

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

1/2022

ročník/volume 32

Recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov.

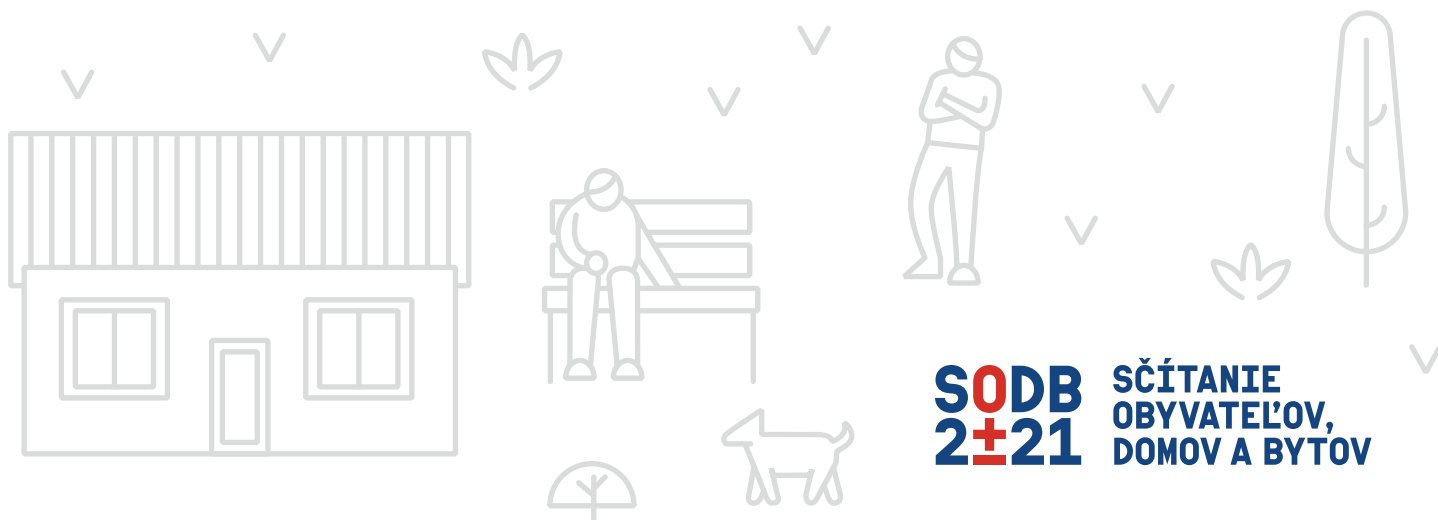
Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 4

Typ článku/Type of article: vedecký článok/scientific article

Strany/Pages: 41 – 54

Dátum vydania/Publication date: 21. marec 2022/March 21, 2022



SODB SČÍTANIE
2±21 OBYVATEĽOV,
DOMOV A BYTOV

Martin KOČIŠ, Cindy ERDELYI
Štatistický úrad Slovenskej republiky

TVORBA INTEGRAČNÉHO ZOZNAMU RODNÝCH ČÍSEL NA ÚČEL AUTENTIFIKÁCIE OBYVATEĽA V ELEKTRONICKOM SČÍTANÍ

CREATION OF AN INTEGRATION LIST OF IDENTIFICATION NUMBERS FOR AUTHENTICATION PURPOSES OF THE ELECTRONIC POPULATION CENSUS

ABSTRAKT

Článok sa zaoberá tvorbou integračného zoznamu rodných čísel na autentifikáciu obyvateľa v elektronickom sčítaní. Príspevok poskytuje detailný pohľad na postup tvorby integračného zoznamu rodných čísel, opisuje jeho využitie v aplikácií na elektronické sčítanie obyvateľov, predovšetkým na autentifikáciu. Následne opisuje možné spôsoby autentifikácie do elektronického sčítacieho formulára.

ABSTRACT

The article deals with the creation of the Integration list of identification numbers for the purpose of population authentication in the electronic census. The article provides a detailed view of the process of creating the Integration list of identification numbers, describes its use in applications for the electronic census, especially for the purpose of authentication. It describes subsequently the possible methods of authentication to the electronic census form.

KLÚČOVÉ SLOVÁ

integračný zoznam rodných čísel, autentifikácia, sčítací formulár, administratívne zdroje údajov

KEY WORDS

integration list of identification numbers, authentication, census form, administrative data sources

1. ÚVOD

Sčítanie obyvateľov, domov a bytov je jedným z najvýznamnejších štatistických zisťovaní, do ktorého majú zákonnú povinnosť zapojiť sa všetci obyvatelia Slovenskej republiky. Toto zisťovanie prinesie nenahraditeľné informácie o stave spoločnosti využiteľné ako v štátnej a verejnej, tak aj v súkromnej sfére, a to pre laickú i odbornú verejnosť. Náročná koncepčná príprava a realizácia sčítania obyvateľov, domov a bytov sa v Štatistickom úrade SR (ďalej aj „ŠÚ SR“) začala v decembri 2014 a až po realizáciu v roku 2021 sa skladala z množstva prípravných činností. Medzi jednu z finálnych prípravných aktivít patrila aj tvorba tzv. integračného zoznamu rodných čísel.

Integračný zoznam rodných čísel (ďalej len „IZRČ“) je databáza všetkých rodných čísel obyvateľov, ktoré boli získané a integrované z administratívnych zdrojov údajov (ďalej len „AZÚ“) použitých na účely sčítania obyvateľov. Okrem rodných čísel IZRČ obsahuje dáta pre potreby autentifikácie obyvateľa v elektronickom sčítaní, vyplnenia trvalého pobytu obyvateľa v sčítacom formulári a priebežného monitoringu sčítania vo webovej aplikácii. IZRČ bol z hľadiska jeho rozsahu vytvorený tak, aby obsahoval

množinu všetkých validných rodných čísel zo všetkých administratívnych zdrojov údajov. Pre každé rodné číslo existuje v tabuľke IZRČ však iba jeden záznam. Hlavným účelom tvorby IZRČ bolo umožniť autentifikáciu obyvateľa v elektronickom sčítacom formulári. Autentifikáciou sa rozumie proces kontroly a overenia identity obyvateľa na základe rodného čísla a následne doplnkového údaj (meno obyvateľa alebo obec pobytu). Po úspešnej autentifikácii bol obyvateľovi umožnený vstup do sčítacieho formulára.

Ďalším podstatným dôvodom na tvorbu IZRČ bolo vytvorenie databázového podkladu na monitoring stavu sčítania obyvateľov. V tomto prípade sa IZRČ použil ako podklad na aplikáciu vytvorenú na účely sledovania priebežného stavu sčítaných obyvateľov k odhadovanému počtu obyvateľov.

Potreba tvorby IZRČ spočívala aj v zámere dosiahnuť jeden z hlavných aspektov sčítania a to zníženie administratívnej záťaže na obyvateľa. Ak autentifikácia prostredníctvom zadania rodného čísla a doplnkového údaj v elektronickom sčítacom formulári prebehla úspešne, znamená to, že obyvateľa sa v databáze IZRČ podarilo cez jeho rodné číslo a doplnkový údaj identifikovať a spárovať. Vďaka úspešnej autentifikácii bol následne obyvateľovi umožnený vstup do tzv. krátkeho sčítacieho formulára, ktorý obsahoval 14 otázok. Pri spracovaní budú tieto informácie doplnené z administratívnych zdrojov údajov. Okrem zníženia administratívnej záťaže na obyvateľov IZRČ poskytuje adresnejšiu formu sčítacieho formulára. Keďže IZRČ obsahoval dáta, ako je napríklad vek alebo pohlavie, otázky, príp. odpovede neaplikovateľné napr. na osoby mladšie ako 18 rokov sa obyvateľovi nezobrazili, príp. si z nich nemohol vybrať. Tento aspekt zaistil aj zvýšenie kvality vyplnených údajov.

