již nerozlišoval druh paliva. Dále se také nezjišťoval důvod neobydlenosti domu. Podrobněji viz kapitola 14. Domovní fond.

Entita byty zahrnuje celkem 8 proměnných: obydlenost bytu, právní důvod užívání bytu, velikost bytu (vyjádřená podlahovou plochou a počtem místností včetně kuchyně), poloha bytu v domě (patro), připojení na plyn, připojení na vodovod, způsob vytápění, zdroj energie/druh paliva používaný k vytápění bytu (jeden hlavní a případně další doplňkové zdroje). Oproti sčítání 2011 byly vyřazeny otázky na teplou vodu, záchod, koupelnu, osobní počítač a připojení k internetu. Nezjišťuje se ani důvod neobydlenosti bytu. Údaje za byty nejsou v administrativních zdrojích k dispozici, s výjimkou obydlenosti bytu tak byly všechny údaje součástí sčítacího formuláře. Podrobněji viz kapitola 15. Bytový fond.

Pro potřeby odvození entity domácností byly na sčítacím formuláři zjišťovány vzájemné vztahy mezi společně bydlícími osobami. Oproti sčítání 2011 se výrazně změnila podoba otázky na sčítacím formuláři, která se skládala ze dvou hlavních částí: rozdělení osob do jednotlivých hospodařících domácností a zápis vztahů mezi osobami. Údaje získané ze sčítacích formulářů byly v průběhu zpracování vhodně doplněny údaji z administrativních zdrojů (především soustava registr obyvatel, informační systém evidence obyvatel a cizinecký informační systém). Následně byly odvozeny základní ukazatele definující domácnost: typ domácnosti, velikost hospodařící domácnosti a počet závislých dětí. Podrobněji viz kapitola 13. Domácnosti.

2.3 Územní příprava, informační pochůzka

Územní přípravou se v terminologii sčítání označuje souhrn tematicky zaměřených činností, kterými se před vlastním provedením sčítání zajišťuje kompletace a aktualizace informací o území, v němž se census koná, zejména o objektech, které jsou předmětem sčítání. Územní příprava Sčítání 2021 započala již více než dva roky před samotným sběrem dat od domácností na začátku roku 2019 a předání souboru referenčních dat z územní přípravy proběhlo v lednu roku 2021.

Společným cílem aktivit územní přípravy bylo připravit co nejkvalitnější bázi podpůrných popisných a lokalizačních dat o sčítacích obvodech, domech, adresách a bytech pro navazující fáze Sčítání 2021, aby evidované údaje v databázi územní přípravy za jednotlivé entity co nejvíce korespondovaly se skutečným stavem v terénu. To znamená, aby byly co nejúplnější, věcně přesné a časově co nejaktuálnější. Výsledný soubor referenčních dat územní přípravy za budovy, adresy, byty a sčítací obvody totiž následně sloužil nejen jako výchozí datový zdroj pro přípravu podkladů pro sčítací komisaře při terénních pracích, ale i jako výchozí seznam adres a bytů pro on-line sčítání a referenční databáze pro ověřování existence adres během informační pochůzky.

Vedle pracovníků ČSÚ se na ověřování a aktualizaci údajů o budovách a jejich adresách podílely vybrané obecní a stavební úřady a osloveni byli i vlastníci bytových domů z řad bytových družstev a ostatních právnických subjektů. V neposlední řadě měli ověřeni správnosti evidence budov a jejich vybraných charakteristik a seznamů bytů provést i sčítací komisaři během informační pochůzky, nicméně kvůli nepříznivé epidemické situaci bylo došetřování těchto údajů zejména v bytových domech nakonec výrazně omezeno. Část aktivit územní přípravy souvisejících s organizací sběru dat v terénu zajišťoval centrálně dodavatel terénních prací, Česká pošta s. p. Jednalo se zejména o výběr provozoven České pošty pro zřízení kontaktních míst pro veřejnost a sběrných míst pro sčítací komisaře, nebo o plánování pochůzkových tras pro sčítací komisaře České pošty.

Specifickým cílem územní přípravy Sčítání 2021, který vycházel z obecného cíle Sčítání 2021 snižovat administrativní zátěž kladenou na participující subjekty, bylo dosažení významného poklesu počtu obcí zapojených do územní přípravy i rozsahu jejich úkolů a povinností ve srovnání se Sčítáním 2011. V souladu s tímto přístupem nebyly při územní přípravě kontaktovány s žádostí o spolupráci všechny obce, ale pouze ty, v jejichž evidencích, identifikacích či lokalizacích budov a adres byly shledány nějaké nedostatky či nekonzistence, které vyžadovaly jejich ověření či došetření.

V zájmu zefektivnění a optimalizace plnění jednotlivých úkolů územní přípravy došlo k nahrazení většiny tištěných podkladů digitálními. To se týkalo zejména elektronizace procesů správy, monitoringu a řešení

většiny úkolů pracovníky ČSÚ ve speciálně vyvinuté GIS aplikaci pro podporu územní přípravy provozované jednak na přenosných počítačích (noteboocích), jednak na mobilních telefonech využívaných při terénních pracích ke sběru popisných a lokalizačních údajů nad digitálními mapovými podklady. Speciální webové aplikace pro ověřování evidovaných a sběr nových dat o budovách však byly vytvořeny i pro stavební úřady či oslovené vlastníky bytových domů.

Do období realizace hlavních činností územní přípravy významně zasáhla pandemie Covid-19. Ta měla negativní dopad zejména na ochotu oslovených subjektů ke spolupráci a také na vnímání realizovaných aktivit veřejností. Z organizačního hlediska měla za následek potřebu pružně přizpůsobovat plánované úkoly aktuální epidemiologické situaci (preferenci využití on-line zdrojů na úkor osobního šetření v terénu a minimalizace fyzického kontaktu s veřejností) a nezbytnost dodržovat platná mimořádná preventivní opatření.

Jelikož územní příprava neprobíhala celoplošně a i kvůli časové prodlevě mezi její realizací a vlastním sčítáním, nemohly její datové výstupy zcela přesně reflektovat aktuální stav území v době sběru dat. Proto bylo tradičním předstupněm provedení sčítání ověření správnosti a úplnosti evidovaných informací o budovách v seznamu objektů a vchodů a o bytech v seznamu bytů sčítacími komisaři přímo v terénu. Tato přípravná fáze se nazývá informační pochůzka a při Sčítání 2021 probíhala od 10. 3. do 19. 3. 2021. Vedlejším účelem informační pochůzky bylo on-site seznámení komisaře s jemu přiděleným sčítacím obvodem, jeho případnými specifiky či posouzení potřeby zajištění specializovaného asistenta pro terénní došetření. Po jejím skončení byl zhruba týden na prověření a promítnutí zjištěných rozdílů do seznamu adres a bytů použitým pro on-line sčítání a terénní došetřování.

2.4 Sběr dat

Sčítání 2021 bylo koncipováno jako kombinované, sběr dat tak byl realizován jednak prostřednictvím sčítacích formulářů, a jednak byly v široké míře využity administrativní zdroje dat (AZD).

2.4.1 Sčítání prostřednictvím sčítacích formulářů

Pro sčítání prostřednictvím sčítacích formulářů byl použit jeden **společný formulář pro domácnost (SF)**. Při sčítání v letech 2001 a 2011 existovaly pro sčítání osob, bytů/domácností a domů zvláštní oddělené formuláře, propojování záznamů z nich prostřednictvím systému identifikátorů pak probíhalo v několika fázích během zpracování údajů. Pro Sčítání 2021 byl sčítací formulář koncipován jako jeden společný pro vyplnění údajů o bytě/bydli a o všech v něm bydlících osobách, včetně jejich příbuzenských vztahů a společného hospodaření. Tento způsob sběru dat zajišťoval propojení údajů o osobách, bytech a domácnostech z velké části již na vstupu. To mimo jiné umožnilo naplnit klíčové požadavky na způsob provedení sběru dat sčítání – realizovat sčítání primárně a převážně elektronicky a minimalizovat nutnost kontaktu povinných osob se sčítacím komisařem či kontaktním místem. Údaje o domech byly pro Sčítání 2021 kompletně získány z AZD, případně odvozeny či doplněny z územní přípravy terénního šetření, nebyly tedy již předmětem zjišťování prostřednictvím formulářů.

Sčítací formulář bylo možno vyplnit v tištěné podobě – **listinný sčítací formulář (LSF)** nebo v elektronické (ESF). Již při sčítání v roce 2011 bylo vůbec poprvé umožněno vyplnit sčítací formulář kromě klasického tištěného, který ovšem zůstal základem terénního zjišťování, také v elektronické podobě. Pro Sčítání 2021 se však **elektronický formulář stal primárním kanálem sběru dat**. Odpovídal tomu nejen důraz při propagaci a zajištění maximální dostupnosti elektronického sečtení, ale především samotná koncepce a časový harmonogram zjišťování. První fáze sčítání (první dva týdny po rozhodném okamžiku) probíhala jako online sčítání. Teprve po několikadenním vyhodnocení online fáze měly být distribuovány listinné formuláře k těm osobám a domácnostem, které se nesečetly elektronicky. Vzhledem k aktuální situaci a panujícím rozsáhlým protiepidemickým opatřením v souvislosti s onemocněním Covid-19 byly organizační plány v týdnech bezprostředně před zahájením



sčítání značně pozměněny a elektronické sčítání nabylo ještě většího významu než v původní koncepci. Možnost sečíst se elektronicky zůstala zachována po celé období plánované pro sběr dat, souběžně s možností vyplnit tištěný formulář. Celkem bylo nakonec odevzdáno 4 217 261 ESF a 816 347 LSF. Z celkového počtu 5 033 608 sebraných formulářů tak elektronické formuláře představovaly 83,8 %.

