

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

4/2026

ročník/volume 36

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

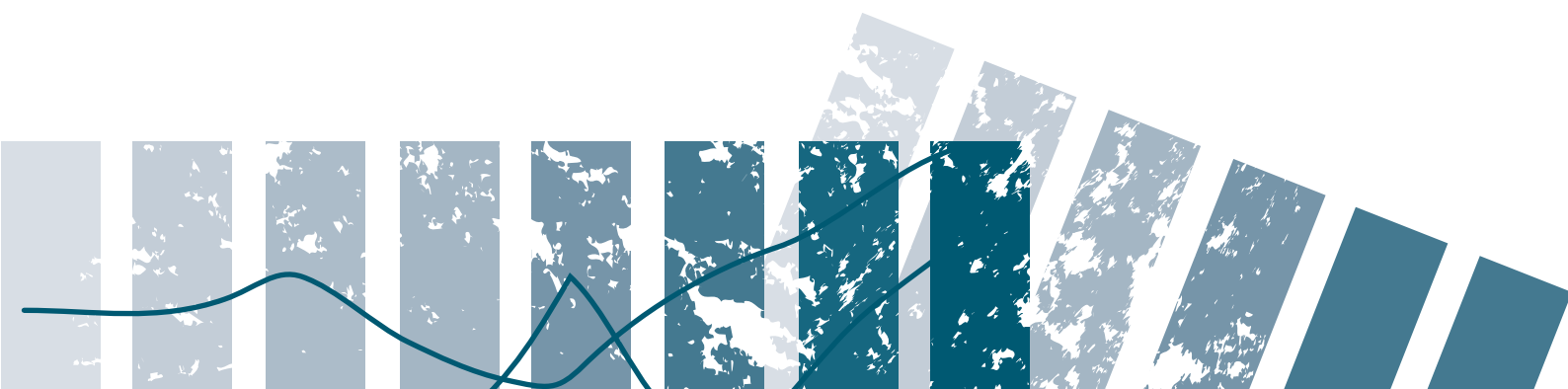
Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 2

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

Strany/Pages: 7 – 26

Dátum vydania/Publication date: 15. október 2026/October 15, 2026



Petra MAZUREKOVÁ, Helena GLASER-OPITZOVÁ
Štatistický úrad Slovenskej republiky

METODOLOGICKÉ ASPEKTY MANAŽÉRSTVA KVALITY PRI TRANSFORMÁCII NA NOVÉ ZDROJE ÚDAJOV V CENOVÝCH ŠTATISTIKÁCH

METHODOLOGICAL ASPECTS OF QUALITY MANAGEMENT IN THE TRANSITION TO NEW DATA SOURCES IN PRICE STATISTICS

ABSTRAKT

Produkcia cenových štatistík prechádza transformáciou v dôsledku využívania údajov zo skenerov obchodných reťazcov, online transakčných údajov a webovo získavaných údajov. Ich implementácia si vyžaduje nielen metodické zmeny pri spracovaní údajov a výpočte cenových indexov, ale aj vytvorenie komplexného systému manažérstva kvality. Cieľom príspevku je predstaviť metodologické aspekty a praktické skúsenosti Štatistického úradu Slovenskej republiky pri zavádzaní nových zdrojov údajov do produkcie indexov spotrebiteľských cien. Príspevok poukazuje na význam spolupráce s poskytovateľmi údajov, automatizácie spracovania, validačných postupov, odborného dohľadu a vhodného výberu metodických postupov na zabezpečenie kvality výsledných štatistických výstupov.

ABSTRACT

Price statistics are undergoing a major transformation as a result of the integration of scanner data from retail chains, online transaction data, and web-scraped data. Their implementation requires not only methodological changes in data processing and compilation of consumer price index but also the establishment of a comprehensive quality management system. This paper presents the methodological aspects and practical experience of the Statistical Office of the Slovak Republic in the integration of new data sources into the production of consumer price indices. It highlights the importance of cooperation with data providers, automated data processing, validation procedures, expert supervision, and the appropriate selection of methodological approaches to ensure the quality of official statistical outputs.

KLÚČOVÉ SLOVÁ

cenové štatistiky, údaje zo skenerov, online transakčné údaje, rámec kvality

KEYWORDS

price statistics, scanner data, online transaction data, quality assurance framework

1. ÚVOD

Rozvoj spolupráce medzi verejným a súkromným sektorom umožnil využívanie nových zdrojov údajov a tým modernizáciu cenových štatistík. Národné štatistické úrady dnes dokážu získavať údaje zo skenerov a online transakčné údaje priamo od komerčných subjektov. Tieto zdroje postupne nahrádzajú tradičné metódy zberu údajov, ako sú fyzické zisťovanie regálových cien. Medzi hlavné výhody nových dátových zdrojov patrí automatizácia procesov, zníženie nákladov na pracovnú silu, zvýšenie produktivity a predovšetkým zlepšenie ekonomickej presnosti cenových štatistík. Na druhej strane nové zdroje údajov prinášajú množstvo výziev, ktoré musia národné štatistické úrady vyriešiť pred ich využitím v oficiálnej štatistike. Medzi najvýznamnejšie výzvy patrí vytvorenie právneho rámca na získanie údajov,

zabezpečenie dôvernosti a bezpečnosti údajov, vybudovanie robustnej informačno-technologickej infraštruktúry, spracovanie veľkých objemov dát a výber najvhodnejšej metodiky spracovania a vzorcov.

Problematika využívania alternatívnych zdrojov údajov patrí v súčasnosti medzi hlavné témy rozvoja oficiálnej štatistiky. Medzinárodné organizácie, najmä Eurostat, Organizácia Spojených národov (OSN) a Organizácia pre hospodársku spoluprácu a rozvoj (OECD), podporujú ich postupnú implementáciu do produkcie cenových štatistík a venujú významnú pozornosť metodickým otázkam spracovania údajov, výberu indexových metód a zabezpečeniu kvality produkčných procesov. Skúsenosti jednotlivých národných štatistických úradov zároveň ukazujú, že úspešnosť implementácie závisí od schopnosti prepojiť metodologické inovácie s efektívnym riadením produkčného procesu.

Štatistický úrad Slovenskej republiky v posledných rokoch postupne implementoval údaje zo skenerov obchodných reťazcov do oficiálnej produkcie indexov spotrebiteľských cien za potraviny, nápoje a tabakové výrobky. Online transakčné údaje a webovo získavané údaje sa zatiaľ využívajú v rámci experimentálnej štatistiky, kde sa overujú možnosti ich spracovania, metodického využitia a potenciálneho uplatnenia pri výpočte cenových indexov. Celkovo tento proces zahŕňa spoluprácu s poskytovateľmi údajov, vývoj produkčného systému a interných nástrojov na automatizované spracovanie, ako aj návrh a testovanie metodických postupov.

Cieľom príspevku je predstaviť metodologické aspekty manažérstva kvality pri implementácii nových zdrojov údajov do produkcie indexov spotrebiteľských cien v Štatistickom úrade Slovenskej republiky. Príspevok opisuje jednotlivé etapy produkčného procesu od nadviazania spolupráce s poskytovateľmi údajov cez zber, validáciu, spracovanie a výpočet cenových indexov až po zabezpečenie kvality výsledných štatistických výstupov. Zároveň sumarizuje praktické skúsenosti získané počas implementácie a poukazuje na metodické riešenia, ktoré môžu byť využiteľné aj pri zavádzaní nových zdrojov údajov v ďalších oblastiach oficiálnej štatistiky.

