

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

2/2022

ročník/volume 32

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

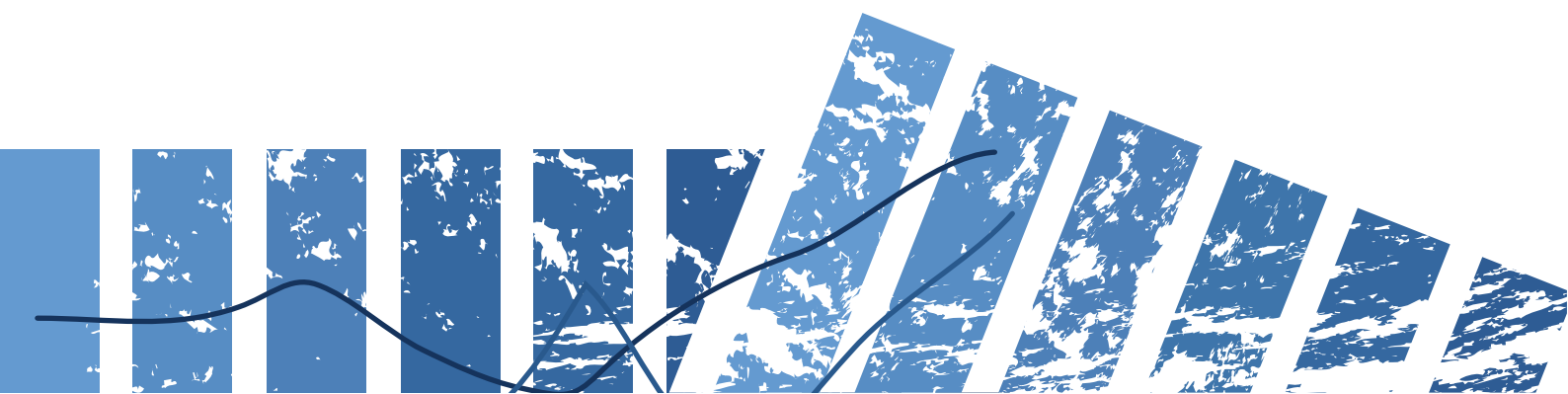
Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 4

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

Strany/Pages: 52 – 66

Dátum vydania/Publication date: 15. apríl 2022/April 15, 2022



Boris VAŇO
INFOSTAT – Výskumné demografické centrum

HODNOTENIE KVALITY ADMINISTRATÍVNYCH ÚDAJOV VSTUPUJÚCICH DO ŠTATISTICKÉHO PROCESU V PODMIENKACH SLOVENSKA

QUALITY EVALUATION OF ADMINISTRATIVE DATA ENTERING THE STATISTICAL PROCESS IN CONDITIONS OF SLOVAKIA

ABSTRAKT

Článok poukazuje na význam hodnotenia kvality administratívnych zdrojov údajov pred ich vstupom do štatistického prostredia. Podrobne opisuje schému na hodnotenie kvality AZÚ navrhutej pre potreby štatistickej praxe na Slovensku. Ide o nástroj, ktorý umožňuje Štatistickému úradu SR vykonávať takéto hodnotenie priamo a štandardizovane, v prípade potreby aj opakovane.

ABSTRACT

The article points to the importance of evaluating the quality of administrative data sources before their entry into the statistical environment. It describes in detail the evaluation scheme for evaluating the quality of the ADS designed for the needs of statistical practice in Slovakia. It is a tool allowing the Statistical Office of the SR to perform such assessment directly and in a standardized manner, if necessary also repeatedly.

KLÚČOVÉ SLOVÁ

administratívne zdroje údajov, hodnotenie kvality, schéma na hodnotenie kvality, indikátor kvality, metóda merania

KEY WORDS

administrative data sources, quality evaluation, evaluation scheme, quality indicator, measurement method

1. ÚVOD

Administratívne údaje nie sú primárne určené na účely, na ktoré ich chce využívať štatistický úrad, preto je veľmi dôležité, hodnotiť ich kvalitu, t. j. vhodnosť využitia na štatistické účely. V snahe vyrovnáť sa s výkyvmi v kvalite administratívnych údajov je potrebné, aby štatistický úrad bol schopný hodnotiť ich kvalitu na pravidelnej báze. Štatistický úrad SR (ďalej ŠÚ SR) má na tento účel k dispozícii procedúru, ktorá umožňuje rýchlo, priamo a štandardizovane určiť kvalitu administratívnych zdrojov údajov (ďalej AZÚ) pri ich vstupe do štatistického procesu¹. Takýto prístup je použiteľný pre každý AZÚ, ktorý sa využíva alebo má využívať na štatistické účely.

¹ Hodnotí sa len kvalita AZÚ pred začiatkom ich využívania na štatistické účely. Do úvahy sa neberie ďalšie využívanie AZÚ v rámci štatistického systému, t.j. fáza procesu alebo fáza výstupu.

2. NÁRODNÝ RÁMEC NA HODNOTENIE KVALITY AZÚ URČENÝCH NA ŠTATISTICKÉ ÚČELY

Hodnotenie kvality administratívnych údajov v kontexte ich využívania na štatistické účely je komplexný a odborne náročný proces. Na hodnotenie kvality administratívnych údajov pri ich vstupe do štatistického procesu možno využiť celý rad indikátorov. Pokiaľ má byť hodnotenie objektívne, je potrebné optimalizovať výber jednotlivých indikátorov kvality na základe univerzálnych aj špecifických kritérií. Preto sa prostredníctvom indikátorov kvality vytvárajú tzv. hodnotiace schémy, ktoré sú nástrojmi na komplexné hodnotenie AZÚ pred ich vstupom do štatistického procesu. V hodnotiacich schémach sa využíva rôzna kombinácia indikátorov kvality a ich rôzna váha, ktoré zodpovedajú potrebám a požiadavkám štatistickej praxe v príslušnej krajine a príslušnom období.

Schéma na hodnotenie kvality AZÚ nemá v Európskom štatistickom systéme ani v iných nadnárodných štruktúrach jednotnú podobu a každý štát si ju v konečnom dôsledku vytvára podľa vlastnej potreby a špecifik. V Európskom štatistickom systéme však existujú odporúčania a príklady dobrej praxe a k dispozícii sú aj skúsenosti z krajín, v ktorých takéto hodnotiace schémy existujú už dlhšie obdobie.