2.4.1.1 Elektronický sčítací formulář

Rozhodný okamžik Sčítání 2021 byl stanoven na 27. března 2021. První fáze sčítání byla připravena jako online. Od 27. 3. se lidé mohli sečíst online přes webovou (na www.onlinescitani.cz) nebo mobilní aplikaci Sčítání21 (na App Store nebo Google Play) bez nutnosti dalšího kontaktu se sčítacím komisařem, pouze s přístupem k internetu, což byl vzhledem k epidemické situaci nejbezpečnější způsob sečtení. Původně byla online fáze sčítání plánována na dva týdny (do 9. 4.), ale později byla prodloužena až do 11. 5. 2021.

Pro přístup do elektronického sčítacího formuláře bylo nutné osobu ověřit. Bylo možné využít tři způsoby pro ověření:

1. zadáním kombinace osobních údajů: číslo osobního dokladu (občanský průkaz nebo cestovní pas), typ osobního dokladu a datum narození
2. přes datovou schránku fyzické osoby
3. prostřednictvím elektronické identity (např. bankovní identity).

Sčítací formulář byl koncipovaný jako společný formulář za domácnost, takže stačilo, když jeden člen domácnosti založil online formulář, vyplnil údaje za sebe a poté za všechny ostatní členy domácnosti. Mohl také vygenerovat odkaz na sčítací formulář a odeslat ho členům domácnosti, kteří mohli data doplnit sami. Elektronický formulář bylo také možné vyplnit částečně, uložit a později se k němu vrátit a chybějící data doplnit.

Online formulář byl k dispozici kromě češtiny v sedmi dalších jazycích: angličtina, němčina, polština, romština, ruština, ukrajinština, vietnamština.

Online formulář byl věcně totožný s listinným formulářem, obsahoval všechny otázky. Byl však interaktivní, takže se uživatel zobrazovaly pouze relevantní otázky, například muži nevyplňovali otázku na počet živě narozených dětí nebo otázka na nejvyšší ukončené vzdělání se zobrazila pouze osobám 15tiletým a starším. Ve formuláři byly zabudované kontroly, které kontrolovaly přípustné hodnoty, například na platné datum. Většina otázek nabízela výběr z možností, kdy respondent vybral jednu, případně více správných odpovědí. U otevřených otázek, například u otázky na zaměstnání, se odpovídalo zápisem do textového pole s pomocí našeptávače, který nabízel různá zaměstnání ze slovníku. Bylo však i možné zapsat text vlastními slovy. U otázek na adresu obvyklého pobytu respondent vybíral platné adresy ze seznamu adres. Pokud šlo o adresu bytového domu s bytovými jednotkami, bylo možné vybrat i konkrétní byt. V případě adresy místa pracoviště nebo školy se také vybíralo ze seznamu adres, pokud respondent přesnou adresu neznal, mohl vybrat alespoň obec. U dobrovolných otázek na národnost a náboženskou víru se našeptávač nevyužíval, bylo nutné vyplnit text odpovědi ručně. Všechny otázky, s výjimkou dvou posledně zmíněných, byly povinné a online formulář vyžadoval jejich vyplnění.

Po odeslání vyplněného formuláře se respondentům zobrazilo potvrzení o odeslání formuláře. V případě odeslání chybných údajů bylo možné založit nový formulář a odeslat ho znovu se správnými údaji. Případné duplicity formulářů byly řešeny později ve zpracování.

2.4.1.2 Listinný sčítací formulář

Osoby, které se z jakéhokoliv důvodu nemohly sečíst online, vyplnily v druhé fázi sčítání klasický listinný formulář, a to od 17. dubna do 11. května. Ten získaly od sčítacích komisařů spolu s odpovědní obálkou. Vyplněné formuláře bylo možné zdarma vhodit do jakékoliv poštovní schránky, poslat na P. O. Box České Pošty nebo odevzdat na kontaktním místě sčítání. Vzhledem k nepříznivé epidemiologické situaci byl plánovaný sběr listinných formulářů sčítacími komisaři zrušen.

Listinný sčítací formulář rovněž odpovídal koncepci společného formuláře. Měl proto podobu sešitu o formátu přibližně A4. V základní podobě – Formulář pro domácnost – měl 16 stran a obsahoval kromě úvodních textů, návodů k vyplnění a vysvětlivek části pro vyplnění údajů o bydlení, resp. o bytě, údajů o složení domácnosti a dvoustranu pro vyplnění údajů za každou osobu. Pro domácnosti, kde žilo více než 5 osob, existoval Dodatečný formulář pro domácnost, který byl podobně koncipovaný (sešit o 16 stranách). Měl být používán ve spojení se základním LSF pro domácnost a umožňoval vyplnit údaje za dalších až 6 osob. Zjednodušenou verzi LSF (dvojlist – 4 strany A4) pak představoval Samostatný formulář pro osobu, který byl určen pro vyplnění za osoby žijící jednotlivě v zařízeních, příp. osoby bez domova. Obsahoval proto pouze dvoustranu určenou k vyplnění údajů za jednu osobu, nezahrnoval údaje o bytě a složení domácnosti.

Distribuce listinných sčítacích formulářů probíhala v rámci sčítacích obvodů, které byly rozděleny do 3 typů. Základním typem byly standardní sčítací obvody, kde terénní práce prováděli sčítací komisaři České pošty. Celkem jich bylo 9 861 a každý sčítací obvod obsluhoval právě jeden sčítací komisař České pošty. Dále měla Česká pošta k dispozici náhradní sčítací komisaře pro případné mimořádné situace. Druhým typem byly samostatné sčítací obvody, které vznikly v hromadných ubytovacích zařízeních, a kde operovali sčítací komisaři Českého statistického úřadu. Celkem jich bylo 1936 a v každém z nich působil alespoň 1 sčítací komisař. Posledním typem byly resortní sčítací obvody, kde měly Sčítání na starosti jednotlivé spolupracující resorty (např. sčítání ve věznicích bylo realizováno ve spolupráci s Vězeňskou službou ČR v kompetenci Ministerstva spravedlnosti).

2.4.2 Sčítání prostřednictvím administrativních zdrojů dat

2.4.2.1 Administrativní zdroje dat za osoby

Škála administrativních zdrojů dat (AZD) poskytujících údaje o osobách se mezi sčítáními 2011 a 2021 výrazně rozšířila. Zatímco pro sčítání 2011 byl k dispozici pouze omezený počet údajů o osobách z *Informačního systému evidence obyvatelstva* (ISEO), o deset let později byla dostupná nejen celá soustava registrů z Ministerstva vnitra, ale i data z dalších rezortů.

Dostupnost AZD pro účely Sčítání 2021 se výrazně zvýšila po přijetí Nařízení Evropského parlamentu a Rady (EU) 2015/759 z dubna 2015, které upravovalo Nařízení (ES) č. 223/2009 o evropské statistice. Tato novela upravuje přístup národních statistických institucí k administrativním zdrojům dat a jejich využívání pro statistické účely. Umožňuje rychlý a bezplatný přístup k administrativním záznamům.

Ve Sčítání 2021 byly AZD využity třemi způsoby:

- 1) Pro doplnění osob, které se nesečetly přímo na sčítacím formuláři – jediných „konstitutivním“ AZD zdrojem byl registr obyvatel.
- 2) Pro signs-of-life analýzu – identifikaci „známek života“ osob v jednotlivých registrech, které slouží pro posouzení, zda na našem území skutečně obvykle žily k rozhodnému okamžiku sčítání.
- 3) Pro doplnění zjišťovaných údajů tj. konkrétních proměnných, které se na sčítacích formuláři přímo nezjišťovaly (například občanství, rodinný stav), nebo které byly na sčítacím formuláři zjišťovány, ale část osob na ně neodpověděla.

Soustava registrů Ministerstva vnitra

Přímo v zákoně 332/2020 Sb. o sčítání lidu, domů a bytů v roce 2021 je zakotveno přebírání dat ze základního *registru obyvatel* (ROB) Ministerstva vnitra a jeho agendových informačních systémů – *agendového informačního systému evidence obyvatel* (AISEO), který eviduje české občany, a *agendového informačního systému cizinců* (AISC). Tato soustava registrů obsahuje klíčové proměnné o obyvatelstvu jako jsou například datum narození, pohlaví, rodinný stav, adresu registrovaného pobytu, občanství, typ a platnost pobytu, nebo rodinné vazby mezi osobami. Zároveň obsahuje identifikátory osob (agendový identifikátor fyzické osoby, rodné číslo, jméno a příjmení), které sloužily k napojování na sčítací formuláře, nebo na další AZD. Šlo o jediné AZD, které bylo tzv. „konstitutivní“ – na základě jeho záznamů bylo možno přidat osobu, která se nesečetla na sčítacím formuláři.