2. PREHĽAD RELEVANTNEJ LITERATÚRY

Téma využívania nových zdrojov údajov v cenových štatistikách patrí v posledných rokoch medzi intenzívne rozvíjané oblasti oficiálnej štatistiky. Metodický rámec ich implementácie vytvárajú najmä príručky a odporúčania Eurostatu a ďalších medzinárodných organizácií. Harmonised Index of Consumer Prices (HICP): Methodological Manual (Eurostat, 2024) a Consumer Price Index Manual: Concepts and Methods (IMF et al., 2020) vymedzujú základné princípy zostavovania indexov spotrebiteľských cien, zatiaľ čo Practical Guide for Processing Supermarket Scanner Data (Eurostat, 2017) poskytuje odporúčania na spracovanie, validáciu, filtrovanie a klasifikáciu údajov zo skenerov. Osobitne problematike webovo získavaných údajov sa venujú Practical Guidelines on Web Scraping for the HICP Eurostat, 2020) a využitiu multilaterálnych indexových metód Guide on Multilateral Methods in the Harmonised Index of Consumer Prices (Eurostat, 2022).

Praktické skúsenosti národných štatistických úradov poukazujú na rozdielne prístupy k implementácii nových zdrojov údajov. Böttcher a Sergeev (2014) opisujú rakúsky projekt využitia skenerových údajov v cenovej štatistike. Skúsenosti Holandska reprezentujú práce Chessu (2026), Chessu, Verburga a Willenborga (2017)

a van der Grieta a de Haana (2010), ktoré sa venujú využívaniu transakčných údajov v holandskom CPI a porovnávaní vhodných indexových metód. Krsinich (2015) prezentuje implementáciu skenerových údajov pre spotrebnú elektroniku v CPI Nového Zélandu, teda v produktovej oblasti charakteristickej vysokou obmenou sortimentu. Problematiku širšieho využívania nových zdrojov údajov v spotrebiteľských cenových štatistikách rozoberá Greenhough (2019) na príklade britského Office for National Statistics. Práca poukazuje na možnosti využitia veľkoobjemových zdrojov, najmä skenerových a webovo získavaných údajov, ako aj na metodické a produkčné výzvy spojené s ich začlenením do cenovej štatistiky.

Významný prínos k problematike implementácie nových zdrojov údajov v podmienkach Slovenskej republiky predstavujú práce Glaser-Opitzovej (2019; 2022). V článku Nové zdroje údajov pre cenovú štatistiku a metódy ich spracovania (Glaser-Opitzová, 2019) autorka analyzuje metodické zmeny spojené s prechodom od tradičného zisťovania cien na využívanie skenerových údajov. V príspevku Use of Scanner Data in Measuring the Consumer Price Index in the Conditions of the Slovak Republic (Glaser-Opitzová, 2022) prezentuje skúsenosti Štatistického úradu Slovenskej republiky s implementáciou skenerových údajov do produkcie indexov spotrebiteľských cien. V článku od Glaser-Opitzovej a Mazurekovej (2023) je uvedená analýza vplyvu parciálneho časového pokrytia skenerových údajov na presnosť cenových indexov a preukazuje, že čiastočné pokrytie údajov môže pri vhodnom metodickom spracovaní poskytovať veľmi presné výsledky. Problematike webovo získavaných údajov sa venuje práca Knížata a Glaser-Opitzovej (2023), ktorá analyzuje možnosti využitia webscrapingu pri zostavovaní indexov spotrebiteľských cien a opisuje postup spracovania denných cien na mesačné cenové indexy. Na tieto výsledky nadväzuje článok Web scraped data in consumer price indices (Knížat, 2023), ktorý podrobnejšie rozoberá metodické otázky využitia webovo získavaných údajov v cenových indexoch a porovnáva alternatívne prístupy k agregácii cien. Vývoj metodiky multilaterálnych cenových indexov reprezentuje štúdia Knížata et al. (2024), ktorá predstavuje nový aditívny spôsob spájania multilaterálnych indexov a analyzuje jeho vlastnosti pri zostavovaní oficiálnych cenových štatistík. Na širší potenciál alternatívnych veľkoobjemových zdrojov údajov v oficiálnej štatistike poukazuje článok Knížata et al. (2025) venovaný nowcastingu indexu priemyselnej produkcie s využitím vysokofrekvenčných mýtnych údajov, a ilustruje možnosti ich využitia pri tvorbe a spresňovaní štatistických výstupov.

Uvedené publikácie dokumentujú postupný posun cenovej štatistiky od tradičného zberu cien k využívaniu veľkoobjemových údajov zo súkromných zdrojov. Zároveň ukazujú, že ich úspešná implementácia nie je iba otázkou výberu indexovej metódy, ale zahŕňa celý proces získavania, validácie, spracovania a kontroly kvality údajov až po ich začlenenie do pravidelnej produkcie oficiálnej štatistiky.

Prezentované publikácie vytvárajú teoretický a metodologický základ na zavádzanie nových zdrojov údajov do cenových štatistík.

3. ZÁKLADNÉ POJMY A METODOLOGICKÉ VÝCHODISKÁ

Modernizácia cenových štatistík je úzko spojená s využívaním nových zdrojov údajov, ktoré dopĺňajú alebo postupne nahrádzajú tradičný terénny zber cien. V porovnaní s klasickým zisťovaním poskytujú nové zdroje údajov výrazne podrobnejšie informácie o realizovaných nákupoch spotrebiteľov, umožňujú častejšiu

aktualizáciu spotrebného koša a lepšie zachytávajú dynamiku maloobchodného trhu. Ich implementácia však prináša nové metodologické výzvy súvisiace so spracovaním veľkoobjemových dát, výberom reprezentatívnych produktov a voľbou vhodných indexových vzorcov.

3.1. ÚDAJE ZO SKENEROV (SCANNER DATA)

Údaje zo skenerov (scanner data) predstavujú transakčné údaje zaznamenávané obchodnými reťazcami pri predaji tovarov prostredníctvom pokladničných systémov. Pri každom nákupe je načítaný čiarový kód výrobku, pričom sa zaznamenáva identifikácia produktu, predajná cena, predané množstvo, hodnota tržieb a čas uskutočnenia transakcie.

Na rozdiel od tradičného zisťovania cien, pri ktorom sa v predajni zisťuje cena vybraného reprezentanta spravidla raz mesačne, údaje zo skenerov zachytávajú všetky uskutočnené predaje daného produktu počas sledovaného obdobia. Vďaka tomu umožňujú zostaviť cenové indexy na základe skutočne realizovaných transakcií a výrazne rozšíriť počet výrobkov zahrnutých do výpočtu indexov spotrebiteľských cien.

Pri využití údajov zo skenerov sa zároveň mení cenový koncept. Do výpočtu nevstupuje jednorazovo zistená cena produktu, ale priemerná realizačná cena vypočítaná ako podiel celkových tržieb a predaného množstva:

priemerná cena = tržby/počet predaných kusov za dané časové obdobie.

Dôležitou vlastnosťou údajov zo skenerov je aj dostupnosť informácií o predaných množstvách, ktoré umožňujú využívať aj vážené indexové vzorce na elementárnej úrovni agregácie.

3.2. ONLINE TRANSAKČNÉ ÚDAJE

Online transakčné údaje predstavujú údaje o predaji produktov uskutočnenom prostredníctvom internetových obchodov. Na rozdiel od webovo získavaných údajov obsahujú informácie o skutočne realizovaných nákupoch vrátane predajných cien, predaných množstiev a tržieb.