Pri tvorbe schémy na hodnotenie kvality AZÚ pre potreby ŠÚ SR sa vychádzalo z odporúčaní Eurostatu [3], [4], [5] aj zo zahraničných skúseností. Postupovalo sa tak, aby sa na jednej strane čo najobjektívnejšie posúdila vhodnosť využitia príslušného AZÚ na štatistické účely a na druhej strane aby sa neznemožnilo jeho využitie pre potreby štatistiky pre menej závažné dôvody. Za základ hodnotiacej schémy pre potreby ŠÚ SR sa zbral teoretický rámec kvality vypracovaný štatistickými úradmi v Holandsku, Taliansku a Švédsku [1], [2], [7], [9]. Pod rámcom kvality treba rozumieť ucelený systém merania kvality s presne zadefinovanou štruktúrou [6], [8].

Na jednoduchšie narábanie s hodnotiacou schémou a na celkovú objektivizáciu hodnotenia kvality AZÚ je potrebné správne a jednotne interpretovať jednotlivé pojmy, ktoré sa nachádzajú v hodnotiacej schéme. Ide hlavne o termíny súvisiace s hierarchickým pohľadom na štatistickú kvalitu administratívnych údajov (hyperdimenzia, dimenzia, indikátor kvality, metóda merania), o termíny spojené s úrovňou aplikácie indikátorov kvality (databáza, objekt, veta, premenná) a o charakteristiky jednotlivých častí administratívnej databázy (objektov, premenných)². V prípade indikátorov kvality sa pri interpretácii venuje pozornosť hlavne kvantitatívnym indikátorom, ktoré sú založené na numerických metódach a ich

² **Hyperdimenzia** predstavuje pohľad na kvalitu na najvyššej hierarchickej úrovni. Každá hyperdimenzia obsahuje najmenej 2 dimenzie.

Dimenzia je druhá najvyššia hierarchická úroveň v kvalitatívnom rámci. Kvalitatívne aspekty obsiahnuté v dimenzii sú výrazne ovplyvnené kontextuálnym pohľadom na kvalitu AZÚ. Každá dimenzia obsahuje najmenej 1 indikátor kvality.

Indikátor kvality charakterizuje špecifickú oblasť, za ktorú sa meria kvalita AZÚ. Pre každý indikátor kvality je k dispozícii najmenej 1 metóda merania.

Metóda merania je kvalitatívna alebo kvantitatívna metóda, ktorá slúži na meranie alebo odhad indikátora kvality.

Štatistická jednotka je jednotka v štatistickom súbore.

Objekt je jednotka v AZÚ. Môže sa líšiť od jednotky v štatistickom súbore, preto je vhodné odlišiť ju aj terminologicky. Ide teda o identifikáciu registrovaných udalostí alebo jednotiek v AZÚ. Z objektov by sa mali dať odvodiť štatistické jednotky.

Veta je skupina údajov, ktoré sa vzťahujú na jeden objekt alebo štatistickú jednotku.

Premenná je konkrétny údaj o objekte alebo štatistickej jednotke, ktorý je súčasťou vety.

interpretácia je podmienená znalosťou týchto metód. Pri kvalitatívnych indikátoroch vyplýva interpretácia viac-menej z názvu indikátora, pričom obvykle ide o otázku na niektorý z aspektov hodnotenia kvality AZÚ.

Pri hodnotení kvality AZÚ využívaných na štatistické účely je veľmi dôležitá spolupráca medzi správcom AZÚ a štatistickým úradom. Preto je pri viacerých indikátoroch kvality potrebná priama komunikácia so správcom AZÚ (väčšinou vo forme poskytnutia určitých informácií).

3. HODNOTIACA SCHÉMA KVALITY AZÚ PRE POTREBY ŠÚ SR

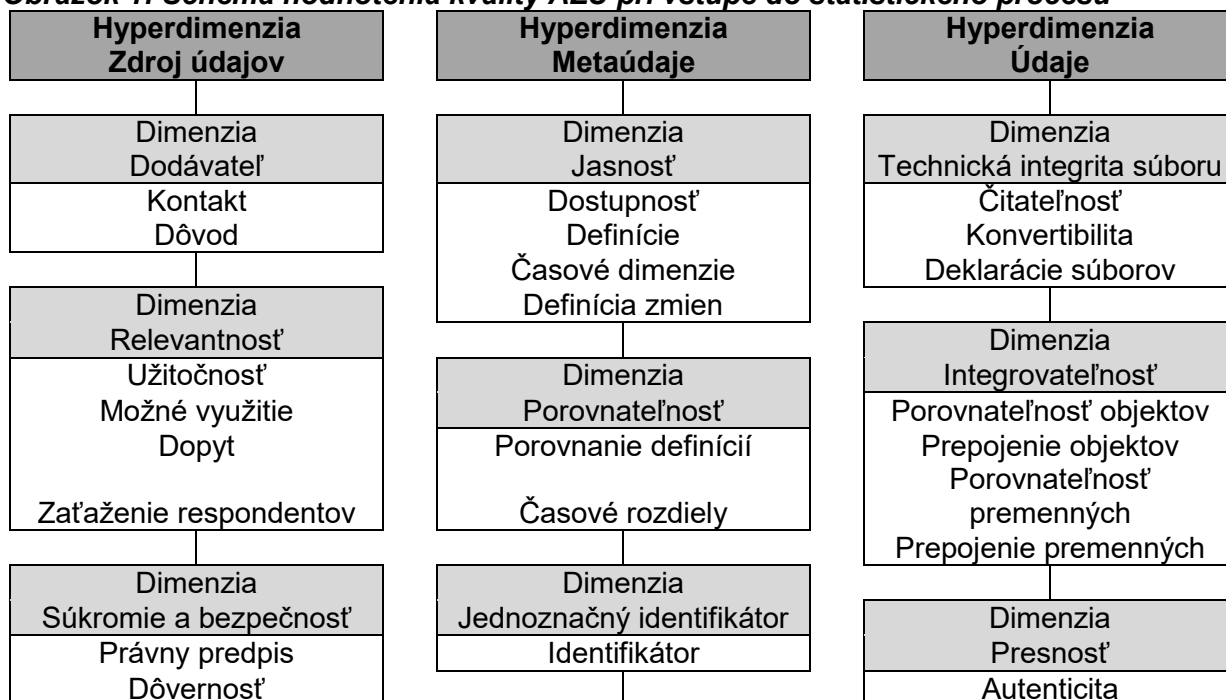
Z teoretického rámca kvality sa v hodnotiacej schéme pre potreby ŠÚ SR využili všetky hyperdimenzie, dimenzie a indikátory kvality, ako aj väčšina metód merania. Hodnotiace kritériá pri jednotlivých indikátoroch kvality aj pri celkovom hodnotení AZÚ sú nastavené v súlade s potrebami štatistickej praxe na Slovensku (s osobitným dôrazom využitie v oblasti sociálnych štatistik a cenzu).