Ostatní administrativní zdroje dat

Dalším klíčovým registrem, který pokrývá téměř celé obyvatelstvo, je *centrální registr pojištěnců (CRP)* v gesci Všeobecné zdravotní pojišťovny ČR. Využití registru ve Sčítání 2021 bylo zásadní pro signs-of-life analýzu a vymezení obvykle bydlicího obyvatelstva. Z *integrovaného informačního systému sociálního zabezpečení (IISSZ)* vedeného Českou správou sociálního zabezpečení (ČSSZ) byly využity tři subsystémy: důchodů (starobní i invalidní důchodci), peněžité pomoci v mateřství (osoby pobírající tuto dávku) a pojistných vztahů (zaměstnané osoby). Z Ministerstva práce a sociálních věcí byly k dispozici ze systému *OK práce* uchazeči a zájemci o zaměstnání, ze systému *OK nouze* informace o osobách pobírajících dávku v hmotné nouzi a z *OK centrum* příjemci rodičovského příspěvku, jejich děti a případně druzí rodiče. Ministerstvo školství, mládeže a tělovýchovy poskytlo data ze *školských matrik* obsahující údaje o žácích základních škol, středních škol, konzervatoří a vyšších odborných škol ve školním roce 2020/2021. Jiným AZD ze stejného rezortu byly *sdílené informace matrik studentů*, které evidují data o vysokoškolských studiích od roku 1999. Dalším administrativním zdrojem dat bylo *přiznání k dani z příjmů fyzických osob*, respektive část údajů z tohoto přiznání, které poskytlo ČSÚ Generální finanční ředitelství. Posledním AZD využitým při posledním sčítání byl *registr ekonomických subjektů* spravovaný ČSÚ.

V návaznosti na Sčítání 2021 ČSÚ zřídil a spravuje nový informační systém veřejné správy – *cenový informační systém*, který je veden podle § 12a zákona č. 89/1995 Sb., o státní statistické službě. Tento systém slouží k uchování vybraných dat ze Sčítání 2021 (místo obvyklého pobytu v rozhodný okamžik, místo bydliště v době narození a nejvyšší dosažené vzdělání) a také pro účely splnění povinností stanovených přímo použitelnými předpisy Evropské unie v oblasti migrace a mezinárodní ochrany, evropské demografické statistiky a sčítání lidu, domů a bytů.

2.4.2.2 Administrativní zdroje dat za byty

V případě bytů byla většina údajů zjišťovaných při Sčítání 2021 odvozena přímo ze sčítacího formuláře, jelikož v Česku chybí plošný referenční registr bytů, který by mohl zcela pokrýt údaje o bytech požadované Eurostatem i domácími uživateli. ČSÚ nicméně vytěžil dostupné AZD s údaji za byty alespoň za účelem kompletace souboru a zpřesnění identifikace bytů pro podporu sběru dat – pro vytvoření seznamu bytů pro on-line sčítání a pro sčítací komisaře. V tomto směru byl využit zejména *informační systém katastru nemovitostí (ISKN)* spravovaný ČÚZK, který eviduje většinu bytů v osobním vlastnictví a v omezené míře i družstevní a obecní byty. Unikátní identifikátory a čísla bytových jednotek z ISKN byly v rámci územní přípravy zpracovány do statistického *registru sčítacích obvodů a budov (RSO)*, který eviduje všechny byty jako samostatné entity s vlastním číslováním a atributy, a proto byl použit jako referenční zdroj dat o bytech pro potřeby provedení sčítání. Prostřednictvím dálkového přístupu do katastru byla vytěžena i jména vlastníků jednotek jako doplňkový údaj pro identifikaci bytů v bytových domech. Identifikační údaje za byty nevidované v ISKN byly částečně došetřeny ve spolupráci s bytovými družstvy, obcemi a ostatními právníky osobami prostřednictvím webové aplikace pro sběr dat od vlastníků bytových domů.

2.4.2.3 Administrativní zdroje dat za domy

Na rozdíl od údajů za obyvatelstvo a byty, které byly zjišťovány převážně prostřednictvím dotazníku přímo od respondentů, byly údaje za domy až na drobné výjimky získávány z dostupných registrů a informačních systémů veřejné správy. Právě **zrušení formuláře s otázkami za dům** bylo jedním z inovativních opatření Sčítání 2021, které mělo za cíl snížit administrativní zátěž kladenou na občany v souvislosti s tímto povinným statistickým zjišťováním. Jednalo se tak o další kvalitativní posun oproti sčítání v roce 2011, při kterém byly AZD využity pouze k předvyplnění sčítacího formuláře za dům, který měl respondent povinnost ověřit a v případě zjištění nesprávného či chybějícího údaje doplnit správnou hodnotu.

Realizaci prvního administrativního sčítání domů umožnilo zejména **spuštění soustavy základních registrů veřejné správy od 1. 7. 2012**, konkrétně pak *registru územní identifikace, adres a nemovitostí (RÚIAN)* vedeného a spravovaného Českým úřadem zeměměřickým a katastrálním (ČÚZK). RÚIAN dle zákona č. 111/2009 Sb., o základních registrech, eviduje a kontinuálně aktualizuje referenční údaje popisného a lokalizačního charakteru územních a územně evidenčních prvků, stavebních objektů a jejich technickoekonomických atributů, adresních míst a ulic, a proto ho bylo možné využít jako referenční zdroj.

Klíčovou roli nejen při územní přípravě sčítání, ale i při zpracování výsledků za domy měl i *registr sčítacích obvodů a budov (RSO)* spadající do soustavy statistických registrů vedených Českým statistickým úřadem. RSO rovněž eviduje potřebné údaje za budovy, na rozdíl od evidence RÚIAN však až do úrovně jednotlivých vchodů, což plně odpovídá metodice sčítání. Výhodou RSO je i vícezdrojové plnění, díky čemuž vykazuje vyšší portfolio i naplněnost technickoekonomických atributů budov než RÚIAN. Vzhledem k přímému napojení RSO na RÚIAN ho tak bylo možné využít jako hlavní zdroj konstitučních, identifikačních, popisných a lokalizačních údajů o budovách, adresách a bytech pro vytvoření souboru referenčních dat územní přípravy (RDÚP) pro potřeby navazujících etap sčítání.

Oba tyto celoplošné zdroje byly využity v rámci zpracování výsledků za domy pro zjištění početnosti domovního fondu i většiny technickoekonomických atributů domů. Jako doplňkový či kontrolní zdroj pro odvození počtu bytů či vybavení domu výtahem byla využita například i *databáze dodacích míst České pošty*. Vlastnická struktura domů zase byla poprvé v historii sčítání plně odvozena z evidovaných dat o vlastnictví nemovitostí v *informačním systému katastru nemovitostí*.

2.5 Zpracování dat

Při zpracování dat ze Sčítání 2021 se (ve srovnání se SLDB 2011) v mnohem větší míře spoléhalo na údaje z administrativních zdrojů. Proměnné, které bylo možné získat za celou populaci z AZD, nebyly na formulářích zjišťovány. Větší množství dat z registrů pomohlo výrazně zdokonalit tzv. signs-of-life analýzu, jež je klíčová k vymezení obyvatelstva. Poprvé byly využity vztahy ze soustavy populačních registrů při odvozování hospodářích domácností, což v kombinaci s tím, že se na sčítacím formuláři zjišťovala celá matice vztahů mezi společně bydlícími osobami (ve sčítání 2011 pouze vztahy k osobě na prvním místě) vedlo ke zvýšení kvality dat za hospodářící domácnosti. Doplnění dat z AZD za proměnné dotazované na sčítacím formuláři zase pomohlo ke snížení podílu nezjištěných odpovědí (např. u počtu živě narozených dětí, nebo u dojížděky). Na druhou stranu vyšší míra využití AZD kladla větší nároky na přesnost propojení dat z různých zdrojů a jejich případnou deduplikaci a prioritizaci.

Výrazně vyšší využití elektronického sčítacího formuláře, resp. online sčítání, oproti sčítání 2011 znamenalo z pohledu zpracování menší objem prací v souvislosti s digitalizací listinných formulářů a kódováním textových zápisů. Vzhledem k existenci listinných formulářů však nebylo možné dotčené procesy zcela vypustit. Jednotný listinný formulář pro domácnost, který sdružoval informace o osobách, vztazích mezi nimi a obydlených bytech výrazně zjednodušil proces propojování jednotlivých entit mezi sebou. Datová burza tak již nemusela být použita v takovém rozsahu jako při předchozím sčítání. Neexistence samostatného bytového listu však neumožnila sběr podrobnějších dat o neobydlených bytech, přičemž odvození samotného počtu neobydlených bytů se ukázalo být jako problematické vzhledem k chybějící přesné evidenci bytů. Navíc protipandemická opatření znemožnila sledovat aspoň čísla bytů při distribuci sčítacích formulářů.

Na rozdíl od sčítání 2011 bylo zpracování dat za domy do značné míry odděleno od zpracování za entity osob, bytů a domácností. Důvodem bylo zejména zrušení listinného formuláře za domy, takže zpracování výsledků za domy probíhalo v první fázi nezávisle na datech ze sčítacích formulářů převážně s využitím dat z AZD a z územní přípravy. K propojení dat za domy s ostatními entitami došlo až v pozdější fázi zpracování po publikaci prvních výsledků za domovní fond, které tak byly následně ještě zpřesněny o zjištěné údaje o bytech a bydlení osob v terénu a prostřednictvím sčítacích formulářů (viz kapitola 14. Domovní fond).