Význam online transakčných údajov postupne rastie v dôsledku rozvoja elektronického obchodu. Poskytujú možnosť sledovať cenový vývoj produktov, ktoré sa predávajú výlučne prostredníctvom internetu alebo majú odlišnú cenovú politiku ako kamenné predajne. Ich využitie v oficiálnej štatistike si vyžaduje obdobné metodické postupy ako pri údajoch zo skenerov.

3.3. WEBOVO ZÍSKAVANÉ ÚDAJE (WEB-SCRAPED DATA)

Webovo získavané údaje (web-scraped data) vznikajú automatizovaným získavaním údajov z verejne dostupných internetových stránok pomocou softvérových nástrojov označovaných ako webscraping. Tieto údaje obsahujú predovšetkým ponukové ceny produktov, ich opis, technické charakteristiky a ďalšie informácie zverejnené internetovými obchodmi. Na rozdiel od údajov zo skenerov alebo online transakčných údajov neobsahujú informácie o skutočne predaných množstvách ani o tržbách.

Výhodou webscrapingu je možnosť veľmi častého získavania údajov a široké pokrytie internetového trhu. Nevýhodou je vyššia volatilita cien, častá obmena produktov a absencia výdavkových váh, čo kladie vyššie nároky na metodiku spracovania a výber vhodných indexových metód. Nezanedbateľná je veľká náročnosť na udržiavanie IT nástroja na získavanie takýchto údajov.

3.4. VYBRANÉ TYPY INDEXOV POUŽÍVANÉ NA VÝPOČET INFLÁCIE

Výber vhodného indexového vzorca patrí medzi najdôležitejšie metodologické rozhodnutia pri implementácii nových zdrojov údajov do cenových štatistík. Metodické odporúčania Eurostatu umožňujú na elementárnej úrovni agregácie použitie viacerých indexových vzorcov, pričom nestanovujú jeden univerzálny vzorec vhodný pre všetky produktové skupiny a typy údajov. Vhodnosť jednotlivých metód závisí od charakteru a spôsobu spracovávaných údajov, dynamiky sortimentu výrobkov, dostupnosti váh a požadovanej ekonomickej interpretácie výsledkov.

Na elementárnej úrovni agregácie¹ možno použiť široké spektrum indexových vzorcov. Vzhľadom na rozsah príspevku nie je možné venovať sa všetkým metódam, preto sa ďalšia pozornosť sústreďuje na troch vybraných zástupcov rozdielnych skupín indexov. Vo všeobecnosti možno rozlišovať bilaterálne indexové vzorce, ktoré porovnávajú ceny medzi dvoma časovými obdobiami, a multilaterálne indexové metódy, ktoré využívajú údaje z viacerých období súčasne a porovnávajú cenový vývoj v širšom časovom okne, spravidla za obdobie trinástich alebo dvadsiatich piatich mesiacov. Osobitnú skupinu bilaterálnych indexov predstavujú superlatívne indexové vzorce, ako sú Törnqvistov alebo Fisherov index, ktoré využívajú informácie o spotrebných váhach z dvoch porovnávaných období.

Jevonsov index

Jevonsov index patrí medzi najčastejšie používané nevážené bilaterálne indexy na elementárnej úrovni agregácie. Je definovaný ako geometrický priemer cenových zmien medzi dvoma po sebe nasledujúcimi obdobiami:

$$I_{Jevons}^{m-1,m} = \prod_{i \in N} \left(\frac{p_i^m}{p_i^{m-1}} \right)^{\frac{1}{N}}, \quad (1)$$

kde m je aktuálne obdobie, $m-1$ je predchádzajúce obdobie a p_i je cena i -tého produktu v danom období. Jeho hlavnou výhodou je metodická jednoduchosť, stabilita a splnenie viacerých axiomatických vlastností. Keďže nevyužíva informácie o predaných množstvách, je vhodný najmä pre produktové skupiny s relatívne stabilným sortimentom. V Štatistickom úrade Slovenskej republiky sa využíva pri výpočte indexov pre potraviny, nápoje a tabakové výrobky.

¹ Elementárna úroveň agregácie je najmenší a relatívne homogénny súbor tovarov alebo služieb, pre ktorý sa určuje indexové číslo bez explicitných výdavkových váh. Môže byť definovaný nielen na základe charakteristických vlastností tovarov a služieb, ale aj s ohľadom na región a typ odbytiska, v ktorom sa nachádzajú a predávajú.

Törnqvistov index

Törnqvistov index patrí medzi superlatívne indexy. Pri výpočte využíva informácie o cenách aj výdavkových váhach v oboch porovnávaných obdobiach, čím lepšie zohľadňuje zmeny spotrebiteľského správania. Törnqvistov index je definovaný takto:

$$I_{Törnqvist}^{m-1,m} = \prod_{i \in N} \left(\frac{p_i^m}{p_i^{m-1}} \right)^{\frac{s_i^{m-1} + s_i^m}{2}}, \quad (2)$$

kde $s_i^{m-1} = p_i^{m-1} q_i^{m-1} / \sum_{i \in N} p_i^{m-1} q_i^{m-1}$ a $s_i^m = p_i^m q_i^m / \sum_{i \in N} p_i^m q_i^m$ sú podiely výdavkov v období $m-1$ a m v danej skupine produktov, pričom q_i sú predané množstvá v danom mesiaci. V porovnaní s neváženými indexmi poskytuje ekonomicky interpretovateľnejšie výsledky, avšak pri reťazení² môže driftovať³.

Metóda GEKS

Metóda GEKS (Gini – Eltetö – Köves – Szulc) predstavuje multilaterálny prístup k výpočtu cenových indexov. Na rozdiel od bilaterálnych metód neporovnáva iba dve obdobia, ale využíva informácie z viacerých časových období súčasne. Hlavnou výhodou je eliminácia javu driftovania a lepšie zachytenie dynamiky sortimentu výrobkov.

Multilaterálna metóda GEKS je založená na súbore bilaterálnych cenových indexov vypočítaných medzi všetkými dvojicami období v rámci vopred definovaného časového okna. Tieto bilaterálne indexy predstavujú základné vstupy do výpočtu multilaterálneho indexu GEKS. Výsledná hodnota indexu preto závisí od voľby bilaterálneho indexového vzorca, ktorý tvorí základ výpočtu, pričom v praxi sa najčastejšie využívajú Jevonsov, Törnqvistov alebo Fisherov index.

Ako dĺžku okna budeme označovať W . V rámci okna sa vyberie základné obdobie označené ako l a vypočíta sa bilaterálny index medzi obdobím l a každým nasledujúcim obdobím v okne. Postup sa opakuje pri všetkých možných alternatívach l . Týmto spôsobom je možné získať maticu bilaterálnych indexov veľkosti $w \times w$ pre všetky možné dvojice jednotlivých období v rámci časového okna. V jednoduchosti možno vzorec zapísať takto:

$$I_{GEKS}^{0,m} = \prod_{l \in W} (I^{0,l} \times I^{l,m})^{\frac{1}{|W|}}, \quad (3)$$

kde $I^{l,m}$ je zvolený bilaterálny vzorec medzi obdobím l a m .

² Na elementárnej úrovni sa počítajú indexy, ktoré porovnávajú zmenu ceny medzi dvoma po sebe idúcimi obdobiami, avšak pre publikačnú prax v oblasti cenovej štatistiky sú dôležité indexy k základnému obdobiu 0, t. j. k decembru predchádzajúceho roku. Preto je potrebné medzimesačné indexy reťaziť, t. j. $I^{0,m} = I^{0,1} \cdot I^{1,2} \dots I^{m-1,m}$.