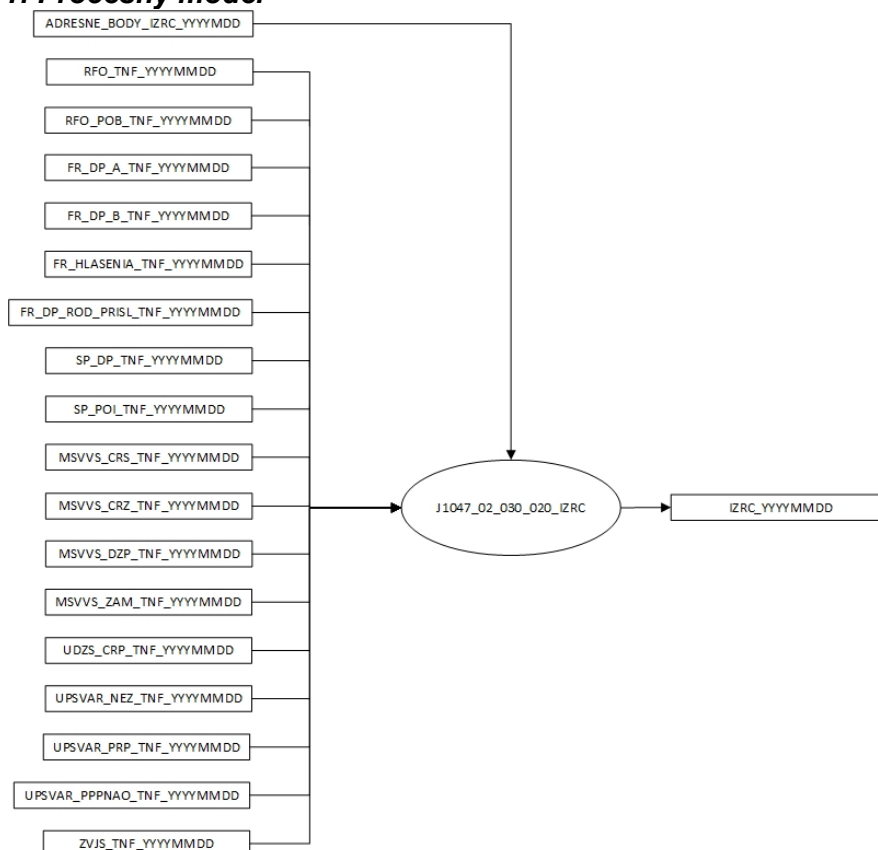
2. ZOZNAM A IMPORT AZÚ NA TVORBU IZRČ

Základným predpokladom na vytvorenie kvalitných dátových podkladov v rôznych etapách SODB 2021 vrátane tvorby IZRČ bolo mať k dispozícii v čo najvyššej možnej miere kvalitné dátové zdroje. Vytvorenie IZRČ na účely SODB 2021 je naviazané na využitie AZÚ, ktoré sú definované v zákone č. 223/2019 Z. z. o sčítaní obyvateľov, domov a bytov v roku 2021 podľa § 20. Základ dátových zdrojov teda tvoria AZÚ, ktoré dodali exportom z informačných systémov (ďalej len „IS“) ich správcovia. Z IS Ministerstva vnútra SR pre register fyzických osôb bol dodaný Register fyzických osôb (ďalej aj „RFO“), ktorý obsahoval 10 dátových tabuliek. Na účely IZRČ boli z RFO potrebné dátové tabuľky, a to tabuľka s názvom RFO_OEX obsahujúca údaje o osobe, tabuľka RFO_POB obsahujúca údaje o pobyte a tabuľka RFO_MOS obsahujúca evidované meno osoby. Z IS Finančného riaditeľstva SR boli dodané a spracované tabuľky daňových priznaní k dani z príjmov fyzických osôb typu A (FR_DP_A), daňové priznania k dani z príjmov fyzických osôb typu B (FR_DP_B) a hlásenia o vyúčtovaní o úhrne príjmov zo závislej činnosti (FR_HLASENIA). Podmnožinou dátového zdroja IS finančného riaditeľstva je tabuľka FR_DP_ROD_PRISL, ktorá obsahuje rodinných príslušníkov osôb evidovaných v tabuľkách FR_DP_A a FR_DP_B. Ministerstvo školstva, vedy, výskumu a športu dodalo export tabuliek z IS pre vysoké školstvo – centrálny register študentov vysokých škôl (MSVVAS_CRIS) a centrálny register zamestnancov vysokých škôl (MSVVAS_CRZ), a export tabuliek z IS pre regionálne školstvo – centrálny register detí, žiakov a poslucháčov (MSVVAS_DZP) a centrálny register pedagogických zamestnancov, odborných zamestnancov a ďalších zamestnancov škôl a školských zariadení (MSVVAS_ZAM). Z IS Sociálnej poisťovne sa využili dátové tabuľky zoznamu poistencov (SP_POI) a zoznamu poberateľov

dávok dôchodkového zabezpečenia (SP_DP). Medzi ďalšie dátové zdroje patrí centrálny register poistencov (UDZS_CRP), ktorý bol dodaný exportom z IS Úradu pre dohľad nad zdravotnou starostlivosťou. Tabuľky o evidovaných nezamestnaných (ÚPSVAR_NEZ), poberateľoch rodičovského príspevku (ÚPSVAR_PRP) a poberateľoch peňažného príspevku na opatrovanie (UPSVAR_PPNAO) boli dodané exportom z IS služieb zamestnanosti a IS riadenia sociálnych dávok od Ústredia práce, sociálnych vecí a rodiny. Posledným využitým zdrojom bol zoznam obyvateľov vo výkone väzby a výkone trestu odňatia slobody (ZVJS) z IS Zboru väzenskej a justičnej stráže.

Všetky zdroje sa ešte pred použitím načítali do extrakčnej vrstvy, v ktorej sú dáta načítané 1:1, tzn. bez vykonania akýchkoľvek zmien. Zmeny v dátach, napríklad ich prvotné spracovanie (transformácie číselníkových hodnôt na hodnoty z číselníkov využiteľných na SODB, deduplikácia záznamov, logické kontroly a pod.) sa vykonali v transformačnej vrstve (TNF). Prvotne spracované dáta z transformačnej vrstvy tvorili základ pre IZRČ. Znáznomený procesný model (obrázok č. 1) poukazuje na jednotlivé dátové vstupy použité na vytvorenie IZRČ. Názvy jednotlivých dátových vstupov sú odvodené od skratky AZÚ a dátumu platnosti AZÚ.

Obrázok č. 1: Procesný model



Zdroj: SAS, ŠÚ SR (vlastné spracovanie)

Vzhľadom na termíny dodania jednotlivých AZÚ od ich správcov na účely SODB sa pre IZRČ k referenčnému dátumu sčítania, t. j. k 31.12.2020 využili AZÚ RFO, MSVVS_CRS, MSVVS_CRZ, MSVVS_DZP, MSVVS_ZAM a ZVJS. Ostatné AZÚ boli platné k dátumu 31.12.2019, čo sa zohľadnilo pri jeho tvorbe vo vzťahu k nastaveniu

priorizácie jednotlivých AZÚ. Priorizácia zohľadňovala obsahovú kvalitu a aktuálnosť jednotlivých AZÚ (bližšie ju opíšeme v nasledujúcej kapitole). Napriek tomu, že nie všetky AZÚ boli platné k rovnakému dátumu, tieto AZÚ sa v čo najvyššej možnej miere použili ako doplňujúce AZÚ na odvodenie chýbajúcich premenných alebo doplnenie rodných čísel do IZRČ.