2.5.1 Digitalizace, vstup dat z elektronických sčítacích formulářů

Prvotní zpracování vyplněných listinných sčítacích formulářů bylo stejně jako všechny ostatní procesy životního cyklu LSF – tisk, distribuce, sběr a logistika LSF vůbec – zajištěno dodavatelsky Českou poštou. ČSÚ přebíral od dodavatele jako výsledek zpracování LSF pouze údaje vytěžené z LSF v elektronické podobě. Veškeré práce spojené s fyzickými papírovými formuláři tedy na rozdíl od předchozích sčítání nebyly prováděny přímo v gesci ČSÚ. Kompletní digitalizace formulářů, včetně jejich zabezpečeného uložení v průběhu digitalizace a také následná skartace LSF byla řešena dodavatelem. ČSÚ ovšem s dodavatelem úzce spolupracoval především na návrzích a přípravách řešení, v průběhu digitalizace pak zajišťoval provádění několika druhů kontrol a přebírání výstupů z digitalizace do IT systémů ČSÚ.

Společný sčítací formulář v podobě sešitu, který byl klíčový pro zajištění sběru dat včetně potřebných propojení, představoval naopak pro digitalizaci určitou komplikaci, resp. větší nároky na přípravu formulářů před skenováním. Sešitý hřbet formuláře bylo nutné před vlastním skenováním oříznout, až poté jednotlivé oddělené listy formuláře procházely velkokapacitním skenováním. Naskenovány byly vždy všechny listy/strany formuláře, prázdné strany byly identifikovány a vyříděny až v rámci následných kroků digitalizace. Celkem bylo naskenováno 814,4 tis. LSF (všech 3 druhů LSF), což představovalo téměř 11,85 mil. stran.

Z naskenovaných obrazů formulářů byly pomocí SW nástrojů automatického rozpoznávání údaje převedeny do datového záznamu. Automaticky rozpoznané údaje byly podle nadefinovaných pravidel rovněž podrobeny manuální validaci prováděné školenými operátory. Kompletní výstup z digitalizace představoval vždy datový záznam údajů vytěžených z jedné strany formuláře a k němu připojený naskenovaný obraz strany. Digitalizovaná data byla v denních dávkách exportována dodavatelem do připraveného datového rozhraní, odkud je přebíral ČSÚ do svého interního vysoce zabezpečeného IT systému určeného pro uložení a zpracování neanonymních dat ze sčítání (Censový IS).

Každá strana formuláře obsahovala unikátní čárový a číselný kód, který umožňoval jednoznačnou identifikaci formuláře i dané strany. Přes tyto kódy bylo možné po provedené digitalizaci v databázi opět propojit části společného formuláře, či spíše datové záznamy za osoby, byt a domácnost, které k sobě ve vyplněném formuláři patřily.

Zatímco digitalizace LSF probíhala několik měsíců (od května do začátku října 2021), u dat z více než 4,2 mil. elektronických formulářů byl jejich vstup do zpracování výrazně jednodušší a rychlejší, včetně kontrol proběhl v řádu několika dnů až týdnů. Data sbíraná prostřednictvím online formulářů byla po odeslání formuláře ukládána do dočasné samostatné databáze ESF. Po skončení sběru dat byla veškerá získaná data dodavatelem ESF jednorázově převedena do Censového IS v prostředí ČSÚ. Data byla již na vstupu uložena v předem dohodnutých strukturách a formátech. Poté, co prošla kontrolami při převzetí údajů od dodavatele, tak byla bez větších zásahů připravena postoupit do kódování a dalších fází zpracování.

2.5.2 Kódování

Proces kódování se týkal těch údajů ve sčítacím formuláři, kde respondent vyplňoval adresu nebo textový, případně číselný zápis. Stejně jako při SLDB 2011 zahrnoval dvě části – kódování automatické a manuální. Pro každou z nich byly vyvinuty speciální SW aplikace, oba procesy ovšem využívaly stejné připravené slovníky pro kódování. Automatické kódování bylo prováděno v dávkách, vždy pro jeden druh údaje nad množinou formulářů připravených ve stavu pro kódování. Přípravu dat ke kódování a spouštění automatického kódování prováděli operátoři zpracování (z řad kmenových zaměstnanců oddělení sčítání). Jednalo se o velmi úzký okruh osob, které měly přístup k neanonymizovaným datům sčítání výhradně na zabezpečeném režimovém pracovišti.

Záznamy nezakódované automaticky přecházely do manuálního kódování. Aplikace pro manuální kódování zobrazovala pouze ty vytěžené údaje, případně výřezy z naskenovaných obrazů formulářů,

kteří byly k zakódování potřeba. Operátorům manuálního kódování tak nebyly přístupné identifikační údaje osob ani další údaje z formuláře.

Existovala i aplikace pro správu a editaci slovníků, v níž bylo možné slovníky dále doplňovat. Doplňovány byly zejména konkrétní zápisy vyskytující se ve větším počtu v již sebraných (a u LSF digitalizovaných) formulářích, na základě analýzy reálných dat ze sčítání. Nejvýznamnější doplnění slovníků probíhalo zpravidla před spuštěním automatického kódování, aby se tak zvýšila jeho úspěšnost, bohatší slovníky ale také významně usnadňovaly práci operátorům manuálního kódování při hledání odpovídajícího číselníkového záznamu a kódu.

Ve srovnání s rokem 2011 bylo kódovaných údajů celkově méně. Zatímco online formulář některé údaje, zejména adresy de facto kódoval již při vyplňování údajů, listinný formulář obsahoval polí ke kódování výrazně více. Úspěšnost automatického kódování také byla u záznamů z ESF zpravidla vyšší než u zápisů z LSF. Celkově se podařilo automaticky zakódovat přibližně 94 % údajů, více než 2 miliony údajů bylo nutné kódovat manuálně. Téměř tři pětiny tohoto počtu tvořily údaje o zaměstnání a odvětví, kde kvůli velké různorodosti zápisů zůstal objem manuálního kódování i přes dodatečné doplňování slovníků největší. Činnost pracoviště manuálního kódování proto byla nakonec prodloužena na cca 8 měsíců.

2.5.3 Zpracování dat za osoby, byty a domácnosti

Jednotný formulář pro domácnost sice urychlil zpracování sčítání, přesto bylo nutné provést v rámci listinných sčítacích formulářů konsolidaci a propojení údajů o osobách a seznamů osob v bytě (tzv. datovou burzu), následně pak byly údaje z LSF a ESF sdruženy do jednotné struktury.

Kromě dat ze sčítacích formulářů vstupovaly do zpracování sčítání i údaje z registru obyvatel (ROB) a jeho agendových informačních systémů AISEO a AISC (viz podkapitola 2.4.2 Sčítání prostřednictvím administrativních zdrojů dat). K tomuto účelu byla aktuální data z jednotlivých systémů propojena. V případě výskytu více záznamů za jednu osobu, byla data deduplikována a byl vybrán prioritní záznam. Data za osoby ze sčítacích formulářů byla následně propojena se záznamy o osobách ze soustavy ROB–AISEO/AISC. To umožnilo identifikaci a doplnění osob, které se nesečetly na sčítacích formulářích, odhalení multiplicitně sečtených osob a dotažení údajů z registrů. Do propojování vstupovaly kombinace identifikátorů osob (např. číslo dokladu, id datové schránky, rodné číslo, jméno, příjmení, datum narození) a na základě analýzy záznamů byla určena hierarchie pravidel, která kombinovala deterministické i pravděpodobnostní postupy. Výsledkem propojení bylo několik subpopulací:

- osob evidovaných v ROB nesečtených prostřednictvím formulářů, sečtených prostřednictvím jednoho nebo více formulářů
- osob neevidovaných v ROB sečtených prostřednictvím jednoho, nebo více formulářů, nebo formuláře, který neobsahoval relevantní identifikační údaje.

Následující kroky zpracování vedly k odstranění duplicit v datech a prioritizaci vybraného záznamu tak, aby za každou sečtenou osobu (ať už ze SF nebo AZD) existoval právě jeden záznam. Sjednocením množin osob ze sčítacích formulářů a z ROBu vznikl tzv. „základní kmen“, tj. referenční seznam unikátních záznamů všech osob, které vstupovaly do dalšího zpracování.

U multiplicit ze sčítacích formulářů odhalených po napojení na ROB byla následně provedena **deduplikace a prioritizace** vybraného záznamu. V případě osob sečtených na sčítacích formulářích a nenalezených v ROBu byly nejprve vyřazeny osoby s nesmyslnou nebo výrazně neúplnou identifikací, aby se zamezilo duplicitnímu sečtení osoby (jednou ze SF, podruhé z ROB). Poté byly nalezeny duplicitní osoby na sčítacích formulářích a na základě analýzy dat byla provedena deduplikace a prioritizace. Na záznamy o osobách z ROBu, které se nenapojily na žádnou osobu ze SF, se navíc napojovaly osoby ze seznamů osob v bytě, které se na SF nesečetly v sekci o osobách. Tento proces umožnil zařazení některých osob z ROBu do bytových domácností.