³ O reťazovom indexe sa hovorí, že driftuje, ak sa nevráti k jednotke, keď sa ceny v bežnom období vrátia na úroveň, ktorú dosiahli v základnom období. Reťazové indexy sú náchylné na drift, keď ceny počas období, ktoré pokrývajú, kolíšu.

Nevýhodou multilaterálnych metód je ich problematická ekonomická interpretovateľnosť.

Index Laspeyresovho typu

Index Laspeyresovho typu predstavuje základnú agregačnú metódu používanú pri zostavovaní indexov spotrebiteľských cien. To znamená, že slúži na výpočet indexov na vyšších úrovniach triedenia agregovaním indexov z nižších úrovní. Na vyšších úrovniach klasifikácie využíva fixné výdavkové váhy vzťahujúce sa na referenčné obdobie, spravidla predchádzajúci kalendárny rok. Tento prístup je základom výpočtu indexu spotrebiteľských cien (CPI) aj harmonizovaného indexu spotrebiteľských cien (HICP) v krajinách Európskej únie a zabezpečuje medzinárodnú porovnateľnosť výsledkov. Možno ho definovať takto:

$$p_A^{y,m/y-1} = \frac{\sum_{a \in A} w_a^{y-1} I_a^{y,m/y-1}}{\sum_{a \in A} w_a^{y-1}}, \quad (4)$$

kde w_a^{y-1} sú váhy založené na ročných výdavkoch za všetky položky patriace do elementárneho agregátu a bez ohľadu na to, či boli zahrnuté do výpočtu po aplikovaní filtrov.

4. PROCESNÝ TOK ÚDAJOV – ŠTATISTICKÝ ÚRAD SLOVENSKEJ REPUBLIKY

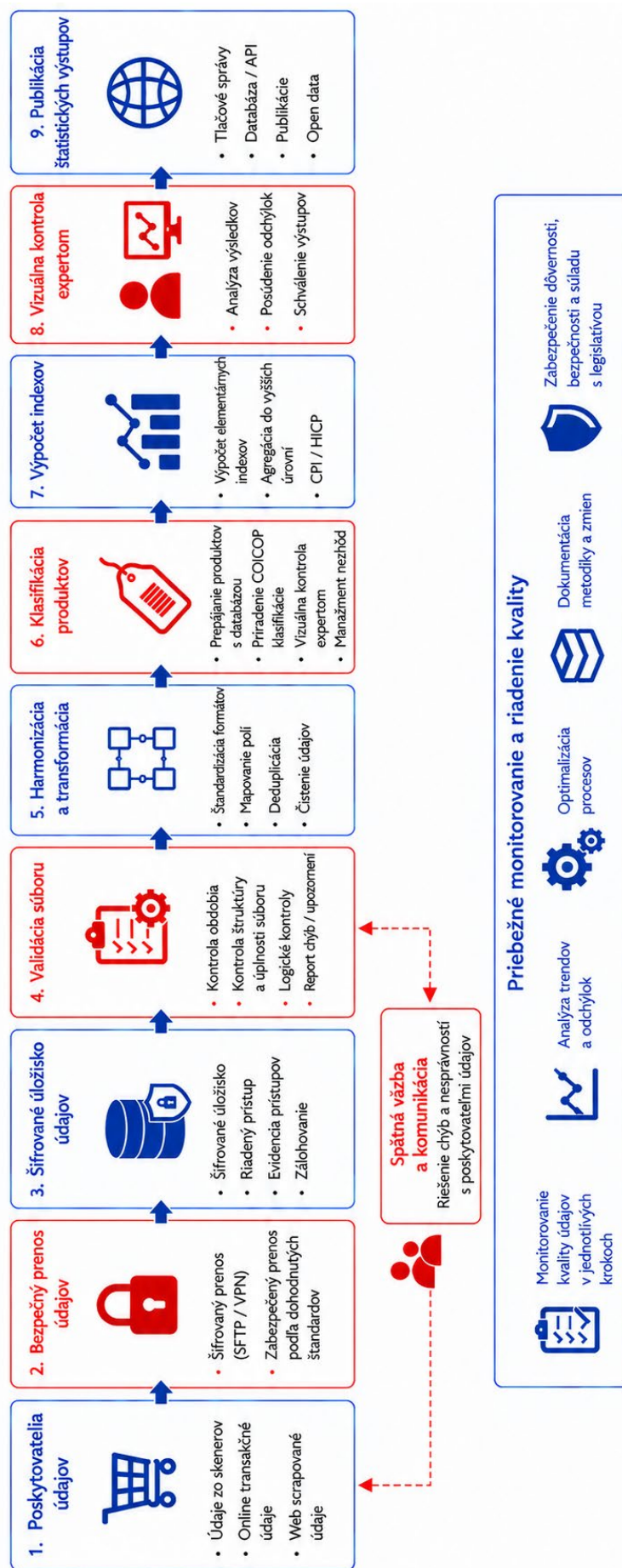
Integrácia nových zdrojov údajov do oficiálnej štatistiky predstavuje komplexný a náročný proces vyžadujúci vybudovanie robustnej infraštruktúry zahŕňajúcej prenos údajov od súkromných poskytovateľov, ich spracovanie a implementáciu nepretržitého zabezpečovania kvality údajov a štatistických produktov. Skúsenosti so zavádzaním údajov zo skenerov ukazujú, že dosiahnutie presnejších a aktuálnejších štatistických výstupov nie je jednorazovým technickým krokom, ale dlhodobým transformačným procesom pozostávajúcim z viacerých navzájom prepojených etáp.

Implementácia nových inovatívnych zdrojov údajov v Štatistickom úrade Slovenskej republiky bola sprevádzaná viacerými výskumnými a vývojovými aktivitami zameranými na transformáciu experimentálnych metodík na udržateľné produkčné procesy. Viaceré štúdie publikované v odborných časopisoch odrážajú praktické skúsenosti aj metodologické inovácie vyvinuté počas tohto prechodu (Glaser-Opitzová, 2019; Glaser-Opitzová & Mazureková, 2023; Knížat, 2023; Knížat & Glaser-Opitzová, 2023; Knížat et al., 2024). Vybrané výsledky týchto štúdií boli implementované do procesného toku údajov.

Na obrázku je znázornený procesný tok údajov demonštrujúci riešenie využívané v Štatistickom úrade Slovenskej republiky.