Okrem dátových vstupov z AZÚ vstupovala do IZRČ aj tabuľka tzv. adresných bodov (ADRESNE_BODY_IZRC_YYYYMMDD), prebraná z územnej prípravy na SODB 2021. Ide o množinu všetkých adresných bodov s lokalizáciou až na x a y súradnice, ktoré vstupovali do procesu sčítania domov a bytov 2021.

Jednotlivé dátové zdroje po prvotnom spracovaní a naviazaní na adresné body vytvorili celkovú množinu rodných čísel a im prislúchajúcich údajov v IZRČ. Pre každé rodné číslo existoval v tabuľke IZRČ len jeden záznam. IZRČ bol vytvorený množinou všetkých iba validných rodných čísel zo všetkých AZÚ. Spracovanie rodného čísla, vyhodnotenie jeho validity a odvodenie ďalších údajov pre potreby autentifikácie obyvateľa v IZRČ opíšeme v nasledujúcej kapitole.

3. SPRACOVANIE ÚDAJOV PRE POTREBY VYTVORENIA IZRČ

Prvotné spracovanie údajov pre IZRČ prebiehalo vo viacerých fázach. Do IZRČ nevstupovali všetky údaje z AZÚ, ale len vybrané premenné, ktoré slúžili prioritne na autentifikáciu obyvateľa alebo na priebežný monitoring stavu sčítania. Štruktúru celého IZRČ zobrazuje tabuľka č. 1.

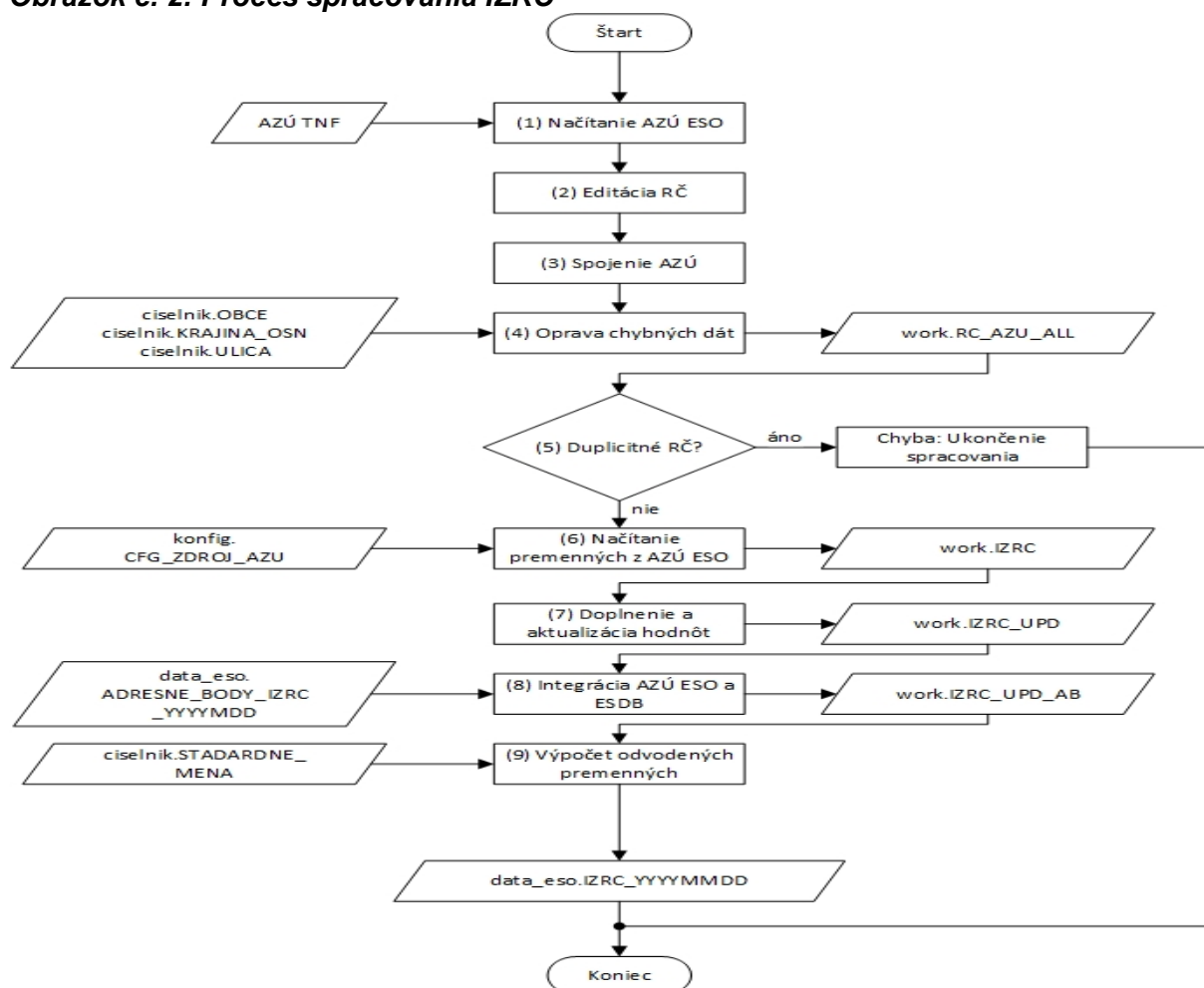
Tabuľka č. 1: Štruktúra IZRČ

Premenná	Popis	Premenná	Popis
RODNE_CISLO	Rodné číslo	ORIENTACNE_CISLO	Orientačné číslo
ZDROJ_RC	Výskyt RČ v AZÚ	STAT_KOD	Štát - kód
RC_EDIT_FLG	Príznak editovaného RČ	STAT_NAZOV	Štát - názov
ID_OEX	Identifikátor osoby v RFO	INDEX_DOMU	Index domu pobytu v RFO
TYP_OSOBY_KOD	Typ osoby – kód	TYP_POBYTU_KOD	Typ pobytu - kód
MENO	Meno	DATUM_PRIHLASENIA_POBYTU	Dátum prihlásenia na pobyt
MENO_SRC	Krstné meno – zdroj	ADRESA_SRC	Adresa - zdroj
MENO_NORM	Krstné meno bez diakritiky	TYP_DOMU_KOD	Typ domu - kód
DATUM_NARODENIA	Dátum narodenia	SURADNICA_LNG	Zemepisná šírka
DATUM_NARODENIA_SRC	Dátum narodenia - zdroj	SURADNICA_LAT	Zemepisná dĺžka
POHLAVIE_KOD	Pohlavie – kód	DUPLICITNE_CISLA_BYTOV_FLG	Príznak duplicitných bytov
POHLAVIE_SRC	Pohlavie – zdroj	MEDZIPOSCHODIA_FLG	Príznak medziposchodia
POHLAVIE_KOD_AZU	Pohlavie - kód AZÚ	ADRESNY_BOD_ID	ID adresného bodu
DATUM_UMRTIA	Dátum úmrtia	ALG_ADRESNY_BOD_IZRC	Algoritmus mapovania adresného bodu do IZRČ
DATUM_UMRTIA_SRC	Dátum úmrtia - zdroj	CENTROID_FLG	Príznak centroidu
OBEC_KOD	Obec – kód	RC_RFO_FLG	Príznak RČ z RFO
OBEC_NAZOV	Obec - názov	RC_VALID_FLG	Príznak validného RČ
CAST_OBCE_KOD	Časť obce - kód	VALID_MENO_FLG	Príznak štandardného krstného mena
CAST_OBCE_NAZOV	Časť obce - názov	VALID_OBEC_FLG	Príznak validnej obce
ULICA_KOD	Ulica – kód	VYBER_DOPLNKOVEHO_UDAJA_KOD	Kód na výber doplnkového údajja
ULICA_NAZOV	Názov ulice	EXPORT_FLG	Príznak exportu
SUPISNE_CISLO	Súpisné číslo		
SUPISNE_CISLO	Súpisné číslo		