Opis schémy na hodnotenie kvality AZÚ pred ich vstupom do štatistického procesu, navrhutej pre potreby ŠÚ SR obsahuje podrobnú informáciu o všetkých navrhnutých indikátoroch kvality a najčastejšie používaných metódach merania. Je usporiadaný podľa hyperdimenzií a dimenzií [6], [8]. Prehľadné zobrazenie všetkých dimenzií a indikátorov kvality pre všetky tri hyperdimenzie v hodnotiacej schéme sa nachádza na obrázku 1. Prvé dve hyperdimenzie (zdroj údajov a metaúdaje) hodnotia AZÚ ako celok a na hodnotenie využívajú väčšinou kvalitatívne metódy. Tretia hyperdimenzia (údaje) je zameraná na hodnotenie jednotlivých údajov, ktoré sa nachádzajú v AZÚ, pričom na hodnotenie sa využívajú skoro výlučne kvantitatívne metódy.

3.1 Hyperdimenzia zdroj údajov

Hyperdimenzia zdroj údajov obsahuje 5 dimenzií, ktoré pomocou kvalitatívnych metód merania poskytujú základnú charakteristiku príslušného AZÚ. Výnimkou je len dimenzia relevantnosť, v ktorej sa využíva kvantitatívna metóda merania.

Obrázok 1: Schéma hodnotenia kvality AZÚ pri vstupe do štatistického procesu



Dimenzia procedúry

Dimenzia procedúry obsahuje 4 indikátory kvality a rovnaký počet metód merania. Hodnotenie v rámci tejto procedúry sa zameriava na špecifické, resp. problémové situácie, ktoré by mohli nastať v AZÚ.

Indikátor kvality zber údajov sa týka spôsobu zberu administratívnych údajov, indikátor kvality plánované zmeny zas prípadných zmien v AZÚ. Indikátor kvality spätná väzba zisťuje komunikáciu s dodávateľom AZÚ a indikátor kvality núdzový scenár hodnotí pripravené riešenia v prípade problémov s dodaním AZÚ.

Všetky štyri metódy merania majú podobu otázky a sú teda kvalitatívne. Otázka pri indikátore kvality zber údajov zisťuje, či je ŠÚ SR informovaný o spôsobe zberu administratívnych údajov. Pri indikátore zmeny sa metóda merania zameriava na zistenie, či je ŠÚ SR včas informovaný o prípadných zmenách v AZÚ. Metóda merania pri indikátore kvality spätná väzba umožňuje zistiť, či má ŠÚ SR možnosť kontaktovať dodávateľa AZÚ v prípade problémov. Metóda merania pri indikátore kvality núdzový scenár zisťuje, či existuje núdzový scenár pre prípad problémov s dodaním AZÚ. Ide predovšetkým o problémy s termínom dodania a kvalitou dodaných údajov.

3.2 Hyperdimenzia metaúdaje

Hyperdimenzia metaúdaje hodnotí kvalitu AZÚ v zmysle definícií, porovnateľnosti, identifikácie a kontrolnej činnosti zo strany poskytovateľa AZÚ. Jasnosť definícií a kompletnosť metainformácií patria medzi hlavné kvalitatívne aspekty v tejto hyperdimenzii. Hyperdimenzia metaúdaje obsahuje 4 dimenzie a 12 metód merania.

Dimenzia jasnosť

Dimenzia jasnosť hodnotí jasnosť údajov z rôznych aspektov. Na tento účel využíva 4 indikátory kvality a rovnaký počet metód merania. Indikátor kvality dostupnosť zisťuje existenciu metaúdajov a či boli dodávateľom ŠÚ SR aj poskytnuté. Indikátor kvality definície sa zameriava na definovanie údajov v AZÚ prostredníctvom metaúdajov. Ide o definície zisťovaných jednotiek, klasifikačných premenných a početností. Indikátor kvality časové dimenzie hodnotí administratívne údaje z časového hľadiska. Indikátor kvality definícia zmien sa zameriava na zmeny a úpravy v definíciách jednotlivých údajov v AZÚ.

Indikátor kvality dostupnosť využíva ako metódu merania otázku, či sú k dispozícii metaúdaje a či boli poskytnuté ŠÚ SR spolu s AZÚ. Metódou merania pri indikátore kvality definície je otázka, či sú údaje nachádzajúce sa v AZÚ jasne definované v metaúdajoch. Otázka, ktorá prislúcha k indikátoru kvality časové dimenzie, zisťuje, či je časové obdobie, na ktoré sa vzťahujú údaje v AZÚ, jasne opísané. Indikátor kvality zmeny v definíciách využíva ako metódu merania otázku, či nastali zmeny v definíciách, a ak áno, či boli jasne a včas odkomunikované.

Dimenzia porovnateľnosť

Dimenzia porovnateľnosť hodnotí porovnateľnosť AZÚ so štandardmi, ktoré sa používajú v ŠÚ SR. V tejto dimenzii sú dva indikátory kvality a dve metódy merania. Indikátor kvality porovnanie definícií hodnotí porovnateľnosť definícií používaných v AZÚ a v ŠÚ SR. Ako už bolo spomenuté pri predchádzajúcej dimenzii, ide o definície

zisťovaných jednotiek, klasifikačných premenných, ako aj početností. Indikátor kvality „Časové rozdiely“ hodnotí porovnateľnosť časových referencií používaných v AZÚ a v ŠÚ SR.

Obidva indikátory kvality sú kvalitatívne, ako metódu merania využívajú otázky, ktoré zisťujú, či sú definície v AZÚ a v ŠÚ SR porovnateľné (indikátor kvality porovnanie definícií) a či sú časové referencie používané v AZÚ a na ŠÚ SR porovnateľné (indikátor kvality časové rozdiely).

Dimenzia jednoznačný identifikátor

Dimenzia jednoznačný identifikátor má jeden indikátor kvality a dve metódy merania. Kvalitatívny indikátor identifikátor hodnotí existenciu a porovnateľnosť jednoznačného identifikátora v AZÚ. Na hodnotenie využíva dve otázky, pomocou ktorých zisťuje, či v AZÚ existuje jednoznačný identifikátor a či je porovnateľný s identifikátormi používanými v ŠÚ SR.

Dimenzia práca s údajmi

Táto dimenzia hodnotí prácu správcu AZÚ s údajmi v zdroji. Ide o všetky vykonané zmeny a kontrolnú činnosť správcu. Dimenzia obsahuje 2 indikátory kvality a 4 kvalitatívne metódy merania. Na hodnotenie v rámci tejto dimenzie sú potrebné podklady od poskytovateľa AZÚ.