Obrázok: Procesný tok údajov

Schéma produkčného procesu spracovania údajov pre indexy spotrebiteľských cien



Zdroj: vlastné spracovanie autoriek na základe produkčného procesu Štatistického úradu Slovenskej republiky

4.1. SÚKROMNÍ POSKYTOVATLIA ÚDAJOV

Prvým predpokladom úspešného využívania údajov zo súkromného sektora je vytvorenie efektívnej spolupráce s ich poskytovateľmi. Keďže primárny zber údajov prebieha mimo tradičného štatistického systému, ich získavanie závisí od jasne definovaného právneho rámca, vzájomnej dôvery a transparentných pravidiel výmeny, odovzdávania a ochrany údajov. V podmienkach Štatistického úradu Slovenskej republiky bolo nevyhnutné nadviazať spoluprácu so zástupcami obchodných reťazcov a predstaviť im zámer využívania údajov zo skenerov obchodných reťazcov na štatistické účely. Cieľom úvodných rokovaní bolo oboznámiť poskytovateľov údajov s prínosmi ich využitia pri zostavovaní indexov spotrebiteľských cien, vysvetliť legislatívny rámec oficiálnej štatistiky a uistiť ich o zachovaní dôvernosti poskytovaných údajov. Obdobný postup bol uplatnený aj pri získavaní online transakčných údajov, keď boli oslovení prevádzkovatelia internetových „porovnávačov“ cien, ako aj významné subjekty pôsobiace na internetovom trhu v Slovenskej republike. Nie všetky tieto aktivity však viedli k nadviazaniu spolupráce a získaniu požadovaných údajov, čo poukazuje na význam komunikácie a ochoty súkromných subjektov spolupracovať pri zavádzaní nových dátových zdrojov do štatistickej praxe. V prípade webového získavania údajov boli prevádzkovatelia príslušných webových stránok písomne informovaní o zámere automatizovaného získavania verejne dostupných údajov v súlade s princípmi etického webscrapingu.

Na úvodné stretnutia nadviazali ďalšie odborné rokovania, počas ktorých bolo potrebné dohodnúť technické a metodické aspekty spolupráce. Predmetom diskusií bola najmä štruktúra odovzdávaných údajov, rozsah požadovaných premenných, frekvencia a spôsob ich zasielania, technické zabezpečenie prenosu údajov, ako aj pravidlá validácie a riešenia prípadných nezrovnalostí. Mimoriadna pozornosť bola venovaná ochrane poskytovaných údajov, keďže obsahujú obchodne citlivé informácie predstavujúce obchodné tajomstvo jednotlivých obchodných reťazcov. Národné štatistické úrady preto musia zabezpečiť úplný súlad s národnou aj európskou legislatívou v oblasti ochrany dôverných údajov a štatistickej dôvernosti, pričom zároveň garantujú, že získané údaje budú využité výlučne na štatistické účely a nebudú sprístupnené v podobe umožňujúcej identifikáciu jednotlivých poskytovateľov. Otvorená komunikácia, transparentnosť celého procesu a minimalizácia administratívnej záťaže poskytovateľov údajov sa ukázali ako kľúčové faktory úspešného nadviazania dlhodobej spolupráce.

4.2. SERVEROVÉ DATABÁZY ŠTATISTICKÉHO ÚRADU SLOVENSKEJ REPUBLIKY – UKLADANIE ÚDAJOV

Centralizovaná infraštruktúra na ukladanie údajov je nevyhnutná na správu veľkých objemov dát získavaných zo skenerových systémov a online transakcií. Štatistický úrad Slovenskej republiky implementoval špecializované serverové databázy na bezpečné uchovávanie surových aj spracovaných dátových súborov pri súčasnom zabezpečení dôvernosti údajov a prevádzkovej stability. Prostredie na ukladanie údajov podporuje efektívnu správu dát, spracovanie údajov a integráciu s ďalšími procesnými tokmi. Výskumné štúdie zdôrazňujú význam škálovateľnej databázovej štruktúry schopnej spracovávať neustále rastúce objemy údajov a obsluhovať viacerých poskytovateľov dát.

Nevyhnutnou súčasťou riešenia je zabezpečenie riadeného prístupu k dátam na základe pridelených oprávnení a evidencie prístupov k údajom v súlade s pravidlami informačnej bezpečnosti a štatistickej dôvernosti.

4.3. ZBER, VALIDÁCIA A SPRACOVANIE ÚDAJOV

Spracovanie nových zdrojov údajov zahŕňa ich validáciu, harmonizáciu, transformáciu, klasifikáciu, filtrovanie a prípravu na výpočet cenových indexov. Skenerové, online transakčné a webovo získavané údaje sa pritom líšia štruktúrou, rozsahom aj kvalitou opisov produktov, čo si vyžaduje štandardizované a vo vysokej miere automatizované postupy spracovania, avšak s prihliadnutím na rozdielnosť zdrojov.

Prvým krokom je automatizovaná validácia prijatých súborov. Kontroluje sa najmä termín doručenia a referenčné obdobie, úplnosť súborov a premenných, dátová štruktúra, počet záznamov a základná logická konzistentnosť údajov. V prípade identifikovaných nedostatkov nasleduje komunikácia s poskytovateľom údajov s cieľom objasniť ich príčinu a zabezpečiť opravu alebo doplnenie údajov. Takto nastavený mechanizmus spätnej väzby prispieva k stabilite produkčného procesu a postupnému zvyšovaniu kvality vstupných údajov.

Jedným z kľúčových krokov ďalšieho spracovania je klasifikácia produktov. Jednotlivé produkty je potrebné zaradiť do príslušných tried klasifikácie individuálnej spotreby podľa účelu (COICOP). Pre potreby výpočtu cenových indexov z údajov zo skenerov vytvoril Štatistický úrad Slovenskej republiky aj podrobnejšiu národnú úroveň členenia produktov do homogénnych skupín, ktoré predstavujú elementárnu úroveň agregácie. V počiatočnej fáze implementácie boli produkty klasifikované manuálne, čo bolo vzhľadom na veľký počet položiek časovo a personálne náročné. Postupne preto boli vyvinuté modely strojového učenia umožňujúce automatizovať proces klasifikácie. Pri údajoch zo skenerov dosahuje úspešnosť automatizovanej klasifikácie približne 70 – 80 % v závislosti od obchodného reťazca. Úspešnosť závisí najmä od rozsahu a kvality dostupných popisov produktov, ako aj od podrobnosti internej klasifikácie. Nové produkty klasifikované pomocou modelov strojového učenia preto následne podliehajú expertnej kontrole a v prípade potreby sa ich zaradenie do COICOP alebo príslušnej homogénnej skupiny upraví. Kombinácia automatizovanej klasifikácie a odborného posúdenia tak umožňuje znižovať manuálnu náročnosť procesu pri zachovaní požadovanej kvality klasifikácie.

Modely strojového učenia boli vyvinuté aj pre online transakčné a webovo získavané údaje, pričom ich ďalšie zdokonaľovanie naďalej prebieha. Klasifikácia a jej následné overovanie sú pri týchto zdrojoch náročnejšie najmä z dôvodu výrazne vyššej obmeny sortimentu a väčšieho počtu nových produktov medzi jednotlivými obdobiami v porovnaní s údajmi zo skenerov. Kvalita automatizovanej klasifikácie preto predstavuje jednu z oblastí ďalšieho metodického a technického vývoja.

Automatizácia spracovania si vyžaduje komplexné aplikačné riešenie, ktoré zabezpečuje riadenie jednotlivých krokov od prevzatia a validácie údajov až po výpočet cenových indexov a tvorbu výstupov. V Štatistickom úrade Slovenskej republiky bola na tento účel vyvinutá interná aplikácia zabezpečujúca automatizované spracovanie údajov zo skenerov. Online transakčné a webovo získavané údaje, ktoré

sa v súčasnosti využívajú v rámci experimentálnej štatistiky, sú spracovávané poloautomatizovaným spôsobom v softvérovom prostredí SAS.

Napriek vysokej miere automatizácie zostáva nevyhnutnou súčasťou procesu odborný dohľad. Experti priebežne monitorujú výsledky spracovania, vykonávajú vizuálne kontroly, analyzujú neštandardné cenové výkyvy a overujú výsledky automatizovanej klasifikácie a validačných postupov. Automatizácia zároveň nepredstavuje jednorazovú aktivitu. Produkčný systém sa priebežne rozvíja v súvislosti so zapájaním nových poskytovateľov a produktových skupín, zmenami metodických požiadaviek, zavádzaním nových kontrol a zdokonaľovaním klasifikačných a výpočtových algoritmov.