Zdroj: ŠÚ SR, vlastné spracovanie

V prvej fáze spracovania údajov bolo potrebné vykonať transformácie premenných vstupujúcich do IZRČ a deduplikáciu rodného čísla vo všetkých AZÚ v transformačnej vrstve (TNF). Do transformačnej vrstvy sa údaje načítali z extrakčnej vrstvy. Transformáciami sa myslia základné úpravy, ktoré sa vykonali na úrovni jednotlivých tabuliek AZÚ v transformačnej vrstve. Bolo potrebné zjednotiť formát premenných pre pohlavie a taktiež pre všetky lokalizačné charakteristiky pobytových premenných na jednotné číselníkové hodnoty, nastavenie rovnakého dátumového formátu pre všetky dátumové premenné a doplnenie chýbajúcich hodnôt pri premenných. Doplnenie sa realizovalo v prípadoch, keď bol napríklad vyplnený kód obce trvalého pobytu, ale chýbal názov obce – v takom prípade sa odvádzal prislúchajúci názov. Takto sa dopĺňalo všade tam, kde to bolo možné. V rámci transformačnej vrstvy patrilo k základným úpravám aj spracovanie rodného čísla, pri ktorom prebehli základné kontroly a deduplikácia. K deduplikácii sa pristúpilo iba pri validných rodných číslach. Rodné číslo bolo vyhodnotené ako validné ak bola jeho dĺžka 9 alebo 10 znakov (bez lomky) a obsahovalo iba číslice. Deduplikácia predstavovala výber platného záznamu za rodné číslo, ak bola na úrovni rodného čísla v AZÚ duplicita. Pri určení výberu platného záznamu sa uprednostňoval napríklad záznam podľa typu alebo aktuálnosti pobytu pre záznamy z RFO, pre záznamy iných AZÚ (napríklad ÚZDS_CRP, SP_POI) sa vyberal záznam s platným a najaktuálnejším poisťným vzťahom a pod. Celý procesný model spracovania IZRČ je zobrazený na obrázku č. 2.

Obrázok č. 2: Proces spracovania IZRČ



Zdroj: SAS, ŠÚ SR (vlastné spracovanie)

Proces spracovania IZRČ sa začína načítaním AZÚ (v procesnom modeli označené ako fáza 1) z transformačnej vrstvy, po aplikovaní základných transformácií a deduplikácii. Z transformačnej vrstvy sa načítajú všetky rodné čísla zo všetkých AZÚ spolu so základnými premennými, medzi ktoré patria:

Tabuľka č. 2: Štruktúra základných premenných

Premenná	Popis
RODNE_CISLO	Rodné číslo
ID_OEX	Identifikátor osoby v RFO
TYP_OSOBY_KOD	Typ osoby – kód
MENO	Meno
DATUM_NARODENIA	Dátum narodenia
POHLAVIE_KOD	Pohlavie – kód
DATUM_UMRTIA	Dátum úmrtia
OBEC_KOD	Obec – kód
OBEC_NAZOV	Obec – názov
CAST_OBCE_KOD	Časť obce – kód
CAST_OBCE_NAZOV	Časť obce – názov
ULICA_NAZOV	Názov ulice
SUPISNE_CISLO	Súpisné číslo
ORIENTACNE_CISLO	Orientačné číslo
STAT_KOD	Štát – kód
STAT_NAZOV	Štát – názov
INDEX_DOMU	Index domu pobytu v RFO
TYP_POBYTU_KOD	Typ pobytu – kód
DATUM_PRIHLASENIA_POBYTU	Dátum prihlásenia na pobyt

Zdroj: ŠÚ SR, vlastné spracovanie

Po načítaní AZÚ sa v procese spracovania IZRČ začala editácia rodného čísla (v procesnom modeli označené ako fáza 2), ktorá sa realizovala podľa prislúchajúceho dátumu narodenia. K tomuto kroku sa pristúpilo z dôvodu, že v niektorých údajoch bola nekonzistentnosť medzi rodným číslom a dátumom narodenia osoby a zároveň pri rodnom čísle neplatila kontrola na modulo 11¹. Pre overenie správnosti, či je nekonzistentnosť spôsobená chybou v rodnom čísle alebo v dátume narodenia, sa rodné číslo vykonštruovalo na novo podľa dátumu narodenia a pohlavia, pričom obidva záznamy sa v IZRČ ponechali. Editácia rodného čísla sa realizovala iba pre rodné čísla s dátumom narodenia 1954 a viac, aby bolo možné overenie cez modulo 11.

Odvedenie sa realizovalo takto:

- prvé dve číslice (predstavujú rok narodenia) v novovytvorenom rodnom čísle boli odvodené z posledných dvoch číslic z roku narodenia osoby. Tretia a štvrtá číslica (predstavuje mesiac narodenia) boli odvodené z mesiaca narodenia osoby, pričom ak pohlavie v danom zázname bolo uvedené 2 (žena), k hodnote mesiaca sa pripočítalo 50. Ak bola hodnota mesiaca 1 – 9 (tzn. január až september) a pohlavie bolo 1 (muž), pridala sa číslica 0. Piata a šiesta číslica (predstavuje deň narodenia) bola odvodená z dňa narodenia

¹ Kontrola modulo 11 – celé desaťmiestne rodné číslo musí byť bezo zvyšku deliteľné číslom 11.

osoby, pričom ak bola hodnota dňa 1 – 9, pridala sa pred dané číslo 0. Posledné štvorčísle sa prebralo z pôvodného rodného čísla bez vykonania úprav.

Ostatné hodnoty premenných zodpovedajúce rodnému číslu v IZRČ ostali nezmenené a prevzali sa zo záznamu rodného čísla pred editáciou. Nový záznam s editovaným rodným číslom bol do IZRČ pridaný iba v prípade, ak editované rodné číslo spĺňalo kontrolu na modulo 11 a zároveň v celom IZRČ neexistovalo rodné číslo s rovnakou hodnotou.

Ďalším z procesov spracovania údajov na vytvorenie IZRČ bolo spojenie AZÚ (v procesnom modeli označené ako fáza 3). Tento proces spočíval v prepojení všetkých rodných čísel so základnými premennými z jednotlivých AZÚ z transformačnej vrstvy a ich spojení do jednej spoločnej tabuľky. Takto sa vytvorila množina komplexných údajov prislúchajúcich k rodným číslam, ktoré sú nevyhnutné pre potreby IZRČ a na následnú autentifikáciu obyvateľa a sledovanie monitoringu sčítania.