Indikátor kvality kontroly hodnotí kontrolnú činnosť správcu AZÚ. Ide o kontrolu zisťovaných jednotiek, kontrolu premenných, kontrolu kombinácie premenných, kontrolu extrémnych hodnôt. Indikátor kvality modifikácie hodnotí zasahovanie do AZÚ zo strany správcu. Za modifikácie sa považuje editovanie, imputácie a dosadzovanie difoltných hodnôt. Hodnotenie kvality sa zameriava na oboznámenosť s modifikáciou údajov, označenie modifikovaných údajov a oboznámenosť s použitím difoltných hodnôt. Metódy merania pri oboch indikátoroch kvality majú podobu otázok. V prípade indikátora kvality kontroly sa otázkami zisťuje, či správca kontroloval údaje v AZÚ a či sa pri kontrole špeciálne zameril na výskyt extrémnych hodnôt. Pri indikátore kvality modifikácie sa zisťuje, či správca modifikoval údaje v AZÚ a či sú modifikované hodnoty v AZÚ aj vyznačené.

3.3 Hyperdimenzia údaje

Hyperdimenzia údaje sa venuje rôznym aspektom kvality údajov, ktoré sa nachádzajú v príslušnom AZÚ. Je zo všetkých hyperdimenzií najrozsiahlejšia, obsahuje 5 dimenzií, 25 indikátorov kvality a 39 metód merania. Na hodnotenie kvality sa v tejto hyperdimenzii využívajú takmer výlučne kvantitatívne metódy merania.

Dimenzia technická integrita súboru

Dimenzia technická integrita súboru obsahuje 3 indikátory kvality a 3 metódy merania. Indikátor kvality čitateľnosť hodnotí dostupnosť súboru a všetkých údajov v súbore. Problémom z hľadiska čitateľnosti môže byť napríklad, keď je súbor v neznámom formáte, keď je súbor poškodený alebo keď nie je možné súbor dekódovať. Za čitateľný sa považuje taký súbor, ktorý má čitateľné všetky vety. AZÚ sa považuje za čitateľný, ak sú čitateľné všetky jeho súbory. Metóda merania, ktorá v hodnotiacej schéme prislúcha indikátoru kvality čitateľnosť, hodnotí podiel nečitateľných súborov z celkového počtu dodaných súborov (obvykle v percentách).

Alternatívne možno nečitateľnosť hodnotiť aj kapacitným vyjadrením (napríklad v MB) ako rozsah poškodených súborov k celkovému rozsahu dodaných súborov.

Indikátor kvality konvertibilita hodnotí možnosť previesť súbor do formátu, ktorý je štandardom v ŠÚ SR. Problémom konvertibility môže byť napr. poškodenie súboru pri dekódovaní alebo chybné údaje v súbore po konverzii. Počas konverzie môžu vzniknúť chyby, v dôsledku ktorých nie je možné súbor otvoriť, alebo v súbore, ktorý je možné otvoriť, sú chybné niektoré údaje. Metóda merania, ktorú využíva indikátor kvality konvertibilita, udáva podiel objektov s chybou pri dekódovaní alebo s chybnými údajmi po konverzii z celkového počtu objektov v AZÚ.

Indikátor kvality deklarácie súborov hodnotí súlad údajov v AZÚ s informáciami v metaúdajoch. Rozdiely medzi údajmi v súbore a metaúdajmi môžu byť v názvoch a formátoch, v obsahu premenných a príznakov, v kategóriách definovaných pre kvalitatívne premenné a v rozsahu pre kvantitatívne premenné. Problémom môže byť napr. chýbajúci popis metaúdajov alebo údaje v súbore s nesprávnou hlavičkou. Zásadnou požiadavkou je, aby boli metaúdaje k dispozícii. Metóda merania použitá v hodnotiacej schéme vyjadruje podiel premenných v AZÚ, ktoré sa líšia od špecifikácie uvedenej v časti metaúdaje.

Dimenzia integrovateľnosť

V dimenzii integrovateľnosť sa nachádzajú 4 indikátory kvality a 9 metód merania. Všetky metódy merania v tejto dimenzii sú kvantitatívne.

Indikátor kvality porovnateľnosť objektov hodnotí podobnosť objektov v AZÚ s objektmi používanými v ŠÚ SR. Na hodnotenie kvality sa využíva len v prípade, ak sa AZÚ používa na nahradenie, doplnenie alebo kontrolu existujúceho štatistického zisťovania. V AZÚ môžu byť identické, zodpovedajúce a neintegrovateľné objekty. Identické administratívne objekty majú rovnakú definíciu a jednotku ako štatistické objekty. Zodpovedajúce administratívne objekty nie sú identické so štatistickými objektmi, ale je možné ich konvertovať. Neintegrovateľné administratívne objekty nie sú zhodné so štatistickými objektmi a nie je ich možné ani integrovať. Metódy merania pri indikátore kvality porovnateľnosť objektov využívajú jednotlivé typy administratívnych objektov. To znamená, že určujú podiel identických, zodpovedajúcich alebo neporovnateľných administratívnych objektov na celkovom počte relevantných objektov v AZÚ.

Indikátor kvality prepojenie objektov hodnotí možnosť prepojenia (párovania) objektov v AZÚ a štatistickom súbore. Identické administratívne objekty sú zhodné so štatistickými objektmi, preto ich prepojenie je bezproblémové. Zodpovedajúce administratívne objekty možno prepojiť so štatistickými objektmi až po konverzii. Nespojené administratívne objekty nie je možné spojiť so žiadnym štatistickým objektom. Za východisko pri prepájaní sa vždy považujú objekty v štatistickom súbore. Metódy merania pri tomto indikátore kvality vyjadrujú podiel identických priradených objektov, zodpovedajúcich priradených objektov, nepriradených objektov alebo nepriradených agregovaných objektov na celkovom počte objektov v AZÚ.

Indikátor kvality porovnateľnosť premenných hodnotí blízkosť hodnôt premenných v AZÚ s hodnotami premenných v iných zdrojoch používaných v ŠÚ SR. Metódou merania je podiel objektov s identickými hodnotami premenných v oboch zdrojoch z celkového počtu objektov v oboch zdrojoch (štatistickom aj administratívnom).