Význam kvalitného spracovania údajov potvrdzujú aj publikované štúdie. Glaser-Opitzová (2019) poukazuje na to, že využitie údajov zo skenerov mení tradičný prístup k tvorbe spotrebného koša a umožňuje pracovať s dynamickým súborom skutočne predávaných produktov, čím možno pružnejšie reagovať na zmeny sortimentu a spotrebiteľského správania. Pri webovo získavaných údajoch Knížat a Glaser-Opitzová (2023) analyzujú agregáciu vysokofrekvenčných cien na mesačnú úroveň a poukazujú na význam spôsobu agregácie pre výsledné hodnoty cenových indexov.

4.4. KRITÉRIÁ DEFINOVANÉ POUŽÍVATEĽOM A ICH IMPLEMENTÁCIA

Dôležitou súčasťou produkčného procesu je možnosť aplikovať metodologické kritériá definované používateľom počas spracovania a analýzy údajov. Táto flexibilita umožňuje prispôbovať jednotlivé spracovateľské postupy špecifikám produktových skupín a umožňuje, aby výsledné cenové indexy čo najvernejšie odrážali reálny vývoj spotrebiteľských cien. Medzi parametre definované používateľom patria najmä prahové hodnoty na výber produktov, pravidlá filtrovania údajov, spôsob identifikácie a ošetrenia odľahlých hodnôt, metódy agregácie údajov, ako aj validačné pravidlá uplatňované počas spracovania.

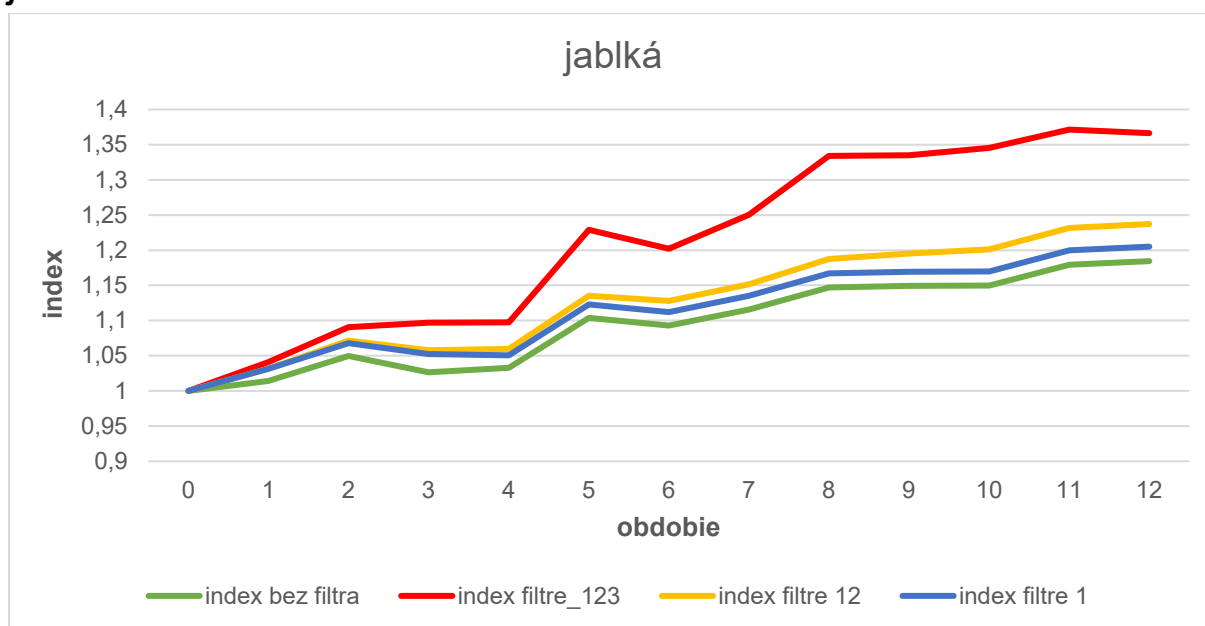
Možnosť definovať jednotlivé parametre však neznamena, že ich možno počas produkcie ľubovoľne meniť. Naopak, jedným zo základných princípov oficiálnej štatistiky je zabezpečenie metodologickej stability a kontinuity produkčného procesu, ktorá je nevyhnutným predpokladom porovnateľnosti výsledkov v čase. Zmeny nastavenia parametrov môžu ovplyvniť hodnoty vypočítaných indexov, a preto musia byť podložené odbornými analýzami, dôkladne zdokumentované a implementované až po overení ich vplyvu na výsledné štatistické ukazovatele.

Pred zavedením spracovania údajov zo skenerov do produkcie prebiehal v Štatistickom úrade Slovenskej republiky rozsiahly metodický výskum zameraný na identifikáciu optimálnych hodnôt jednotlivých parametrov. V rámci testovacích výpočtov boli analyzované rôzne kombinácie prahových hodnôt, pravidiel filtrovania a validačných kritérií s cieľom posúdiť ich vplyv na stabilitu a reprezentatívnosť vypočítaných cenových indexov. Výsledky testovania umožnili nastaviť parametre tak, aby sa minimalizoval vplyv extrémnych cien, jednorazových cenových výkyvov a dopredajových produktov, ktoré by mohli neprimerane ovplyvniť výsledné indexy.

Osobitná pozornosť bola venovaná výberu produktov vstupujúcich do výpočtu cenových indexov. Rozhodnutie aplikovať pri spracovaní údajov zo skenerov filtračné pravidlá vychádzalo z metodických odporúčaní Eurostatu uvedených v Practical Guide

for Processing Supermarket Scanner Data (Eurostat, 2017). Cieľom bolo zabezpečiť, aby výsledné indexy odrážali predovšetkým cenový vývoj výrobkov, ktoré tvoria rozhodujúcu časť spotrebiteľských výdavkov domácností. Preto boli metodické pravidlá navrhnuté tak, aby mali dominantný vplyv na výpočet produkty s vysokým objemom predaja, zatiaľ čo výrobky s minimálnym predajom alebo krátkodoboponúkané produkty neovplyvňovali výsledné indexy neprimeraným spôsobom.

Graf 1: Vplyv jednotlivých filtrov na hodnoty cenového indexu – elementárna skupina jablká



Zdroj údajov: Štatistický úrad Slovenskej republiky, výpočty autoriek

Graf znázorňuje vplyv postupného uplatňovania filtrov pri spracovaní skenerových údajov na hodnoty mesačných cenových indexov pre produktovú skupinu jablká. Zelená krivka predstavuje index vypočítaný bez aplikácie filtrov, zatiaľ čo ďalšie krivky zobrazujú výsledky po postupnom a súbežnom zavedení jednotlivých filtrov. Filter 1 predstavuje odstránenie extrémnych cenových hodnôt, filter 2 eliminuje dopredajové produkty a filter 3 odstraňuje produkty s nízkym objemom predaja. Z výsledkov je zrejmé, že aplikácia filtrov ovplyvňuje priebeh časového radu cenového indexu. Najvýraznejší vplyv na hodnoty indexu má odstránenie málo predávaných produktov.