Po procese spojenia AZÚ sa znova pristúpilo ku kontrolám a opravám chybných dát (v procesnom modeli označené ako fáza 4), ktoré sa nerealizovali na úrovni transformačnej vrstvy. Kontroly a opravy sa týkali prevažne dát o adrese pobytu. V tejto fáze sa detailnejšie analyzovali rôzne neštandardné zápisy v adrese využitím aliasingu ulíc² alebo dopĺňanie resp. oprava kódov a názvov obcí a štátov napríklad pre duplicitné obce v rámci SR alebo aj sveta. Po aplikovaní spomenutých opráv sa začal proces vyhodnotenia duplicitných rodných čísel (v procesnom modeli označené ako fáza 5). Táto kontrola sa realizovala preto, že ak by bol identifikovaný duplicitný záznam v IZRČ, celý proces by sa bol zastavil a spracovanie by sa bolo ukončilo. Ako sme spomenuli, IZRČ nemôže obsahovať duplicitu na úrovni rodného čísla. Ak sa v IZRČ nenachádzali duplicitné rodné čísla, pristúpilo sa k načítaniu premenných z AZÚ (v procesnom modeli označené ako fáza 6). Týmto krokom sa uskutočnil výber a načítanie finálnych premenných k jednotlivým rodným číslam so zohľadnením prioritizácie jednotlivých AZÚ. Pri pobytových premenných sa na úrovni výberu správneho záznamu zohľadňovala prioritizácia takým spôsobom, aby bol stále vybratý iba jeden pobyt pre každú osobu na úrovni rodného čísla. K tomuto kroku sa pristúpilo z dôvodu, že za jedno rodné číslo sa mohlo vyskytovať viacero záznamov o pobyte (napríklad súbeh trvalých pobytov, resp. prechodných pobytov). Podľa zákona o sčítaní sa sčítaním zisťovali údaje o obyvateľoch Slovenskej republiky, pričom obyvateľ je štátny občan Slovenskej republiky s trvalým pobytom alebo prechodným pobytom na území Slovenskej republiky, občan EÚ s obvyklým bydliskom na území SR alebo štátny príslušník tretej krajiny s trvalým pobytom, prechodným pobytom alebo tolerovaným pobytom na území SR okrem cudzinca používajúceho diplomatické výsady a imunitu na území SR podľa medzinárodného práva [1]. Preto bol základom výberu platného pobytu v SR v prioritizácii trvalý pobyt v SR → prechodný pobyt v SR → tolerovaný pobyt v SR → trvalý pobyt v zahraničí → prechodný pobyt v zahraničí → tolerovaný pobyt v zahraničí → trvalý pobyt bez obce → prechodný pobyt bez obce → tolerovaný pobyt bez obce. V Slovenskej republike sa za referenčný register považuje register fyzických osôb (RFO), ktorý je základným registrom verejnej správy, obsahujúcim súbor údajov o fyzických osobách. Údaje o fyzických osobách vedené v registri fyzických osôb sa považujú za úplné a zodpovedajúce skutočnosti, kým nie je

² *Aliasingom ulíc je štandardizácia rôznych zápisov a tvarov ulíc pod jeden spoločný názov a ich priradenie k oficiálnemu, a teda správne názvu.*

preukázaný opak. Proti osobe, ktorá sa v dobrej viere spolieha na údaje uvedené v registri fyzických osôb, nemôže iná osoba namietat', že tieto údaje nie sú úplné alebo nezodpovedajú skutočnosti. Výber lokalizačných premenných pre pobyt v IZRČ zachytáva tabuľka č. 3.

Tabuľka č. 3: Postupnosť výberu pobytových premenných

I)	Pobyt z RFO s úplnou adresou, kde je vyplnený kód obce, názov ulice, súpisné číslo a orientačné číslo
II)	Pobyt z RFO s adresou, kde je vyplnený kód obce a súpisné číslo alebo orientačné číslo
III)	Pobyt z iného AZÚ s adresou, kde názov alebo kód obce sa zhoduje s obcou v RFO a je vyplnený kód obce, názov ulice, súpisné číslo a orientačné číslo
IV)	Pobyt z iného AZÚ s adresou, kde názov alebo kód obce sa zhoduje s obcou v RFO a je vyplnený kód obce a súpisné číslo alebo orientačné číslo
V)	Pobyt z RFO s neúplnou adresou
VI)	Pobyt z iného AZÚ s úplnou adresou, kde je vyplnený kód obce, názov ulice, súpisné číslo a orientačné číslo
VII)	Pobyt z iného AZÚ s adresou, kde je vyplnený kód obce a súpisné číslo alebo orientačné číslo
VIII)	Pobyt z iného AZÚ s neúplnou adresou

Zdroj: SAS, ŠÚ SR

V prípade, keď informácie o pobyte neboli dostupné v RFO, pre IZRČ sa načítal pobyt s najvyššou početnosťou v AZÚ podľa kódu obce, súpisného a orientačného čísla. Ak početnosť výskytu hodnôt v AZÚ bola rovnaká, načítaval sa pobyt z AZÚ podľa prioritizácie jednotlivých AZÚ, pričom v prioritizácii bola zohľadnená aktuálnosť AZÚ a ich kvalita. Najvyššiu prioritu mal aktuálny a referenčný register RFO. Nasledovali rovnako aktuálne (tzn. platné k 31.12.2020) AZÚ Ministerstva školstva, vedy výskumu a športu v postupnosti MSVVS_CRS, MSVVS_CRZ, MSVVS_DZP a MSVVS_ZAM a AZÚ Zboru väzenskej a justičnej stráže (ZVJS). Ďalšie AZÚ neboli aktuálne (platné k 31.12.2019), preto sa zohľadnila ich dátová kvalita z hľadiska obsahu, kompletnosti a úplnosti, a to v poradí Centrálny register poistencov (UDZS_CRP), AZÚ Sociálnej poisťovne SP_DP a SP_POI, ďalej AZÚ Ústredia práce sociálnych vecí a rodiny UPSVAR_NEZ, UPSVAR_PRP, UPSVAR_PPPNAO a nakoniec AZÚ Finančného riaditeľstva FR_DP_A, FR_DP_B, FR_HLASENIA a FR_DP_ROD_PRISL.

Pri výbere ostatných finálnych premenných pre IZRČ (meno, dátum narodenia, dátum úmrtia, pohlavie) sa postupovalo obdobne ako pri pobytových premenných. Aj tu sme vychádzali zo skutočnosti, že RFO je referenčný a zároveň aktuálny register, preto mal pri výbere najvyššiu prioritu. Ak nebola v RFO dostupná hodnota napríklad pohlavia, načítala sa hodnota najčastejšie sa vyskytujúca v AZÚ. Ak bola početnosť vyskytujúcich sa hodnôt rovnaká, výber sa uskutočnil na základe prioritizácie AZÚ, ktorá bola nastavená rovnako v prípade pobytových premenných.