Indikátor kvality prepojenie premenných hodnotí užitočnosť prepájacích premenných (napr. identifikačných čísel) pre objekty v AZÚ. Problémom môžu byť objekty bez prepájacích premenných alebo objekty s prepájacími premennými, ktoré sa líšia od

tých, ktoré používa štatistický úrad. Ako metódu merania možno využiť podiel objektov s neprepojenými premennými, podiel objektov s prepojenými premennými, ktoré sú rozdielne od premenných používaných v štatistickom úrade, alebo podiel objektov s korektne konvertovateľnými prepojenými premennými z celkového počtu objektov s prepojenými premennými v AZÚ.

Dimenzia presnosť

Presnosti, ako kritériu kvality sa v hodnotiacej schéme venuje veľká pozornosť. V tejto dimenzii sa nachádza 6 indikátorov kvality a 7 kvantitatívnych metód merania. Ako alternatíva k jednej z kvantitatívnych metód merania sa môže využiť 5 kvalitatívnych metód, ktoré majú podobu otázok.

Indikátor kvality autenticita je zameraný na hodnotenie legitimacy objektov v AZÚ. Autenticita zahŕňa syntaktickú správnosť použitých identifikátorov (ak existujú) a zhodu objektov v AZÚ s príslušnými reálnymi objektmi. Problémom je objekt s nesprávnym identifikátorom (syntaktická chyba), resp. syntakticky správny ale nesprávne priradený identifikátor. Pokiaľ existuje v zdroji viac identifikátorov (napr. identifikátor domácnosti a identifikátor osoby), treba skontrolovať správnosť každého z nich. Na identifikáciu objektov s nesprávne priradeným identifikátorom je potrebné porovnanie s referenčným zoznamom. Objekty vyhodnotené ako problematické z hľadiska tohto indikátora kvality sa označujú ako neautentické. Na hodnotenie ako metóda merania slúži podiel objektov so syntakticky nesprávnym identifikátorom z celkového počtu objektov v AZÚ alebo podiel objektov, ktoré v súbore obsahujú protichodné informácie v porovnaní s informáciami v referenčnom zozname pre tieto objekty (z celkového počtu objektov v AZÚ).

Indikátor kvality nekonzistentné objekty je zameraný na celkovú konzistentnosť objektov v AZÚ. Nekonzistentné objekty sú zapojené do nelogických vzťahov s inými objektmi. Indikátor sa používa, len ak sú v zdrojovom súbore viacnásobné objekty. Problémom je napr., ak je jedna osoba priradená k viacerým domácnostiam alebo lokálna obchodná jednotka k viac ako jednému podniku. Metódou merania sa v tejto dimenzii zisťuje podiel objektov zapojených do nelogických vzťahov s inými objektmi z celkového počtu objektov v AZÚ.

Indikátor kvality pochybné objekty sa zameriava na existenciu nedôveryhodných objektov. Ide o objekty, ktoré majú nepravdepodobný, avšak nie jednoznačne nesprávny vzťah s inými objektmi. Identifikácia pochybných objektov si vyžaduje dobrú znalosť súboru. Indikátor má zmysel len v prípade, že súbor obsahuje viacnásobné objekty. Za príklad pochybného objektu možno označiť domácnosť, ktorá má 25 členov, t. j. 25 osôb (objektov) je priradených k tej istej domácnosti, alebo manželov vo veku nad 100 rokov. Metódou merania sa zisťuje podiel objektov zapojených do nepravdepodobných (nie však nevyhnutne nesprávnych) vzťahov s inými objektmi z celkového počtu objektov v AZÚ.

Indikátor kvality chyby merania hodnotí korektnosť hodnôt premenných v AZÚ. Meria odchýlku aktuálne nameraných údajov od ideálnych bezchybných meraní (zisťovaní), resp. rozsah, v akom nameraná hodnota zodpovedá hodnote, ktorá mala byť nameraná. Existuje viacero dôvodov, pre ktoré nemá zistená hodnota potrebnú validitu. Môže to byť napríklad chyba pri hlásení (evidované objekty majú záujem byť evidované určitým spôsobom) alebo chyba pri spracovaní (administratívna prax alebo

spôsob spracovania môže mať za následok skreslené údaje). Keďže všetky dôvody súvisia s procesom získavania alebo spracovania údajov, je nevyhnutné získať potrebné informácie na výpočet tohto indikátora od poskytovateľa AZÚ. Pri tomto indikátore kvality sa v hodnotiacej schéme nachádza jedna kvantitatívna metóda merania, ktorá udáva podiel neoznačených chybných hodnôt zo všetkých hodnôt v AZÚ. Táto metóda je použiteľná len v prípade, že chybné hodnoty sú v AZÚ vyznačené. Ak poskytovateľ AZÚ chybné hodnoty nevyznačil, je potrebné ho kontaktovať a hodnotenie získať na základe niekoľkých otázok, ktoré sú zamerané na uplatnenie stanoveného dizajnu v procese zberu údajov, kontrolu údajov počas fázy reportingu, používanie štandardov pre príslušné premenné, kontrolný proces pre vstup údajov a kontroly počas spracovania alebo ukladania údajov.

Indikátor kvality nekonzistentné premenné hodnotí premenné z hľadiska konzistentnosti hodnôt v prípade kombinácií premenných v AZÚ, t. j. do akej miery sú údaje v AZÚ vnútorne konzistentné. Hodnotenie konzistentnosti si vyžaduje dobrú znalosť údajov v súbore. Príkladom nekonzistentnosti je napr. kombinácia vek 10 rokov a rodinný stav rozvedený. Inak povedané, ide o všetky kombinácie hodnôt premenných, ktoré odporujú niektorej definícii. Metóda merania pre tento indikátor kvality udáva podiel objektov, ktoré obsahujú nelogickú kombináciu hodnôt premenných z celkového počtu objektov v AZÚ.

Indikátor kvality pochybné premenné zisťuje existenciu nepravdepodobných kombinácií hodnôt premenných objektu. Ide o kombinácie, ktoré sú podozrivé, nie však nevyhnutne nesprávne. Obvykle ide o kombinácie hodnôt, ktoré výrazne prevyšujú zvyčajné hodnoty, napríklad podnik, ktorý má 10-krát vyšší obrat, ako je obvyklé pri danom type podnikov, alebo matka s 15 deťmi. Posúdenie pochybnosti hodnôt vyžaduje dobrú znalosť údajov v AZÚ. Metóda merania v tomto prípade udáva podiel objektov, ktoré obsahujú nepravdepodobné kombinácie hodnôt premenných z celkového počtu objektov v AZÚ.

Dimenzia úplnosť

Dimenzia úplnosť hodnotí rôzne stránky úplnosti súboru, nie len v zmysle chýbajúcich alebo nadbytočných hodnôt, ale aj ich podhodnotenia alebo nadhodnotenia ako aj metód riešenia úplnosti. Obsahuje 6 dimenzií a 7 metód merania. Všetky metódy merania sú kvantitatívne.