Samotné odporúčanie Eurostatu aplikovať filtre však neurčuje automaticky optimálne nastavenie ich parametrov na konkrétne podmienky produkcie. Úlohou testovania a analýz realizovaných v Štatistickom úrade Slovenskej republiky preto bolo stanoviť vhodné prahové hodnoty jednotlivých filtrov a posúdiť ich vplyv na výsledné cenové indexy. Testovali sa rôzne hodnoty parametrov s cieľom eliminovať nereprezentatívne cenové pozorovania a zároveň zachovať dostatočný rozsah údajov a reprezentatívnosť výpočtu. Pri údajoch zo skenerov a online transakčných údajoch sa do výpočtu vyberajú produkty tak, aby pokrývali približne 80 % tržieb v každej skupine COICOP, čím sa zabezpečuje významné zastúpenie skutočne predávaných produktov. Pri webovo získavaných údajoch informácie o predaných množstvách a tržbách nie sú k dispozícii, preto nie je možné rovnakým spôsobom posúdiť predajný význam jednotlivých produktov. V tomto prípade sa pri filtrovaní vychádza najmä

z vývoja pozorovaných cien, pričom produkty s výraznými cenovými výkyvmi môžu byť na základe stanovených pravidiel z výpočtu vylúčené.

Pri nastavovaní metodických parametrov je zároveň potrebné zohľadniť charakter jednotlivých produktových skupín. Spotrebné správanie, frekvencia obmeny sortimentu aj cenová dynamika sa výrazne líšia medzi potravinami, elektronikou, odevmi alebo drogeriovým tovarom. Zatiaľ čo pri potravinách je charakteristická vysoká frekvencia predaja a časté cenové zmeny, pri elektronike alebo odevoch zohrávajú významnejšiu úlohu sezónnosť, uvádzanie nových modelov či dopredaje. Z uvedeného dôvodu nie je vhodné uplatňovať jednotné prahové hodnoty na všetky produktové skupiny, ale je potrebné ich prispôsobiť charakteru spracovávaných údajov. V prípade online transakčných údajov, ako aj webovo získavaných údajov preto naďalej prebiehajú testy a analýzy, keďže tieto údaje obsahujú viacero druhov tovarov, ako napr. elektroniku, oblečenie, nábytok...

Uvedené skutočnosti potvrdzujú, že návrhu produkčného systému musí predchádzať rozsiahle testovanie, experimentálne výpočty a metodické analýzy. Až na základe ich výsledkov je možné stanoviť také parametre spracovania, ktoré zabezpečia stabilitu produkčného procesu, vysokú kvalitu výsledných cenových indexov a ich dlhodobú porovnateľnosť. Súčasne je potrebné tieto parametre pravidelne prehodnocovať s ohľadom na vývoj maloobchodného trhu, rozširovanie spracovania o nové produktové skupiny a skúsenosti získané z produkčnej prevádzky.

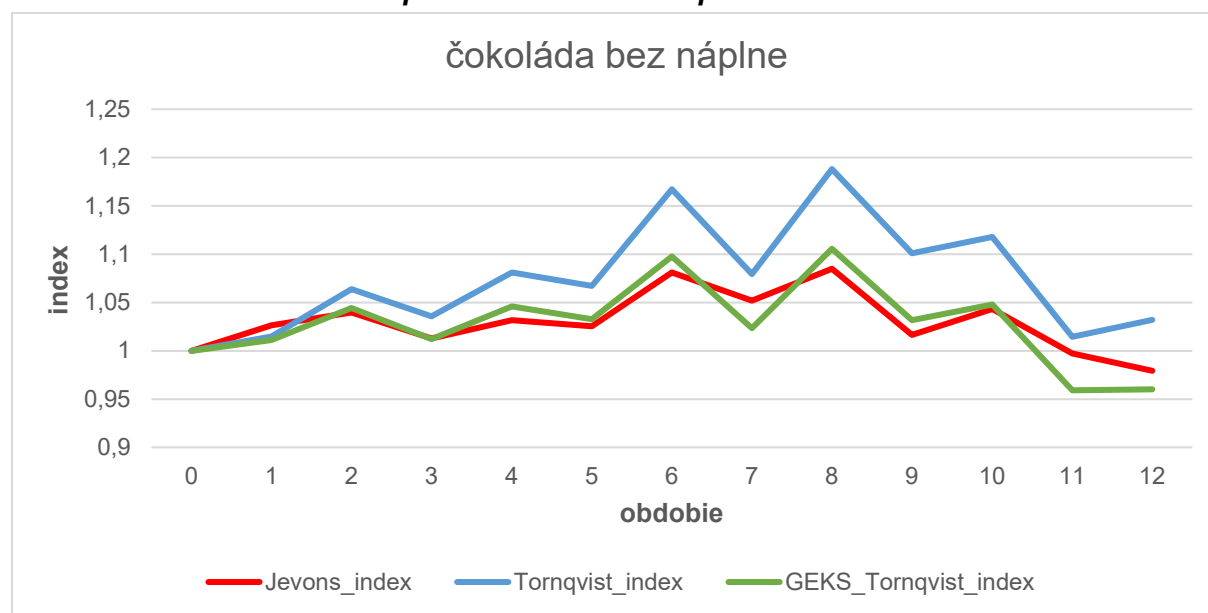
4.5. VÝBER VHODNÉHO CENOVÉHO INDEXU

Jedným z hlavných výstupov opísaného pracovného postupu je zostavovanie cenových indexov založených na údajoch zo skenerov obchodných reťazcov a online transakčných údajoch. Výber najvhodnejšieho indexového vzorca nie je univerzálny a musí vychádzať z viacerých kritérií. Pri rozhodovaní je potrebné zohľadniť výsledky empirických testov, odporúčania medzinárodných organizácií a skúseností iných krajín, ekonomickú interpretovateľnosť indexov, ich stabilitu v čase, schopnosť reagovať na zmeny sortimentu výrobkov, ako aj výpočtovú náročnosť jednotlivých metód. Zároveň je potrebné zabezpečiť, aby zvolený metodický prístup spĺňal požiadavky na reprodukovateľnosť výsledkov a umožňoval dlhodobú porovnateľnosť cenových indexov.

Skúsenosti získané pri implementácii údajov zo skenerov ukázali, že výber indexového vzorca závisí od charakteru produktovej skupiny a spôsobu výberu produktov vstupujúcich do výpočtu. Z axiomatického hľadiska vykazuje Jevonsov index veľmi dobré vlastnosti a patrí medzi všeobecne odporúčané elementárne indexy. Törnqvistov index síce využíva informácie o výdavkových podieloch, pri dynamickom koši a potrebe reťazenia však môže dochádzať k reťazovému driftu. Multilaterálne metódy umožňujú lepšie zohľadniť dynamiku sortimentu, ich výpočet a ekonomická interpretácia sú však zložitejšie. Na základe metodických odporúčaní a výsledkov testovania preto Štatistický úrad Slovenskej republiky zvolil na produkčný výpočet indexov za potraviny, nápoje a tabakové výrobky Jevonsov index, najmä pre jeho vhodné axiomatické vlastnosti, jednoduchú ekonomickú interpretáciu, stabilitu výsledkov, jednoduchú implementáciu a absenciu problému reťazového driftu. Pri online transakčných a webovo získavaných údajoch testovanie vhodných indexových metód naďalej pokračuje. Osobitnú pozornosť si pritom zasluhujú multilaterálne indexy, ktoré umožňujú pracovať s informáciami z viacerých období a sú

vhodné najmä pre produktové skupiny s vyššou mierou obmeny sortimentu a niektoré typy môžu zohľadniť aj vlastnosti produktov.