Kromě AZD ze soustavy ROB–AISEO/AISC byly pro zpracování sčítání využity i tzv. ostatní administrativní zdroje dat (OAZD). Jejich výčet je uveden v podkapitole 2.4.2 Sčítání prostřednictvím administrativních zdrojů dat. V této fázi zpracování byly propojeny na základní kmen osob podle specifických identifikátorů v OAZD obsažených. V některých případech byla propojení jednoznačná (například data od ČSSZ obsahovala agendový identifikátor fyzické osoby), někdy byla méně jasná (propojování podle kombinace jmen, příjmení a dat narození). Žádné OAZD nebylo konstitutivním zdrojem, které by do kmene osob přidávalo další osoby, ale byly zásadní pro potřeby signs-of-life analýzy a k doplnění údajů, které se na SF přímo nezjišťovaly (postavení v zaměstnání), nebo na které respondent neodpověděl (například vzdělání, nebo ekonomická aktivita).

Zpracování pokračovalo klíčovou fází, jejíž cílem bylo vymezení obvykle bydlícího obyvatelstva. **Obvykle bydlící obyvatelstvo** Česka byly vymezeno ze tří subpopulací:

1. Osoby sečtené na formuláři a evidované v ROBu s platným pobytem k rozhodnému okamžiku s výjimkou osob, které deklarovaly obvyklý pobyt v zahraničí.
2. Osoby sečtené na formuláři a nenalezené v ROBu, nebo s neplatným pobytem v ROBu musely mít kompletně vyplněné identifikační údaje, splňovat pravidla pro zařazení do obvykle bydlícího obyvatelstva a zároveň deklarovat první pobyt po narození v zahraničí. Měli to být především cizinci, kteří na našem území pobývají, ale nejsou přihlášení k pobytu. Občanství bylo dostupné ale pouze z ROBu, takže jako náhrada byl použit první pobyt po narození.
3. Osoby evidované v ROBu s platným pobytem k rozhodnému okamžiku, ale nesečtené na formulářích.

Vybrat z této skupiny osoby, které na našem území skutečně žijí, bylo cílem signs-of-life analýzy. Populační registry obvykle obsahují záznamy osob, které na daném území již nežijí – buď se odstěhovaly, nebo zemřely. Signs-of-life analýza využívala různé administrativní zdroje dat k vyhodnocení příznaků života osoby na našem území. Pravidla byla nastavena specificky pro různé subpopulace podle věku i občanství (občan ČR vs. cizinec).

Konkrétní adresu obvyklého pobytu deklarovaly osoby na sčítacím formuláři a ta nemusela být totožná s adresou registrovaného pobytu. V případě osob nesečtených na formuláři byla za adresu obvyklého pobytu považována adresa registrovaného pobytu z ROBu.

Obyvatelé s registrovaným pobytem jsou ve Sčítání 2021 všichni obvykle bydlící s výjimkou osob sečtených na formuláři a nenalezených v ROBu, nebo s neplatným pobytem v ROBu. Jinými slovy musely osoby s registrovaným pobytem mít platný pobyt k rozhodnému okamžiku a buď se sečíst na formuláři, nebo být vyhodnoceny signs-of-life analýzou s pozitivními známkami života na našem území.

Při odvozování jednotlivých proměnných za **osoby** byly aplikovány specifické kontroly přípustnosti a logické kontroly. Cílem následující fáze zpracování bylo prioritizovat zdroje jednotlivých proměnných. V případě některých proměnných byly AZD jediným zdrojem (např. státní občanství), někdy byly k dispozici údaje ze sčítacího formuláře i AZD. Ve většině případů byly prioritizovány informace ze sčítacího formuláře. Data z AZD byla použita spíše v případech, kdy údaj ze sčítacího formuláře nebyl vyplněn, nebo byl v průběhu kontrol zneplatněn. Podrobnosti lze nalézt v jednotlivých analytických kapitolách. Níže je uvedena tabulka proměnných ze Sčítání 2021 spolu s dostupnými zdroji dat.

Tab. 2.1 Proměnné za osoby ze SLDB 2021 podle zdroje dat

Proměnná	Sčítací formuláře	Administrativní zdroje dat
pohlaví	X	X
věk	X	X
rodinný stav		X
občanství		X
místo obvyklého pobytu	X	X
místo bydliště jeden rok před sčítáním	X	X
místo bydliště v době narození	X	X
místo registrovaného pobytu		X
nejvyšší dosažené vzdělání	X	X
počet živě narozených dětí	X	X
mateřský jazyk	X	
národnost	X	
náboženská víra	X	
způsob bydlení	X	X
ekonomická aktivita	X	X
postavení v zaměstnání		X
odvětví ekonomické činnosti	X	X
zaměstnání	X	
místo pracoviště/školy	X	X
frekvence dojížděky	X	
dopravní prostředek dojížděky	X	

Vstupem do zpracování dat za **byty** byla data za bydlení propojená se seznamem osob v bytech ze sčítacích formulářů a seznam bytů z referenčních dat územní přípravy Sčítání 2021 (RDÚP) aktualizovaný během informační pochůzky sčítacími komisaři. I v případě bytů bylo nutné provést proces deduplikace a prioritizace. K identifikaci multiplicit sloužil zejména seznam osob v bytě (bytová domácnost) na stejné adrese. Identifikátor bytu nemohl být využit díky neexistující evidenci distribuce LSF na úroveň bytu (v době distribuce LSF byla v platnosti přísná protipandemická opatření, která to znemožňovala) a velké chybovosti identifikace bytu z ESF. Po odstranění duplicitních záznamů došlo k propojení se seznamem bytů z RDÚP a obohacení o nově zjištěné obydlené byty ze sčítacích formulářů. Na základě propojení osob ze základního kmene a bytů byla určena obydlenost bytu. Byty, které nebyly obydleny žádnou osobou, byly po splnění několika podmínek určeny jako neobydlené byty.

Pro odvození **hospodařících domácností** byly využity vztahy mezi všemi společně bydlícími osobami a informace o společném hospodaření těchto osob, to vše ze sčítacích formulářů. Ze soustavy registrů ROB–AISEO/AISC byly k dispozici klíčové rodinné vztahy (manželé/registrovaní partneři, rodiče/děti a po odvození prarodiče/vnuci). Podle ne(existence) rodinných jader byly následně identifikovány jednotlivé typy hospodařících domácností (např. úplná rodina tvořená manželským párem, neúplná rodina osamělého otce s dítětem, vícečlenná nerodinná domácnost, domácnost jednotlivce apod.). Ve výsledcích sčítání jsou pro hospodařící domácnosti tvořené 1 rodinou odvozeny počty závislých dětí. Závislé dítě je každá osoba v domácnosti tvořené 1 rodinou, jejíž postavení v domácnosti je „dítě“, není zahrnuta mezi pracovní sílu a zároveň je ve věku 0–25 let.

2.5.4 Zpracování dat za domy

Pro odvození **počtu domů** byl využit export všech budov z referenčních dat územní přípravy (RDÚP), který byl následně po prověření aktualizován o potvrzené nové či neexistující budovy identifikované sčítacími komisaři v rámci informační pochůzky a o dodatečné přírůstky nových budov do RSO po

předání souboru RDÚP. Následně proběhlo podchycení duplicitních záznamů a výběr unikátních záznamů k rozhodování o jejich zařazení do domovního fondu. Vzhledem k vymezení domovního fondu při Sčítání 2021, zahrnujícího pouze domy určené k bydlení, tj. budovy, v nichž je alespoň jeden byt nebo jiné prostory určené k bydlení nebo dlouhodobějšímu ubytování, jako jsou různá ubytovací zařízení, vstupovaly do podmínky pro zařazení budovy do domovního fondu také atributy počet bytů, druh domu a u některých i obydlenost domu. Na závěr byly provedeny ještě nezbytné logické kontroly a porovnání vývoje počtu domů mezi sčítáními 2011 a 2021 se statistikou nové bytové výstavby v tomto období.

Zařazení budovy do domovního fondu rozhodovalo o zpracování dalších charakteristik za tyto domy – obydlenosti domu a druhu domu – které jsou dostupné pro všechny budovy zahrnuté do domovního fondu. Všechny další zjišťované atributy domů byly v souladu s předchozími sčítáními odvozeny a zveřejněny jen za množinu obydlených domů. Mezi atributy, které byly zpracovány výhradně z administrativních zdrojů, patří druh domu a typ vlastníka domu. Technickoekonomické atributy materiál nosných zdí domu, způsob odvádění odpadních vod a vybavenost domu výtahem byly převážně odvozeny z údajů evidovaných v administrativních zdrojích, částečně ale byly došetřeny i v rámci územní přípravy buď stavebními úřady nebo spolupracujícími vlastníky bytových domů či terénními pracovníky sčítání. U zbývajících charakteristik domů (připojení domu na vodovod, plyn, způsob vytápění domu, počet nadzemních podlaží) byly výsledné hodnoty odvozeny z významově odpovídajících údajů zjištěných ze sčítacích formulářů za byty v kombinaci s evidovanými hodnotami v administrativních zdrojích (využity zejména v případě nezjištěných odpovědí za byty).

2.6 Diseminace výstupů

Závěrečnou a do jisté míry tou nejviditelnější fází projektu Sčítání 2021 byla diseminace výsledků, která je v širším pojetí chápána jako šíření získaných a zpracovaných statistických informací všemi možnými dostupnými způsoby. Strategie diseminace sčítání byla předem nastavena a schválena v souladu s celkovou diseminační politikou Českého statistického úřadu, která vychází z ustanovení *Zákona č. 89/1995 Sb., o státní statistické službě*.