Indikátor kvality podhodnotenie hodnotí, či v AZÚ chýbajú objekty, ktoré sa nachádzajú v referenčnom zozname. Referenčným zoznamom, s ktorým sa porovnávajú zdrojové údaje, je cieľový zoznam, ktorý berie do úvahy ŠÚ SR. Príkladom chýbajúcich objektov sú napríklad osoby, ktoré boli sčítané v sčítaní obyvateľov a spĺňajú podmienku na zaradenie do registra obyvateľov, ale v čase sčítania sa v registri nenachádzajú. Metódou merania je podiel objektov z referenčného zoznamu, ktoré chýbajú v AZÚ (z celkového počtu objektov v referenčnom zozname).

Indikátor kvality nadhodnotenie hodnotí existenciu iných ako cieľových objektov v AZÚ. Ide o také objekty, ktoré v príslušnom referenčnom období nepatria do cieľového súboru štatistického úradu. Napríklad, keď sa v registri obyvateľov v čase sčítania nachádzajú osoby, ktoré neboli sčítané. Metódou merania je podiel objektov v AZÚ, ktoré sa nenachádzajú v referenčnom zozname (z celkového počtu objektov).

Indikátor kvality selektívnosť sa týka štatistického pokrytia a reprezentatívnosti objektov v AZÚ. Indikátor je zameraný hlavne na objekty, ktoré nechýbajú náhodne. Na posúdenie selektivity je potrebná informácia o celom súbore. Ako referenčný súbor sa použije cieľový zoznam príslušnej štatistiky. Selektivita sa v hodnotiacej schéme meria pomocou indikátora reprezentatívnosti (R – indikátor) alebo prostredníctvom tabuľkového diagramu.

Indikátor kvality nadbytočnosť hodnotí viacnásobnú evidenciu identických objektov v AZÚ, a to buď na základe duplicity identifikátorov, ktoré majú byť evidované len raz alebo na základe duplicitnej hodnoty pre vybrané premenné, ktoré slúžia na identifikáciu. Za duplicitné objekty sa považujú také objekty, ktoré majú rovnaký identifikátor alebo ktoré majú rovnakú hodnotu pre vybrané premenné. Na meranie duplicitnosti sa v hodnotiacej schéme využíva podiel duplicitných objektov z celkového počtu objektov v AZÚ.

Indikátor kvality chýbajúce hodnoty sa týka chýbajúcich hodnôt premenných v AZÚ. Môže ísť o jednotlivú premennú alebo o kombináciu premenných. Hodnotenie si vyžaduje dobrú znalosť súboru. Na meranie sa využíva podiel objektov s chýbajúcou hodnotou pre príslušnú premennú (zo všetkých objektov v AZÚ).

Indikátor kvality imputované hodnoty sa týka existencie doplnených hodnôt v AZÚ zo strany poskytovateľa. Štatistický úrad vie identifikovať imputované hodnoty len v prípade, ak ich poskytovateľ označí. Pokiaľ sú imputované hodnoty vyznačené v súbore, hodnotenie nevyžaduje ďalšiu spoluprácu zo strany poskytovateľa AZÚ. Pokiaľ nie sú imputované hodnoty vyznačené, musí ich poskytovateľ pred hodnotením vyznačiť. Hodnotenie sa robí na základe podielu imputovaných hodnôt, ktoré pripadajú na jednu premennú.

Dimenzia časové hľadisko

Posledná dimenzia v hyperdimenzii dáta je zameraná na časové hľadisko dodania AZÚ do ŠÚ SR a na zmeny, ktoré sa udiali v porovnaní s predchádzajúcim dodaným súborom (v prípade, že sa AZÚ nedodáva do ŠÚ SR prvý raz). Ide o dôležitú stránku hodnotenia kvality, pretože oneskorené doručenie môže ovplyvniť použiteľnosť údajov. Dimenzia časové hľadisko obsahuje 6 indikátorov kvality a 8 metód merania. Všetky metódy merania v tejto dimenzii sú kvantitatívne.

Indikátor kvality včasnosť hodnotí časový rozdiel medzi koncom referenčného obdobia a okamihom prevzatia AZÚ štatistickým úradom. Metódou merania je časový rozdiel (obvykle v dňoch), ktorý sa určí ako rozdiel medzi dátumom doručenia AZÚ do ŠÚ SR a dátumom konca referenčného obdobia, na ktoré sa viažu údaje v AZÚ.

Indikátor kvality presnosť hodnotí časový rozdiel medzi dátumom doručenia AZÚ do ŠÚ SR a dohodnutým dátumom doručenia. Metódou merania je časový rozdiel (obvykle vyjadrený v dňoch), ktorý sa pri tomto indikátore kvality počíta ako rozdiel medzi skutočným a dohodnutým dátumom doručenia AZÚ do ŠÚ SR.

Indikátor kvality celková doba oneskorenia hodnotí celkový časový rozdiel medzi koncom referenčného obdobia a okamihom, keď štatistický úrad potvrdil, že súbor sa môže začať využívať. Na rozdiel od predchádzajúcich indikátorov kvality, tento indikátor berie do úvahy aj čas potrebný na vyhodnotenie dodaných údajov. Metódou

merania je pri tomto indikátore celkový časový rozdiel (v dňoch), ktorý sa vypočíta ako rozdiel medzi dátumom, keď ŠÚ SR deklaruje, že AZÚ môže začať využívať, a dátumom konca referenčného obdobia, na ktoré sa viažu údaje v AZÚ.

Indikátor kvality oneskorenie hodnotí rýchlosť, s akou sa prípadné zmeny v AZÚ zachytia v evidencii. Týka sa to zmien v zložení súboru aj zmien v hodnotách premenných pre objekty v súbore. Súbory s veľkým oneskorením poskytujú štatistickému úradu viac alebo menej zastarané údaje. Využitie tohto indikátora vyžaduje podklady od poskytovateľa AZÚ. Metódou merania je časový rozdiel v dňoch, ktorý sa vypočíta ako rozdiel medzi dátumom zachytenia zmeny v AZÚ zo strany poskytovateľa a dátumom, keď zmena nastala.

Indikátor kvality dynamika objektov hodnotí identifikáciu zmien v súbore v priebehu času. Ide o zmeny, ktoré sú spôsobené vznikom a zánikom objektov. Objekty, ktoré zostávajú v súbore, sa nazývajú pretrvávajúce objekty. Adekvátne zaraďovanie nových objektov a vyradovanie zaniknutých objektov je nevyhnutný predpoklad na kvalitnú aktualizáciu súboru.