Po výbere finálnych premenných nastúpil proces spracovania, v ktorom sa doplnili a aktualizovali hodnoty (v procesnom modeli označené ako fáza 7). Ide o proces kontroly úplnosti údajov po načítaní údajov z AZÚ do IZRČ. Tento proces sa vykonal v prípade, ak ani po transformáciách a kontrolách na úrovni transformačnej vrstvy a následne po výbere finálnych premenných pre konkrétne rodné číslo neboli úplne doplnené všetky údaje. Pre nedostupný dátum narodenia alebo starší dátum

narodenia ako 1.1.1800 sa dopĺňal dátum narodenia z rodného čísla: ak malo dĺžku 10 znakov, platila kontrola na modulo 11 a odvodený dátum narodenia nebol po referenčnom dátume a zároveň bol väčší ako 1.1.1800. V niektorých prípadoch sa dátum narodenia odvádzal aj pre 10-miestne rodné čísla s koncovkou 0000, ktorá sa v určitých prípadoch dopĺňa pre cudzincov. Pre rodné čísla s dĺžkou 9 znakov sa odvodil dátum narodenia pre všetky chýbajúce záznamy, ak nebol po referenčnom dátume a zároveň väčší ako 1.1.1800. Pre rodné čísla s chýbajúcim dátumom narodenia sa deň narodenia odvádzal z 5. a 6. číslice v rodnom čísle. Mesiac narodenia bol odvodený z 3. a 4. číslice rodného čísla, ak však bola hodnota na 3. a 4. číslici väčšia ako 50 a menšia ako 63, potom sa od tejto hodnoty odpočítalo 50. Pri roku narodenia sa brala do úvahy aj dĺžka rodného čísla. Ak bola dĺžka 9 znakov a hodnota prvých dvoch číslic bola väčšia ako 54 alebo sa rovnala 54, potom bol rok narodenia hodnota 1800 + hodnota prvých dvoch číslic. Ak dĺžka rodného čísla bola 9 číslic a hodnota prvých dvoch číslic bola menšia ako 54, potom bol rok narodenia hodnota 1900 + hodnota prvých dvoch číslic. Pri dĺžke rodného čísla 10 znakov a hodnote prvých dvoch číslic 54 alebo viac bol rok narodenia hodnota 1900 + hodnota prvých dvoch číslic. Ak pri 10-miestnom rodnom čísle bola hodnota prvých dvoch číslic menšia ako 54, potom bol rok narodenia hodnota 2000 + hodnota prvých dvoch číslic.

Rodné číslo sa v procese doplnenia a aktualizácii hodnôt využívalo aj na odvodenie pohlavia. Ak pohlavie nebolo dostupné, tak sa odvádzalo z rodného čísla. Odvádzanie pohlavia sa realizovalo, ak rodné číslo bolo 10-miestne, platila kontrola na modulo 11 a odvodený dátum narodenia nebol po referenčnom dátume alebo starší ako 1.1.1800, prípadne pre 10-miestne rodné čísla bola koncovka 0000, ale odvodený dátum narodenia nebol po referenčnom dátume alebo starší ako 1.1.1800. Pri 9-miestnych rodných číslach sa pohlavie odvádzalo, ak opäť platilo, že odvodený dátum narodenia nebol po referenčnom dátume alebo starší ako 1.1.1800. Pre odvodenie platilo, že ak na 3. mieste rodného čísla je číslica 5 alebo 6, potom bolo pohlavie 2 (žena), inak pohlavie 1 (muž). V prípadoch, keď pohlavie nebolo možné odvodiť ani z rodného čísla, odvodilo sa z krstného mena osoby pomocou nástroja SAS DQ funkcie DQGENDER.

Nevyhnutným krokom pre kompletizáciu IZRČ bola integrácia údajov z AZÚ, ktoré boli po prechádzajúcich výberoch obsiahnuté v IZRČ, s údajmi z územnej prípravy. Išlo teda o proces priradenia osoby z AZÚ na konkrétny adresný bod z elektronického systému sčítania domov a bytov (ďalej aj ako „ESDB“) na základe pobytu osoby, ktorý mu bol odvodený v IZRČ. Toto priradenie slúžilo aj na priebežný monitoring sčítania. Proces sa realizoval spojením údajov z ESDB s údajmi v IZRČ cez rôzne prepájacie kľúče. Pri dostupnosti sa prepojenie realizovalo cez tzv. building index – ide o jedinečný identifikátor adresného bodu. V ostatných prípadoch sa prepájanie realizovalo cez ďalšie lokalizačné premenné (obec, časť obce, súpisné číslo, orientačné číslo) v rôznych kombináciách. Prepojeniu predchádzala štandardizácia názvov časti obcí a ulíc, čiže transformácia názvov na veľké písmená bez diakritiky a odstránenie všetkých znakov okrem písmen a číslic. Výsledkom prepojenia bolo priradenie adresného bodu (ID_AB) pre záznamy v IZRČ a odvodených premenných prislúchajúcich k danému adresnému bodu – kód ulice, typ domu, súradnice, príznak duplicitných bytov, príznak medziposchodia, príznak centroidu.

Posledným krokom v procese spracovania bol výpočet odvodených premenných, ktoré zaznamenáva tabuľka č. 4.

Tabuľka č. 4: Odvodené premenné v IZRČ

Názov	Popis
ZDROJ_RC	Výskyt RČ v AZÚ
RC_EDIT_FLG	Príznak pre editované RČ
MENO_NORM	Krstné meno bez diakritiky
RC_RFO_FLG	Príznak pre RČ z RFO
RC_VALID_FLG	Príznak pre validné RČ
VALID_MENO_FLG	Príznak pre štandardné krstné meno
VALID_OBEC_FLG	Príznak pre validnú obec
VYBER_DOPLNKOVEHO_UDAJA_KOD	Kód pre výber doplnkového údajaja
EXPORT_FLG	Príznak pre export

Zdroj: ŠÚ SR, vlastné spracovanie

Jednotlivé odvodené premenné sa neprebrali priamo z AZÚ, ale odvádzali sa z údajov v IZRČ a signalizovali určitý príznak konkrétneho záznamu. Príznak výskytu rodného čísla v konkrétnych AZÚ popisovala premenná ZDROJ_RC. Bola tvorená tzv. maskou výskytu, v ktorej prítomnosť rodného čísla v maske vyjadrovala hodnota 1, neprítomnosť hodnota 0. Napríklad ak ZDROJ_RC = 110001000000000000, tak rodné číslo existovalo v RFO, FR_DP_A a SP_DP. Ďalšou premennou vzťahujúcou sa na rodné číslo bola RC_EDIT_FLG, v ktorej bol príznak o tom, či bolo rodné číslo editované. Príznak o validite rodného čísla bol nastavený v premennej RC_VALID_FLG, prítomnosť rodného čísla v RFO zasa v premennej RC_RFO_FLG. Prvé meno osoby v štandardizovanom tvare (veľké písmená a odstránená diakritika) sa nachádzalo v premennej MENO_NORM. Tento tvar mena slúžil na overenie mena pri autentifikácii obyvateľa pri vstupe do formulára. Autentifikácia menom v prvom pokuse prihlásenia sa do formulára však bola umožnená iba obyvateľom, ktorých meno malo príznak validity v premennej VALID_MENO_FLG. Za validné mená boli označené tzv. štandardné mená, ktorých výber pozostával z tých mien, ktoré sa podarilo spárovať so štandardnými kalendárovými menami obsiahnutými v oficiálnom kalendáriu mien vydávanom Ministerstvom kultúry SR alebo sa ich podarilo spárovať s najčastejšie používanými menami v Českej republike (vybraných 200 najčastejších mužských a 200 ženských mien), alebo sa spárovali s najčastejšie používanými menami v Maďarsku (vybraných 100 najčastejších mužských a 100 ženských mien). Do zoznamu boli doplnené aj ďalšie mená, ktoré neboli spárované v predchádzajúcich troch krokoch, ale sa bežne používajú (napr. Janka, Danko, Nikoleta a pod.). Ďalšou z premenných s informáciou o validite bola premenná VALID_OBEC_FLG, ktorá signalizovala validitu obce v konkrétnom zázname. Obec bola vyhodnotená ako validná, ak bol vyplnený kód obce a zároveň bol dátum prihlásenia na pobyt do referenčného dátumu sčítania. Na základe validity mena, resp. obce sa odvádzal príznak v premennej VYBER_DOPLNKOVEHO_UDAJA, ktorý označil, akým spôsobom sa môže obyvateľ po prihlásení do formulára autentifikovať. Poslednou premennou, ktorá signalizovala platný alebo neplatný záznam v celom IZRČ, bola premenná EXPORT_FLG. Záznam bol označený ako platný, čo znamenalo, že obyvateľovi bolo umožnené prihlásiť sa do formulára rodným číslom, keď bol pre dané rodné číslo dostupný dátum narodenia, ktorý je menší ako referenčný dátum sčítania alebo mu je rovný, pobyt osoby nebol ukončený do referenčného dátumu sčítania, typ pobytu bol vyplnený, premenná VYBER_DOPLNKOVEHO_UDAJA_KOD neobsahovala 00 (tzn. mala platné údaje potrebné na autentifikáciu), dátum úmrtia bol nevyplnený a adresný bod bol priradený.