V průběhu všech činností souvisejících s diseminací výsledků byly uplatňovány veškeré principy, kterými jsou mimo jiné profesionální nezávislost, objektivita a transparentnost nebo přesnost a spolehlivost. Ve všech aspektech byl kladen důraz na naplnění potřeb širokého spektra uživatelů a dodržení jejich práva na svobodný přístup k informacím. V neposlední řadě byl respektován a zajištěn relevantní princip pro ochranu důvěrných dat.

Diseminace dat úzce souvisí s obsahem sčítání. Základním východiskem pro Sčítání 2021 bylo postupné zveřejňování definitivních výsledků. Na rozdíl od posledního sčítání o deset let dříve tak nebyly publikovány předběžné výsledky. Zveřejněná data, zpracovaná v tradičním rozsahu tabulek v různém územním detailu, plně umožňovala porovnání s výsledky minulého sčítání za obvykle bydlící obyvatelstvo. Kromě některých dílčích obsahových změn (viz Metodické vysvětlivky) byly pro Sčítání 2021 charakteristické zejména rozsáhlejší možnosti prezentace dat.

Celosvětovým trendem moderní společnosti je zvyšování dostupnosti a srozumitelnosti poskytovaných statistických informací nejen směrem k odborné, ale také k široké veřejnosti. Způsoby zveřejňování výsledků ze Sčítání 2021 v elektronické podobě proto maximálně využívaly nová technologická řešení a nabízely flexibilní a dynamické nástroje pro zobrazování.

První představení definitivních výsledků proběhlo na tiskové konferenci v lednu 2022. Od té doby – v návaznosti na průběh zpracování - probíhalo kontinuální zveřejňování dat za obyvatelstvo, domovní a bytový fond, hospodařící domácnosti a v neposlední řadě i za vyjížděku a dojížděku do zaměstnání a školy. Důležitým milníkem diseminace bylo splnění závazků vůči Statistickému úřadu Evropské unie (Eurostat), který požadoval včasné dodání mezinárodně srovnatelných dat za stanovená témata a ukazatele – ve formě takzvaných hyperkostek. Jednou z novinek tohoto sčítání, vycházejících z potřeb všech členských států Evropské unie, byla diseminace výsledků v takzvaných kilometrových

čtvercích. To umožňovalo zveřejnit údaje v prakticky libovolném územním členění, bez ohledu na administrativní celky nebo národní hranice.

Hlavním rozcestníkem ke všem informacím o Sčítání 2021 se stala webová stránka na adrese www.scitani.cz. Zde, v rubrice s Výsledky, měl uživatel možnost dostat se jednoduše ke zveřejněným datům. Tabulkové výstupy byly doplněny doprovodnými komentáři pro snadnější a rychlejší porozumění jednotlivým výstupům. Vizualizace pomocí kartogramů umožňovala zase lepší představu o územních rozdílech u konkrétního jevu. Významnou roli při prezentaci výsledků hrály i další grafické vizualizace, které vhodně doplnily tabulkové či textové výstupy. Obsah webových stránek byl kompletně k dispozici nejen v českém, ale i anglickém jazyce.

Pro diseminaci výsledků v podrobnějším územním členění byla, stejně jako pro sčítání v roce 2011, využívána Veřejná databáze ČSÚ (zkráceně VDB). Ke zveřejňovaným tématům a jejich ukazatelům byly předem spočítány stanovené agregace věcných dimenzí pro různé územní úrovně. Nad agregovanými daty byly poté vytvořeny tabulky umožňující v některých případech interaktivní výběr parametrů, například podle věku či pohlaví. Výhodou databáze byl mimo jiné také metodický popis zveřejněných údajů a dostupné metainformace. Přístup byl podporován v českém i v anglickém jazyce. Pro odborné uživatele byly postupně zveřejňovány výsledky také ve formě otevřených dat.

V rámci přípravy jednotlivých výstupů pro výše popsané diseminační kanály se zároveň připravovaly datové sady ze sčítání v podobě, ve které budou uživatelům k dispozici prostřednictvím nové aplikace DataStat. Projekt je zaměřen na uživatelem definované výstupy; kromě zobrazení předdefinovaných datových výstupů má umožnit především sestavení vlastních výsledných tabulek právě podle požadavků uživatele. Podobné aplikace jsou využívány v řadě zemí nejen pro výsledky censů, ale i pro další statistické oblasti.

Stejně jako v minulosti byly připravovány a postupně zveřejňovány také datové a analytické publikace. Výstupy ve formě publikací jsou vhodné nejen pro zkušené uživatele s dobrou orientací ve zveřejňovaných výsledcích, tedy se znalostí publikovaných tabulek, ale i pro ty uživatele, kteří se zatím v datových výstupech tolik neorientují. Všechny publikace byly zveřejněny v elektronické podobě přes Katalog produktů ČSÚ (např. Základní výsledky Sčítání lidu, domů a bytů 2021 nebo krajské analýzy výsledků sčítání). Tištěné výtisky v omezeném rozsahu byly využívány především v rámci Odboru informačních služeb, protože umožňují předložit uživateli o datech ze sčítání souhrnný přehled.

Zcela nový způsob prezentace výsledků ze sčítání nabídl Statistický geoportál. Výsledná data v něm byla geoprostorově vizualizována s využitím sofistikovaných nástrojů geografického informačního systému (GIS). Koncové webové aplikace - Statistický atlas, Dojíždka, Statistické georeporty - pak umožnily uživatelům zobrazení vybraných ukazatelů v jednotlivých tématech za určité území v přehledných interaktivních mapách.

Doplňkovou možnost zpřístupnění dat ze sčítání, především těch nejzajímavějších, představovaly údaje průběžně prezentované prostřednictvím sociálních sítí (Facebook, sociální síť X - dříve Twitter, YouTube, LinkedIn nebo Instagram) nebo přes mobilní aplikaci ČSÚ. Takovéto formy prezentace výsledků byly praktikovány vůbec poprvé v historii sčítání.

Popsané diseminační kanály představovaly základní nebo doplňkové způsoby zveřejňování výsledků Sčítání 2021 a byly navrženy a vytvořeny tak, aby různým skupinám uživatelů umožnily co nejkomfortnější a nejrychlejší cestu k požadovaným informacím a výsledkům. V průběhu celé diseminace byl kladen důraz na srozumitelnost a metodickou srovnatelnost výstupů v čase i na moderní grafickou formu.

2.7 Hodnocení kvality

Nezbytnou součástí prací na tak složitém a nákladném projektu jakým sčítání lidu je, jsou kontrolní procesy a vyhodnocování jeho kvality. Český statistický úřad se zaměřuje na dvě dimenze kvality:



faktickou (věcnou) a statistickou kvalitu. Hodnocení věcné kvality je založeno na kvantitativní a kvalitativní analýze kontrolních zpráv. Audity zkoumají všechny procesy a subjekty podílející se na realizaci sčítání. Statistické hodnocení kvality je založeno na kvantitativní analýze získaných dat.

2.7.1 Řízení kvality sčítání

Řízení kvality bylo zaměřeno na všechny procesy, které se podílely jak na přípravě sčítání (jak terénních prací, tak i podkladů pro pochůzky sčítacích komisařů, tisk listinných formulářů a dalších materiálů), tak jeho realizaci – distribuci a sběru listinných formulářů, kontrola digitalizace dat, kontroly na základě reportů předávaných z informačních systémů České pošty (např. objektů označených za neexistující či neobydlené, kontrola terénního došetření).

Kontroly kvality byly prováděny při zkušebním sčítání v druhé polovině roku 2020 i během reálného sčítání v první polovině roku 2021. Kontroly kvality při zkušebním sčítání pomohly zpřesnit odhad náročnosti jednotlivých prací, především ale odhalily, na které procesy je třeba zaměřit pozornost, aby sebraná data měla co nejvyšší vypovídací hodnotu. Průběžné vyhodnocování kontrol mělo za cíl objevit slabá místa jednotlivých procesů a možnost včas reagovat na případné problémy a zajistit jejich nápravu.

Podkladem pro kontroly kvality byly Metodické listy kontrol, které přesně popsaly místo kontroly, kontrolní období, předmět kontroly, kontrolní a procesní postup, požadovanou součinnost s kontrolovaným objektem, způsob výběru kontrolovaného objektu a velikost kontrolního vzorku, standardní dobu trvání kontroly, limitní hodnoty. Na ně navazovaly pracovní postupy, které kontrolorům ČSÚ podrobně a po jednotlivých po sobě následujících krocích určily průběh a obsah kontroly. O každé provedené kontrole byl veden záznam.

Ve zkušebním sčítání bylo provedeno celkem 635 kontrol na kontaktních a sběrných místech České pošty (dále „ČP“), dále 131 kontrol sběru listinných sčítacích formulářů (dále „LSF“) ve vybraných sčítacích obvodech, 135 kontrol vyplnění LSF, kontroly rozvozu LSF na sběrná místa, kontroly ověřeného a doplněného popisu a mapy standardních sčítacích obvodů – objektů. Kvůli šíření onemocnění Covid-19 bylo mnoho kontrol zrušeno – jak na jednotlivých sběrných místech ČP, tak i úplné zrušení kontroly ověřeného a doplněného seznamu bytů (aby kontroloři ČSÚ nevcházeli do uzavřených prostor v domech).