Pre objekty v súbore platí nasledujúca klasifikácia (t je aktuálne obdobie):

nové objekty (t) = objekty existujúce v čase t avšak nie v čase $t-1$

pretrvávajúce objekty (t) = objekty existujúce v čase t a zároveň v čase $t-1$

zaniknuté objekty (t) = objekty existujúce v čase $t-1$ avšak nie v čase t

všetky objekty (t) = nové objekty (t) + pretrvávajúce objekty (t)

všetky objekty ($t-1$) = pretrvávajúce objekty (t) + zaniknuté objekty (t)

Dynamiku objektov je možné hodnotiť pomocou troch metód merania. Ide o podiel nových, zaniknutých alebo pretrvávajúcich objektov z celkového počtu objektov v AZÚ.

Indikátor kvality stabilita premenných hodnotí zmeny premenných v čase. Má zmysel len pre pretrvávajúce objekty. Pre niektoré premenné je nevyhnutné, aby sa v čase nemenili (napr. klasifikačné kódy a pod.). Kvalita sa v tomto prípade hodnotí pomocou podielu objektov so zmenenými hodnotami z celkového počtu objektov v AZÚ. Do hodnotenia sa zaraďujú len pretrvávajúce objekty a stabilné premenné, teda tie, ktoré by sa v čase nemali meniť.

4. POUŽITIE HODNOTIACEJ SCHÉMY

Opísaná schéma na hodnotenie kvality AZÚ vstupujúcich do štatistického systému obsahuje 51 indikátorov kvality a 74 metód merania. Ku každému indikátoru kvality je priradená najmenej jedna metóda merania. Pre niektoré indikátory kvality je v hodnotiacej schéme k dispozícii viac metód merania, ktoré však môžu byť navrhnuté alternatívne. Pokiaľ sú metódy merania navrhnuté alternatívne, vyberie sa len jedna z metód a tá sa použije na hodnotenie. Výber jednej z alternatívnych metód je plne v kompetencii hodnotiteľa. Kritériom pri výbere môže byť napríklad dostupnosť údajov alebo zložitosť výpočtu. Pokiaľ nie sú metódy merania navrhnuté alternatívne, môže sa pre príslušný indikátor kvality použiť viac metód merania.

Na základe kvalitatívnej alebo kvantitatívnej hodnoty, ktorá sa získa pomocou príslušnej metódy merania (odpoveď na otázku alebo výpočet ukazovateľa), sa indikátoru kvality pridelia body. Bodové hodnotenie je nastavené tak, aby čo najviac zohľadňovalo realitu, priority a požiadavky štatistickej praxe na Slovensku. Pokiaľ je pre indikátor kvality navrhnutých viac metód merania (a nie sú navrhnuté alternatívne),

tak sa indikátoru priradí hodnotenie za každú použitú metódu merania. Takto sa postupuje pri každom indikátore kvality a celkové hodnotenie pre AZÚ sa získa ako súčet hodnotení za všetky indikátory kvality.

Nie je nutné využiť na hodnotenie AZÚ všetky indikátory kvality. To znamená, že ľubovoľný indikátor kvality možno vynechať, či už z dôvodu nepotrebnosti, alebo z dôvodu neexistencie požadovaných údajov (môže nastať hlavne v prípade, keď poskytovateľ administratívnej databázy nedodá potrebné informácie na hodnotenie). Viaceré indikátory kvality sú bezpredmetné v prípade, že v štatistickom systéme neexistuje štatistické zisťovanie s rovnakým alebo podobným zameraním ako hodnotený AZÚ. Pokiaľ však hodnotený AZÚ vstupuje do štatistického systému s cieľom nahradiť, doplniť alebo kontrolovať niektoré existujúce štatistické zisťovanie, sú relevantné prakticky všetky indikátory kvality, teda aj tie, ktoré pri hodnotení berú do úvahy viaceré zdroje údajov.

V praxi sa na hodnotenie kvality AZÚ využíva automatizovaná verzia hodnotiacej schémy [8]. Ide o používateľský softvér vytvorený ako nadstavba nad MS Excel, v programovacom jazyku VBA (Visual Basic for Applications). Pri hodnotení sa do vyznačených polí jednoducho zadávajú požadované vstupy (odpovede na otázky alebo údaje na výpočet kvantitatívnych metód merania) alebo sa metóda merania označí ako nehodnotená. Tieto vstupné informácie majú podobu externých vstupov a pripravujú sa mimo hodnotiacej schémy. Príprava vstupných informácií vyžaduje dobrú znalosť hodnoteného súboru a v niektorých prípadoch treba na získanie požadovaných údajov použiť aj zložitejšie procedúry.

Na základe zadaných vstupov program priradí metóde merania kvalitatívnu alebo numerickú hodnotu a na jej základe priradí následne aj počet bodov indikátoru kvality. Po prejdení všetkých indikátorov kvality sa prechádza na záverečné zhrnutie, ktoré poskytuje informáciu o celkovom hodnotení AZÚ v percentách. Ide o podiel dosiahnutých bodov z celkového počtu možných bodov. Okrem celkového skóre zhrnutie obsahuje aj prehľadné hodnotenie jednotlivých indikátorov kvality v podobe podielu dosiahnutých bodov pre príslušný indikátor z maximálne možného počtu bodov pre príslušný indikátor. Do celkového hodnotenia sa započítajú len použité indikátory kvality, vynechané indikátory sa nezarátavajú a v hodnotení sú zobrazené ako nehodnotené.

5. ZÁVER

Hodnotiaca schéma kvality AZÚ je určená na hodnotenie AZÚ, ktoré vstupujú do štatistického systému. Umožňuje priamo, cielene a štandardizovane hodnotiť všetky AZÚ, ktoré chce ŠÚ SR využívať na štatistické účely. Napĺňa sa tým potreba hodnotiť kvalitu AZÚ pred vstupom do štatistického prostredia, pretože AZÚ nie sú primárne určené na využitie pre potreby štatistickej praxe.

Dôležitým aspektom hodnotenia kvality AZÚ je fakt, že v prípade niektorých AZÚ prebieha hodnotenie kvality opakovane a výsledky hodnotenia sa premietajú do komunikácie s jednotlivými správcami AZÚ s cieľom zlepšiť využitie AZÚ na štatistické účely.

Opísaný spôsob hodnotenia kvality AZÚ bol do značnej miery implementovaný aj na účely SODB 2021, a to tak v etape prípravy, ako aj v etape spracovania údajov.