Do formulára sa teda cez rodné číslo mohli prihlásiť obyvatelia, ktorým v IZRČ v premennej EXPORT_FLG bola priradená hodnota 1.

4. AUTENTIFIKÁCIA OBYVATEĽA V ELEKTRONICKOM SČÍTACOM FORMULÁRI

Povinnosť sčítať sa prostredníctvom vyplnenia sčítacieho formulára mali všetci obyvatelia Slovenskej republiky. Na splnenie tejto povinnosti mohol obyvateľ využiť webovú aplikáciu dostupnú na webstránke www.scitanie.sk alebo bezplatnú mobilnú aplikáciu pre operačné systémy iOS a Android. V oboch prípadoch bolo potrebné prejsť procesom autentifikácie, po ktorom bol obyvateľovi umožnený vstup do sčítacieho formulára. Možností autentifikácie bolo viac.

1. Rodné číslo a doplnkový údaj

V prípade autentifikácie rodným číslom a doplnkovým údajom musel obyvateľ pri vstupe do sčítacieho formulára vyplniť svoje rodné číslo v zobrazenom poli. Následne aplikácia od obyvateľa vyžiadala jeden z doplnkových údajov – obec pobytu alebo prvé meno. Ak systém vyžiadala od obyvateľa ako doplnkový údaj obec pobytu, aplikácia pracovala spôsobom autosuggestu pri postupnom vpisovaní jednotlivých písmen obce do vyhradeného poľa. Ak systém od obyvateľa žiadal pri autentifikácii meno, do vyhradeného poľa mohol obyvateľ zadávať veľké aj malé písmená s diakritikou aj bez diakritiky (pozri obr. č. 3).

Obrázok č. 3: Autentifikácia obyvateľa v elektronickom sčítacom formulári

Prihlásenie do sčítania obyvateľov 2021

Prihláste sa rodným číslom

Zadajte rodné číslo

Nie som robot

reCAPTCHA

Prihlásiť

Nemám rodné číslo

Prihlásiť sa elektronickým občianskym preukazom

Použite občiansky preukaz, alebo doklad o pobyte

Prihlásiť

Zdroj: ŠÚ SR

Ak obyvateľ zadal nesprávny prvý z náhodne vybraných doplnkových údajov, aplikácia si vyžiadala druhý. Ak obyvateľ zadal aj druhý doplnkový údaj nesprávne, autentifikácia bola neúspešná. Pri úspešnej autentifikácii bol obyvateľovi umožnený vstup do sčítacieho formulára a zobrazila sa mu úvodná strana formulára. Táto forma autentifikácie bola dostupná vo webovej aj v mobilnej aplikácii.

Pri autentifikácii rodným číslom sa z dôvodu bezpečnosti využil mechanizmus Recaptcha Enterprise, ktorý pri webovej verzii aplikácie vyžadoval manuálne overenie. V mobilnej aplikácii overenie prebehlo na pozadí bez potreby zásahu obyvateľa.

2. *Elektronický občiansky preukaz (eID)*

Druhým spôsobom, ako sa mohol obyvateľ v elektronickom sčítacom formulári autentifikovať, bolo použitie elektronického občianskeho preukazu. Pre úspešnú autentifikáciu musel obyvateľ disponovať:

- občianskym preukazom s čipom,
- pripojenou čítačkou čipových kariet,
- aktívnymi prihlasovacími údajmi do eID konta,
- nainštalovanou a spustenou aplikáciou pre eID,
- elektronickým OP správne vloženým do čítačky.

Ak si obyvateľ zvolil túto možnosť autentifikácie a splňal uvedené podmienky, po kliknutí na tlačidlo Prihlásiť v samostatnom poli ho systém presmeroval na stránku www.slovensko.sk, kde bol od neho vyžiadaný tzv. BOK – bezpečnostný osobný kód. V prípade zadania správneho BOK systém obyvateľa úspešne autentifikoval a následne ho presmeroval na prvú stranu sčítacieho formulára. Táto forma autentifikácie bola dostupná výhradne vo webovej aplikácii.

3. *Fotografia občianskeho preukazu (MRZ)*

Autentifikácia prostredníctvom odfotoграфovania, resp. naskenovania občianskeho preukazu bola možná výhradne v mobilných aplikáciách. Ak si obyvateľ v mobilnej aplikácii zvolil kliknutím na pole Prihlásiť možnosť Prihláste sa občianskym preukazom, aplikácia automaticky spustila fotoaparát s rámikom, do ktorého bolo potrebné umiestniť zadnú stranu preukazu, t. j. naskenovať tzv. Machine Readable Zone (MRZ). Systém následne načíta rodné číslo a meno, ktoré využil na autentifikáciu prvým opísaným spôsobom, t. j. rodným číslom a doplnkovým údajom – menom. Ak bola autentifikácia menom ako doplnkovým údajom neúspešná, aplikácia od obyvateľa vyžiada aj zadanie druhého doplnkového údaja – obce pobytu. Po úspešnej autentifikácii bol obyvateľ presmerovaný na prvú stranu sčítacieho formulára. Pri autentifikácii rodným číslom sa z dôvodu bezpečnosti opäť využil mechanizmus Recaptcha Enterprise, ktorého overenie prebehlo v mobilnej aplikácii na pozadí bez potreby zásahu obyvateľa.