V období reálného sčítání proběhlo celkem 2719 kontrol kontaktních míst, a 62 kontrol sběrných míst (kontrola dokumentace, monitoring činností, bezpečnost nakládání s listinnými formuláři). Dále proběhlo 469 kontrol pochůzek (distribuce LSF).

Na základě náhodně vybraného vzorku 258 standardních sčítacích obvodů s více než 12,2 tisíci náhodně vybranými objekty proběhla kontrola popisu a map standardních sčítacích obvodů za objekty. Kontrolu prováděli krajsí pracovníci SLDB. Pracovní postup byl rozdělen do dvou fází - přípravné a terénní. V přípravné fázi proběhla na pracovišti ČSÚ prvotní kontrola souboru budov s využitím dostupných online zdrojů (např. Registr územní identifikace, adres a nemovitostí, Registr sčítacích obvodů, Katastr nemovitostí apod.). Jejím účelem bylo identifikovat rozdíly v obdržení datech z informační pochůzky s evidovanými administrativními daty a na tyto rozdíly se následně zaměřit při kontrole v terénu. Následně kontroloři ČSÚ provedli fyzickou kontrolu evidovaných hodnot atributů vybraných budov zjištěním skutečného stavu v terénu, jelikož při posuzování správnosti údaje byl vždy rozhodující zjištěný stav v terénu, kterým se řídil i sčítací komisař při informační pochůzce.

Dalšími důležitými kontrolami byly kontroly informační pochůzky. V období průběhu informační pochůzky (od 20. 3. 2021 do 14. 4. 2021) byly primárně kontrolovány objekty, které sčítací komisaři ČP označili jako neexistující či neobydlené. Důvodem této kontroly byl fakt, že neobydlené a neexistující objekty byly následně vyřazeny z terénního došetření a neprobíhala zde distribuce listinných sčítacích formulářů. Cílem této kontroly tedy bylo minimalizovat případy chybného vyřazení objektů z terénního došetření. Celkem bylo takto zkontrolováno 12 sčítacích obvodů s vysokou koncentrací neexistujících objektů, 6 sčítacích obvodů s vysokou koncentrací neobydlených objektů, 178 neexistujících bytových

domů a 189 neobydlených bytových domů. Po konci informační pochůzky bylo provedeno vyhodnocení podkladů od krajských pracovníků a objekty, u kterých byl zjištěn rozpor mezi zjištěním sčítacího komisaře a příslušných krajských pracovníků, byly reklamovány České poště, tzn., byla požadována změna stavu objektu.

Kontrola terénního došetření byla realizována obdobně jako kontrola informační pochůzky. První z kontrolovaných oblastí byl počet distribuovaných formulářů sčítacími komisaři na úrovni jednotlivých sčítacích obvodů. Cílem této kontroly bylo nalezení oblastí s potenciálním problémem v kvalitě distribuce listinných formulářů sčítacími komisaři. Tento ukazatel byl sledován průběžně a byl také komunikován s Českou poštou.

2.7.2 Kvalita dat

Důležitým aspektem kvality dat ze sčítání je podíl zjištěných údajů za danou proměnnou. Vysoká míra non-response vede k tomu, že jsou data méně důvěryhodná a jejich vypovídající schopnost může být zkrácená. Ani využití pouze relativních údajů bez nezjištěno nemusí pomoci. Není totiž jisté, zda by subpopulace, za kterou data nemáme, měla relativně stejné odpovědi jako subpopulace se zjištěnou odpovědí.

Tab. 2.2 Podíl zjištěných údajů u proměnných za osoby^{*)} a zdroj dat, SLDB 2021 (v %)

Proměnná	Podíl zjištěných odpovědí	Zdroj dat
pohlaví	100,0	SF / AZD
věk	100,0	SF / AZD
rodinný stav	99,5	AZD
občanství	99,7	AZD
místo obvyklého pobytu	100,0	SF / AZD
místo bydliště jeden rok před sčítáním	99,6	SF / AZD
místo bydliště v době narození	100,0	SF / AZD
místo registrovaného pobytu	100,0	AZD
druh registrovaného pobytu	100,0	AZD
nejvyšší dosažené vzdělání	94,2	SF / AZD
počet živě narozených dětí	97,9	SF / AZD
mateřský jazyk	92,8	SF
národnost	68,4	SF
náboženská víra	69,9	SF
způsob bydlení	100,0	SF / AZD
ekonomická aktivita	98,8	SF / AZD
postavení v zaměstnání	90,7	AZD
odvětví ekonomické činnosti	99,2	SF / AZD
zaměstnání	87,0	SF
místo pracoviště/školy	97,4	SF / AZD
frekvence dojížděky	85,6	SF
dopravní prostředek dojížděky	87,0	SF

^{*)} Pouze za osoby, které měly na danou otázku odpovídat.

Nižší podíl zjištěných údajů byl evidován u proměnných, jejichž zdrojem byl pouze sčítací formulář (viz tab. 2.2) – zejména u dobrovolných otázek na národnost (68,4 %) a náboženskou víru (69,9 %). Naopak proměnné vycházející z kombinace zdrojů, nebo jejichž zdrojem dat byla pouze AZD, dosahovaly vyšší úrovně zjištěnosti (výjimkou je pouze postavení v zaměstnání s 90,7 %). Místo obvyklého bydliště bylo zjištěno u všech osob s obvyklým pobytem, avšak u osob, kde jsme neměly

informaci o obvyklému pobytu ze sčítacího formuláře bylo ztotožněno s místem registrovaného pobytu. Všechny osoby s registrovaným pobytem mají zjištěné místo tohoto pobytu. Plnou vyplněnost měly i proměnné pohlaví, věk a způsob bydlení. Místo bydliště v době narození a druh registrovaného pobytu sice nebyly plně zjištěné, ale po zaokrouhlení se hodnota dostala na 100,0 %. Ve srovnání se sčítáním 2011 se zvýšila úroveň zjištěnosti zejména na ekonomické charakteristiky (ekonomická aktivita, odvětví ekonomické činnosti, místo pracoviště/školy) nebo u proměnné počet živě narozených dětí. Důvodem bylo zejména zapojení administrativních zdrojů dat.

Ženy měly podíl zjištěných údajů vyšší než muži téměř u všech proměnných (viz tab. 2.3 a 2.4). Jedinou výjimkou byl dopravní prostředek, kde měly míru zjištěnosti o 0,4 p. b. nižší. Naopak proměnnou s nejvyšším rozdílem ve prospěch žen bylo zaměstnání – o 5 p. b. více než muži.

Tab. 2.3 Podíl zjištěných údajů u proměnných za osoby^{*)}, muži podle věkových skupin, (v %)

Proměnná	Věková skupina					
	Celkem	0–14	15–29	30–49	50–64	65+
pohlaví	100,0	100,0	100,0	100,0	100,0	100,0
věk	100,0	100,0	100,0	100,0	100,0	100,0
rodinný stav	99,5	100,0	99,1	99,4	99,6	99,4
občanství	99,7	99,8	99,4	99,6	99,8	99,9
místo obvyklého pobytu	100,0	100,0	100,0	100,0	100,0	100,0
místo bydliště jeden rok před sčítáním	99,5	99,8	99,3	99,4	99,6	99,8
místo bydliště v době narození	100,0	100,0	100,0	100,0	100,0	100,0
místo registrovaného pobytu	100,0	100,0	100,0	100,0	100,0	100,0
druh registrovaného pobytu	100,0	100,0	100,0	100,0	100,0	100,0
nejvyšší dosažené vzdělání	93,1	x	94,1	91,8	92,4	95,1
počet živě narozených dětí	x	x	x	x	x	x
mateřský jazyk	91,6	87,6	91,3	91,1	92,7	95,6
národnost	67,1	61,7	62,1	65,3	71,3	75,6
náboženská víra	69,2	62,9	69,7	69,9	70,2	72,4
způsob bydlení	100,0	100,0	100,0	100,0	100,0	100,0
ekonomická aktivita	98,5	99,9	98,2	97,5	98,1	99,8
postavení v zaměstnání	89,7	x	83,9	90,8	93,0	80,9
odvětví ekonomické činnosti	99,1	x	98,7	99,2	99,3	98,1
zaměstnání	84,7	x	84,4	86,6	85,3	65,1
místo pracoviště/školy	97,3	98,2	96,4	97,7	97,7	93,6
frekvence dojížděky	84,9	57,4	85,9	93,6	93,6	75,9
dopravní prostředek dojížděky	87,2	57,6	87,5	95,0	94,9	80,3

^{*)} Pouze za osoby, které měly na danou otázku odpovídat.

Dalším diferencujícím faktorem, který ovlivňoval podíl zjištěných údajů, byl věk osoby. Nelze ovšem identifikovat jednoznačný směr vlivu napříč většinou otázek. Z administrativních zdrojů dat bylo možné v případě některých proměnných doplnit údaj pouze za určitou věkovou subpopulaci, což ovlivňuje výsledné hodnoty. Například vyšší podíl zjištěných počtů živě narozených dětí lze sledovat spíše u mladších matek. Také data o nejvyšším dosaženém vzdělání byly z AZD doplňovány zejména za mladší osoby. Jednoznačnou vazbu mezi věkem a podílem zjištěných údajů tak lze pozorovat pouze u sociokulturních charakteristik zjišťovaných jenom na sčítacím formuláři – mateřského jazyka, národnosti a náboženské víry. S rostoucím věkem rostl podíl zjištěných údajů a to u obou pohlaví.