Napomohol najmä pri konečnom rozhodnutí o kvalite AZÚ a jeho následnom využití na SODB 2021, čo sa premietlo do zákona o sčítaní. Ďalej napomohol pri určení, ktoré údaje sa budú zbierať a ktoré sa plne prevezmú z AZÚ. Pomohol tiež vymedziť, ktoré AZÚ, resp. premenné z AZÚ sa použijú na kontrolu premenných v SODB 2021.

LITERATÚRA

- [1] Cerroni, F., Di Bella, G., Galié, L.: Evaluating administrative data quality as input of the statistical production process. In: *Rivista di Statistica Ufficiale*, 2014, č. 1-2, s. 117-146.
- [2] Daas, P., Ossen, S., Vis-Visschers, R., Arends-Tóth, J.: Checklist for the quality evaluation of administrative data sources. The Hague/Heerlen, Statistics Netherlands, 2009, 18 s.
- [3] Eurostat. 2003. Quality assessment of administrative data for statistical purposes. Luxembourg. 22 s.
- [4] Eurostat. 2011. Kódex postupov pre európsku štatistiku. Luxembourg. 8 s.
- [5] Eurostat. 2014. ESS handbook for quality reports, 2014 edition. Luxembourg, Office for official publications of the European communities, 139 s. ISBN 978-92-79-45487-5.
- [6] Ivančíková, Ľ., Vaňo, B.: Teoreticko-metodologické aspekty hodnotenia administratívnych zdrojov údajov využívaných na štatistické účely. *Slovenská štatistika a demografia*, 2021, č.(3), s. 57-71.
- [7] Laitila, T., Wallgren, A., Wallgren, B.: Quality assessment of administrative data. Stockholm, Statistics Sweden, 2011. 33 s. ISSN 1653-7149.
- [8] Vaňo, B.: Národný rámec na hodnotenie kvality administratívnych údajov určených na štatistické účely. Bratislava, INFOSTAT, 2017. 39 s.
- [9] Zhang, L.: Topics of statistical theory for register-based statistics. *Statistica Neederlantica*, 2012, 66, č. 1, s. 41-66.

RESUMÉ

Administratívne údaje nie sú primárne určené na účely, na ktoré ich chce využívať štatistický úrad. Pred vstupom do štatistického prostredia je preto veľmi dôležité, hodnotiť ich kvalitu, t. j. vhodnosť využitia na štatistické účely. Štatistický úrad SR má na tento účel k dispozícii procedúru, ktorá umožňuje rýchlo, priamo a štandardizovane určiť kvalitu administratívnych zdrojov údajov.

Hodnotenie kvality administratívnych údajov v kontexte ich využívania na štatistické účely je komplexný a odborne náročný proces. Na hodnotenie kvality administratívnych údajov možno využiť celý rad kvalitatívnych aj kvantitatívnych indikátorov. Pokiaľ má byť hodnotenie objektívne, je potrebné optimalizovať výber jednotlivých indikátorov kvality na základe univerzálnych aj špecifických kritérií. Preto sa prostredníctvom indikátorov kvality vytvárajú tzv. hodnotiace schémy, ktoré sú nástrojom na komplexné hodnotenie administratívnych zdrojov údajov.

Z hierarchického hľadiska sú hodnotiace schémy usporiadané do hyperdimenzií a dimenzií. V každej dimenzii sa nachádzajú indikátory kvality (kvalitatívne alebo kvantitatívne) a ku každému indikátoru kvality prináleží najmenej jedna metóda merania.

Hodnotiacia schéma navrhnutá pre Štatistický úrad SR využíva na hodnotenie kvality administratívnych zdrojov údajov vstupujúcich do štatistického procesu, 3 hyperdimenzie, 14 dimenzií, 52 indikátorov kvality a 74 metód merania. Prvé dve hyperdimenzie (zdroj údajov a metaúdaje) hodnotia administratívny zdroj údajov ako celok a na hodnotenie využívajú väčšinou kvalitatívne metódy. Tretia hyperdimenzia

(údaje) je zameraná na hodnotenie jednotlivých údajov, ktoré sa nachádzajú v administratívnom zdroji údajov, pričom na hodnotenie sa využívajú skoro výlučne kvantitatívne metódy.

V praxi sa na hodnotenie kvality AZÚ využíva automatizovaná verzia hodnotiacej schémy. Ide o používateľský softvér, ktorý zjednodušuje a zrýchľuje hodnotenie a eliminuje formálne chyby.

RESUME

Administrative data are not primarily used for the purposes for which the Statistical office intends to use them. Therefore it is very important to evaluate their quality, i.e. suitability for statistical purposes, before entering the statistical environment. For this purpose, the Statistical Office of the SR has a procedure at its disposal, enabling fast, direct and standard determination of the quality of administrative data sources.

Evaluating the quality of administrative data in the context of their use for statistical purposes is a complex and professionally demanding process. A number of qualitative and quantitative indicators can be used to assess the quality of administrative data. If the evaluation is to be objective, it is necessary to optimally determine the selection of individual quality indicators on the basis of universal and specific criteria. Therefore, through quality indicators, the so-called evaluation schemes are created, serving as a tool for a comprehensive evaluation of administrative data sources.

From a hierarchical perspective, evaluation schemes are arranged in hyperdimensions and dimensions. Each dimension contains quality indicators (qualitative and quantitative) and there is at least one measurement method for each quality indicator. The evaluation scheme, designed for the Statistical Office of the SR, uses for the quality evaluation, administrative data sources entering the statistical process and it has 3 hyperdimensions, 14 dimensions, 52 quality indicators and 74 measurement methods. The first two hyperdimensions (Data Source and Metadata) evaluate the administrative data source as a whole and mostly use qualitative methods for evaluation. The third hyperdimension (Data) focuses on the evaluation of individual data contained in an administrative data source, using almost exclusively quantitative evaluation methods.

In practice, an automated version of the evaluation scheme is used for the quality assessment of the ADS. It is a user software simplifying and accelerating the evaluation and eliminating the formal errors.

PROFESIJNÝ ŽIVOTOPIS

Ing. Boris Vaňo vyštudoval Vysokú školu ekonomickú v Bratislave, následne absolvoval postgraduálne štúdium z demografie na Karlovej Univerzite v Prahe. Od roku 1980 pracuje v Inštitúte informatiky a štatistiky ako výskumný pracovník v oblasti demografie. V rokoch 2000 – 2014 bol vedúcim Výskumného demografického centra, v období rokov 2006 – 2010 podpredsedom Slovenskej štatistickej a demografickej spoločnosti pre demografiu. Špecializuje sa na hodnotenie populačného vývoja, demografické prognózy a populačnú politiku.

KONTAKT

vano@infostat.sk