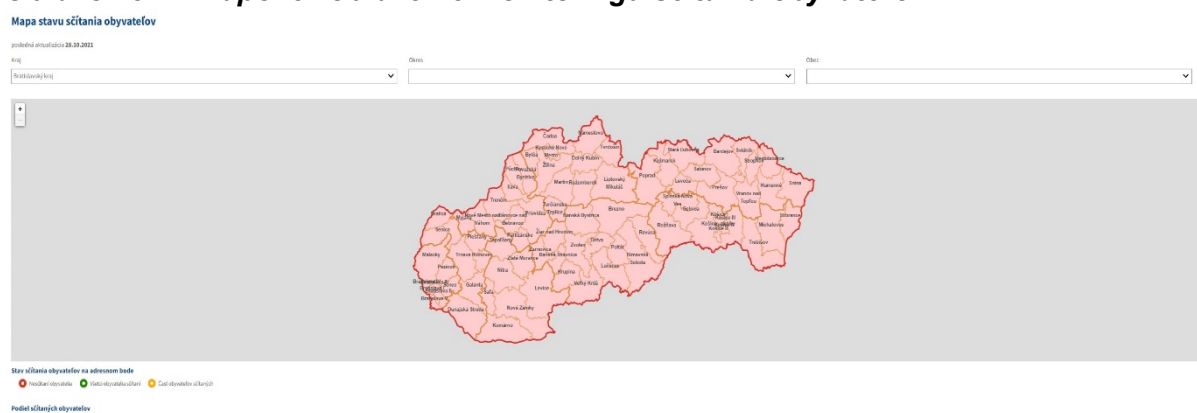
4. *Dátum narodenia a pohlavie*

Ďalšou možnosťou, ako sa autentifikovať v sčítacom formulári, bolo zadanie dátumu narodenia a pohlavia. Táto forma autentifikácie bola dostupná predovšetkým pre obyvateľov, ktorí nedisponujú rodným číslom, príp. spôsoby autentifikácie uvedené predtým v jeho prípade neboli úspešné. V tomto prípade však obyvateľ vyplňal všetky otázky, t. j. aj tie, ktoré sa pri autentifikácii rodným číslom v spracovaní obyvateľovi doplnili z administratívnych zdrojov údajov. V tomto spôsobe autentifikácie po kliknutí na tlačidlo Nemám rodné číslo obyvateľ zvolil z možností hodnoty pre deň a mesiac narodenia. Rok narodenia sa vpisoval ručne do vyhradeného poľa, pohlavie vybral z tzv. radiobuttonu. Po úspešnej autentifikácii bol obyvateľ presmerovaný na prvú stranu dlhého sčítacieho formulára. Táto forma autentifikácie bola dostupná vo webovej aj v mobilnej aplikácii. Rovnako aj pri autentifikácii dátumom narodenia a pohlavím sa kvôli bezpečnosti využil mechanizmus Recaptcha Enterprise, ktorý pri webovej verzii aplikácie vyžadoval manuálne overenie zo strany obyvateľa.

5. VYUŽITIE INTEGRAČNÉHO ZOZNAMU RODNÝCH ČÍSEL NA MONITORING SČÍTANIA

Dôvodom na tvorbu IZRČ bolo aj vytvorenie databázového podkladu na monitoring stavu sčítania obyvateľov. Vo webovej aplikácii v časti monitoring bolo umožnené oprávneným osobám sledovať vývoj stavu sčítaných obyvateľov k odhadovanému počtu obyvateľov až na úroveň obce. Odhadované počty obyvateľov pre jednotlivé obce boli stanovené na základe údajov z registra fyzických osôb. Do odhadovaných počtov vstupovali osoby, ktorých rodné číslo bolo dostupné v RFO a zároveň v tomto administratívnom zdroji mali platný trvalý pobyt. Nevstupovali tam osoby z RFO, ktoré sa narodili po referenčnom dátume, zomreli do referenčného dátumu a ktorých trvalý pobyt bol ukončený do referenčného dátumu. Osoby, ktoré boli oprávnené vstupovať do webovej aplikácie určenej na sledovanie stavu sčítania, mali k dispozícii grafické, tabuľkové aj mapové zobrazenie priebežného stavu sčítania obyvateľov.

Obrázok č. 4: Mapové zobrazenie monitoringu sčítania obyvateľov



Zdroj: ŠÚ SR, IS Sčítania obyvateľov

6. ZÁVER

Cieľom článku bolo detailne opísať účel tvorby, využitie a proces konštrukcie databázy IZRČ a jeho naviazanie na sčítací formulár. Čiastočne sme v článku opísali aj prácu s AZÚ, najmä vo vzťahu k tvorbe IZRČ. Administratívne zdroje údajov predstavujú databázy, ktorých správcami sú orgány štátnej správy. Takýmto zdrojom údajov je napríklad register fyzických osôb Ministerstva vnútra SR či Centrálny register poistencov zdravotného poistenia, ktorého správcom je Úrad pre dohľad nad zdravotnou starostlivosťou. Jedným z dôležitých predpokladov zmeny spôsobu sčítania od tradičného k integrovanému je dostupnosť práve takýchto administratívnych zdrojov údajov. Nový koncept realizácie SODB priniesol nové formy využitia údajov z AZÚ a s nimi zavádzanie nových postupov a iného, efektívnejšieho využitia AZÚ a práce s nimi. Na základe dostupnosti všetkých dátových zdrojov opísaných v 3. kapitole bola možná realizácia rôznych etáp prípravy SODB 2021 vrátane tvorby integračného zoznamu rodných čísel. Výsledná forma IZRČ obsahovala všetky rodné čísla, ktoré sa podarilo integráciou administratívnych zdrojov údajov získať. Tvorba tejto databázy umožnila čo najväčšiemu množstvu obyvateľov autentifikovať sa v elektronickom sčítaní, čo postačovalo vyplniť tzv. krátky formulár s menším počtom otázok, čím sa výrazne znížila administratívna záťaž na obyvateľa. IZRČ zároveň poslúžil ako prostriedok na zvýšenie kvality vyplnených údajov. Keď sa obyvateľa podarilo cez jeho rodné číslo a doplnkový údaj identifikovať a spárovať, bola mu poskytnutá adresnejšia forma sčítacieho formulára napr. vzhľadom na jeho vek alebo pohlavie, príj. iné údaje ktoré boli dostupné v databáze IZRČ.

LITERATÚRA

Zákon č. 223/2019 Z. z. o sčítaní obyvateľov, domov a bytov v roku 2021 a o zmene a doplnení niektorých zákonov.

RESUMÉ

Vzhľadom na zmenu spôsobu sčítania z tradičného na integrovaný tvoril integračný zoznam rodných čísel nevyhnutnú súčasť Sčítania obyvateľov, domov a bytov v roku 2021. Obsahoval komplexnú množinu údajov, ktorá v čo najvyššej možnej miere pokryla údaje potrebné na zníženie administratívnej záťaže obyvateľstva pri SODB 2021. Proces spracovania údajov obsiahnutých v AZÚ zahŕňal značné množstvo práce s údajmi, v konečnom dôsledku však zabezpečil základný predpoklad nového konceptu sčítania – elektronického sčítania. Ako sa opísalo v článku, takmer všetky ďalšie činnosti od autentifikácie cez vypĺňanie formulára až po monitoring stavu sčítania sa opierali práve o IZRČ.

RESUME

Due to the change in the method of census from traditional to integrated, the Integration list of identification numbers formed an essential part of the 2021 Population and Housing Census. It contained a comprehensive set of data covering as much as possible the data needed to reduce the administrative burden of the population at the 2021 Census21. The data processing contained in the administrative data sources covered a significant amount of work with the data, nevertheless, ultimately it provided the basic premise of the new census concept – the electronic census. As it is described in the article, almost all the other activities, from authentication, through the completion of forms to the monitoring of the census status, were based on the Integration list of identification numbers.

PROFESIJNÝ ŽIVOTOPIS

Mgr. Martin Kočiš absolvoval magisterské štúdium na Fakulte humanitných a prírodných vied Prešovskej univerzity v odbore geografia a regionálny rozvoj. Od roku 2016 pôsobí v Štatistickom úrade SR, kde najprv pracoval v oddelení demografickej štatistiky, od roku 2018 pôsobí v oddelení sčítania obyvateľov, domov a bytov a prierezových štatistík. V oblasti sčítania obyvateľov, domov a bytov sa venuje dátovým analýzám, najmä v súvislosti s prípravou a spracovaním AZÚ, ako aj spracovaním samotných výsledkov SODB.

Mgr. Cindy Erdelyi absolvovala magisterské štúdium na Fakulte sociálnych a ekonomických vied Univerzity Komenského v Bratislave v Ústave verejnej politiky. Od roku 2019 pôsobí v Štatistickom úrade SR, v oddelení sčítania obyvateľov, domov a bytov a prierezových štatistík. V oblasti sčítania obyvateľov, domov a bytov sa venuje dátovým analýzám, najmä v súvislosti s prípravou a spracovaním AZÚ, ako aj spracovaním samotných výsledkov SODB.

KONTAKTY

martin.kocis@statistics.sk

cindy.erdelyi@statistics.sk