Tab. 2.4 Podíl zjištěných údajů u proměnných za osoby^{*)}, ženy podle věkových skupin, (v %)

Proměnná	Věková skupina					
	Celkem	0–14	15–29	30–49	50–64	65+
pohlaví	100,0	100,0	100,0	100,0	100,0	100,0
věk	100,0	100,0	100,0	100,0	100,0	100,0
rodinný stav	99,6	100,0	99,1	99,6	99,8	99,5
občanství	99,8	99,8	99,3	99,7	99,9	100,0
místo obvyklého pobytu	100,0	100,0	100,0	100,0	100,0	100,0
místo bydliště jeden rok před sčítáním	99,7	99,8	99,6	99,7	99,8	99,8
místo bydliště v době narození	100,0	100,0	100,0	100,0	100,0	100,0
místo registrovaného pobytu	100,0	100,0	100,0	100,0	100,0	100,0
druh registrovaného pobytu	100,0	100,0	100,0	100,0	100,0	100,0
nejvyšší dosažené vzdělání	95,3	x	95,9	95,0	95,3	95,3
počet živě narozených dětí	97,9	x	99,0	98,8	98,1	95,8
mateřský jazyk	93,9	87,4	92,9	94,5	95,9	96,4
národnost	69,7	61,6	63,2	69,4	74,8	75,5
náboženská víra	70,7	63,0	71,0	72,2	72,0	72,6
způsob bydlení	100,0	100,0	100,0	100,0	100,0	100,0
ekonomická aktivita	99,0	99,9	97,9	98,1	99,3	99,9
postavení v zaměstnání	91,9	x	82,4	94,5	95,2	74,8
odvětví ekonomické činnosti	99,3	x	99,0	99,5	99,5	98,1
zaměstnání	89,8	x	87,3	91,8	91,2	67,7
místo pracoviště/školy	97,5	98,3	96,5	98,1	97,8	91,3
frekvence dojížděky	86,2	57,1	87,0	95,5	95,2	75,7
dopravní prostředek dojížděky	86,8	57,4	87,5	95,8	95,5	77,4

^{*)} Pouze za osoby, které měly na danou otázku odpovídat.

2.8 Rozpočet sčítání

Návrh rozpočtu sčítání byl od počátku součástí projednávání věcného záměru zákona o sčítání lidu, domů a bytů v roce 2021, který schválila vláda 19. června 2018. Upravený návrh rozpočtu byl součástí vládního návrhu zákona o sčítání lidu, domů a bytů v roce 2021, drobné úpravy spočívaly zejména ve změnách rozložení prostředků do jednotlivých let a snížení objemu na výdaje ostatních ústředních orgánů o 1,1 mil. Kč. Předpokládané výdaje na Sčítání 2021 tak měly činit celkem 2 225,1 mil. Kč, což zhruba odpovídalo nominální úrovni výdajů na sčítání v roce 2011.

Schválené výdaje z rozpočtové kapitoly 345 – Český statistický úřad byly v letech 2018 až 2023 ve výši **2 156,8 mil. Kč**. Výdaje ostatních ústředních orgánů byly plánovány ve výši 35,3 mil. Kč a výdaje obcí prostřednictvím účelové dotace z rozpočtové kapitoly Všeobecná pokladní správa ve výši 33,0 mil. Kč.

Tab. 2.5 Schválený rozpočet ČSÚ v letech 2018 až 2023 (v mil. Kč)

Druh výdaje	2018	2019	2020	2021	2022	2023	Celkem
Běžné výdaje celkem	45,0	293,0	626,6	953,6	29,6	6,0	1 953,8
v tom							
osobní výdaje	17,0	82,7	158,3	152,2	13,2	0,0	423,4
věcné výdaje	28,0	210,3	468,3	801,4	16,4	6,0	1 530,4
Kapitálové výdaje	0,0	56,0	132,0	15,0	0,0	0,0	203,0
Výdaje celkem	45,0	349,0	758,6	968,6	29,6	6,0	2 156,8



První prostředky schváleného rozpočtu účelově určené na provedení Sčítání 2021 byly uvolněny ve druhém pololetí roku 2018. Veškeré potřebné činnosti na přípravě sčítání realizované do této doby byly kryty z běžného rozpočtu ČSÚ. Výše a struktura přidělených prostředků na přípravu, provedení, zpracování a diseminaci výsledků sčítání byla každoročně zpřesňována při přípravě státního rozpočtu na následující rok ve spolupráci s Ministerstvem financí ČR.

Největší část výdajů Českého statistického úřadu na realizaci Sčítání 2021 tvořily výdaje:

- na úhradu smluvní odměny za provedení terénních prací partnerem projektu Českou poštou, s. p., která zajišťovala zejména tyto činnosti:
 - o tisk a rozvoz listinných sčítacích formulářů a dalších tiskovin,
 - o informační pochůzku,
 - o distribuci a sběr listinných sčítacích formulářů,
 - o digitalizaci, skladování a skartaci listinných formulářů,
- na vytvoření, implementaci a provoz 3 páteřních IT systémů sčítání – systému distribuce a sběru online elektronických formulářů (ONLINE), systému pro bezpečné uložení a zpracování neanonymních dat (Censový IS) a systému monitoringu a vyhodnocení průběhu fáze distribuce, sběru a zpracování formulářů (CMS SLDB),
- na vysoce zabezpečenou infrastrukturu pro zřízení, provoz a správu systému pro realizaci online sběru dat a systému pro uložení, zpracování a anonymizování dat získaných v rámci sčítání prostřednictvím služeb Státní pokladny Centra sdílených služeb, s. p.,
- na územní přípravu sčítání vč. realizace zkušebního sčítání,
- na zřízení a provoz dočasných pracovišť,
- na platy a ostatní osobní výdaje dočasných pracovníků,
- na komunikační a informační kampaň, provoz kontaktního centra,
- na zpracování výsledků, tj. výdaje na zřízení a provoz dočasného režimového pracoviště pro kódování, pořízení vstupů, kontrol a zpracování výsledků,
- na vytvoření, implementaci a provoz 3 IT systémů diseminace výsledků – Geoportálu pro diseminaci geografických dat SLDB 2021 (GIS), Webové aplikace pro uživatelem definované výstupy dat SLDB 2021 (UDV) a Portálu pro prezentaci výsledků SLDB 2021 (PDI),
- na služby projektového managementu, služby technické kanceláře, bezpečnostní testování systémů.

Tab. 2.6 Výdaje ČSÚ na sčítání v letech 2018 až 2023 (v tis. Kč)

Druh výdaje	2018	2019	2020	2021	2022	2023	Celkem
Běžné výdaje celkem	3 137,2	130 453,8	438 040,5	1 000 616,6	74 248,4	16 878,6	1 663 375,1
v tom							
osobní výdaje	1 549,8	62 657,5	146 110,8	132 237,8	15 775,4	3 667,5	361 998,8
věcné výdaje	1 587,4	67 796,3	291 929,7	868 378,8	58 473,0	13 211,1	1 301 376,3
Kapitálové výdaje	0,0	10 395,3	74 649,2	33 699,2	40 448,8	4 155,3	163 347,8
Výdaje celkem	3 137,2	140 849,1	512 689,7	1 034 315,8	114 697,2	21 033,9	1 826 722,9

Čerpání výdajů na projekt nebylo v době vydání publikace ukončeno, bude pokračovat do doby vypořádání závazků ze smluv uzavřených z titulu SLDB 2021, zejména na dokončení IT projektů diseminace a zajištění jejich technické podpory.

Přesto lze konstatovat, že došlo k výrazným úsporám oproti předpokládaným výdajům. **Nejvyšší úspora** byla dosažena ve **výdajích České pošty**, s. p., a to zejména v důsledku epidemické situace v souvislosti s onemocněním Covid-19, a s tím souvisejícími mimořádnými opatřeními, která vedla ke zmenšení rozsahu terénních prací. V této souvislosti došlo dodatkem smlouvy ke změnám postupů při realizaci terénních prací spočívajícím v tom, že důraz bude kladen na online sběr dat a činnost sčítacích komisařů při terénním došetření bude v součinnosti ČSÚ a České pošty modifikována tak, aby došlo k výraznému omezení osobních kontaktů mezi sčítacími komisaři a obyvatelstvem. Úspora z těchto

změn byla vyčíslena na **93,4 mil. Kč**. Dále je potřeba mezi úspory zahrnout i zjištěné nadměrné vyrovnání v rámci kontroly a ověření výše vyrovnávací platby poskytnuté v rámci projektu Sčítání 2021 České pošty ve výši 132 mil. Kč. K vrácení nadměrné platby došlo zatím ve výši **63,9 mil. Kč**.

Ve srovnání s předchozím sčítáním v roce 2011 budou konečné výdaje na Sčítání 2021 nižší o více než 400 mil. Kč.

Graf 2.1 Porovnání výdajů SLDB 2011 a 2021 v jednotlivých letech projektu