Graf 2: Porovnanie vývoja cenových indexov vypočítaných rôznymi indexovými vzorcami – elementárna skupina čokoláda bez náplne



Zdroj údajov: Štatistický úrad Slovenskej republiky, výpočty autoriek

Graf porovnáva vývoj mesačných cenových indexov pre elementárnu skupinu čokoláda, ktoré boli vypočítané použitím troch rôznych typov vzorcov – Jevonsovho indexu, Törnqvistovho indexu a GEKS-Törnqvistovho indexu. Cieľom porovnania je poukázať na vplyv výberu indexového vzorca na výsledné hodnoty cenových indexov pri spracovaní údajov zo skenerov obchodných reťazcov.

Význam výberu vhodnej indexovej metódy potvrdzujú aj vedecké štúdie Chessa (2026) a Greenhaougha (2019). Autori poukazujú na to, že pri niektorých druhoch tovarov predávaných prostredníctvom internetu dochádza k veľmi častej obmene sortimentu. Nové výrobky priebežne vstupujú na trh, staršie výrobky z neho miznú alebo sa po určitom čase opätovne objavujú v inovovaných verziách. Takéto prostredie predstavuje významnú metodologickú výzvu pre tradičné bilaterálne indexové vzorce, ktoré predpokladajú existenciu identických produktov v porovnávaných obdobiach. V empirickej časti štúdie autori porovnávajú viacero indexových metód na produktovej skupine chladničky a preukazujú, že výber indexového vzorca môže významne ovplyvniť hodnotu aj časový vývoj cenového indexu. Výsledky zároveň potvrdzujú, že multilaterálne indexové metódy dokážu lepšie zachytiť cenovú dynamiku pri produktoch s častou obmenou sortimentu a predstavujú vhodné riešenie na spracovanie moderných transakčných údajov.

4.6. PRESNOSŤ A INTEGRITA

Konečným cieľom integrácie súkromných zdrojov údajov do produkcie oficiálnej štatistiky je zvýšenie presnosti, reprezentatívnosti a aktuálnosti štatistických výstupov pri súčasnom zachovaní ich metodologickej konzistentnosti. Údaje zo skenerov, ako aj online transakčné údaje od obchodných reťazcov poskytujú podrobné informácie o realizovaných nákupných transakciách vrátane predajných cien, predaných množstiev a obratu jednotlivých výrobkov. V porovnaní s tradičným zisťovaním cien

umožňujú podstatne lepšie zachytiť skutočné spotrebiteľské správanie a pružnejšie reagovať na zmeny sortimentu, sezónnosť predaja či cenové akcie.

Zavedenie nových zdrojov údajov však samo osebe nezaručuje vyššiu kvalitu štatistických výstupov. Nevyhnutným predpokladom ich úspešného využitia je vytvorenie systému priebežného monitorovania kvality údajov počas celého produkčného procesu. Národné štatistické úrady musia pravidelne hodnotiť úplnosť a konzistentnosť vstupných údajov, reprezentatívnosť spracovávaných dát, stabilitu použitých metodických postupov a porovnateľnosť výsledných cenových indexov v čase. Dôležitou súčasťou manažérstva kvality je aj transparentné dokumentovanie metodických postupov, validačných pravidiel a zmien v produkčnom procese, čím sa posilňuje dôveryhodnosť a reprodukovateľnosť oficiálnych štatistických výstupov.

Vo vedeckej štúdii autorky Glaser-Opitzová a Mazureková (2023) analyzujú vplyv časového rozsahu dostupných transakčných údajov na presnosť výpočtu cenových indexov. Autorky porovnali výsledky výpočtu indexov spotrebiteľských cien založených na databázach s úplným a parciálnym časovým pokrytím údajov za produktovú skupinu potraviny a nealkoholické nápoje. Výsledky empirickej analýzy preukázali, že priemerný absolútny percentuálny rozdiel medzi hodnotami cenových indexov vypočítaných z databázy s parciálnym časovým pokrytím a databázy s úplným časovým pokrytím bol vo všetkých analyzovaných obdobiach nižší ako 1 %. Zistenia naznačujú, že vhodne navrhnutá metodika výberu a spracovania údajov dokáže zabezpečiť vysokú presnosť výsledných indexov aj v prípadoch, keď nie je možné využiť transakčné údaje za celé sledované obdobie.

4.7. PRECHOD OD EXPERIMENTÁLNEJ ŠTATISTIKY K OFICIÁLNEJ ŠTATISTIKE

Poslednou etapou produkčného procesu je rozhodnutie o zavedení nového metodického postupu do oficiálnej štatistiky. Ak má nový zdroj nahradiť tradičný spôsob zberu cien, jedným z dôležitých predpokladov je jeho dostatočné pokrytie príslušného segmentu trhu. Napríklad v prípade údajov zo skenerov využívaných Štatistickým úradom Slovenskej republiky pri výpočte indexov za potraviny a nealkoholické nápoje predstavuje pokrytie približne 80 % tržieb v danej oblasti. Pri hodnotení reprezentatívnosti je však potrebné zohľadniť nielen podiel pokrytého trhu, ale aj rozsah produktov zahrnutých do nového zdroja údajov. V tomto smere existuje významný rozdiel oproti tradičnému zberu cien. Pri tradičnom zbere bolo možné sledovať iba vybranú vzorku reprezentatívnych produktov, keďže manuálne zisťovanie cien všetkých produktov dostupných v jednotlivých predajniach by bolo personálne a časovo neuskutočniteľné. Údaje zo skenerov, naopak, poskytujú informácie o všetkých predaných produktoch zapojených obchodných reťazcov na území Slovenskej republiky. Obdobne online transakčné údaje umožňujú pracovať s informáciami o celom rozsahu realizovaných predajov príslušného internetového predajcu. Nové zdroje tak prinášajú nielen širšie pokrytie produktov, ale aj podrobnejšie informácie o skutočne realizovaných predajoch, cenách a množstvách.

Pred samotnou implementáciou je nevyhnutné overiť metodickú správnosť navrhovaného riešenia, zdokumentovať produkčný proces a vyhodnotiť vplyv metodickej zmeny na výsledné hodnoty cenových indexov. Súčasťou overovania je aj spätný prepočet indexov podľa novej metodiky minimálne za obdobie jedného roka. Takto vypočítané indexy sa porovnávajú s už publikovanými hodnotami založenými na

údajoch získaných tradičným zberom cien. Porovnanie umožňuje posúdiť rozdiely medzi pôvodným a novým postupom, stabilitu výsledkov a kvantifikovať vplyv metodologickej zmeny pred jej zavedením do produkcie.

Pri harmonizovanom indexe spotrebiteľských cien (HICP) je potrebné o plánovanej metodologickej zmene vopred informovať EUROSTAT. Súčasťou procesu je vypracovanie metodologickej dokumentácie, kvantifikácia vplyvu navrhovanej zmeny na hodnoty indexov a preukázanie, že nový postup spĺňa požiadavky európskej legislatívy a metodických usmernení. Dôležitá je zároveň transparentná komunikácia s používateľmi štatistických výstupov. Pri implementácii údajov zo skenerov Štatistický úrad Slovenskej republiky informoval používateľov o metodologickej zmene a na svojom webovom sídle, ako aj pri zverejňovaných údajoch sprístupnil informáciu o zmene spolu so stručným opisom novej metodiky (Štatistický úrad Slovenskej republiky, 2026). Až po úspešnom ukončení testovania a schválení metodologickej zmeny môže byť nový produkčný proces zavedený do pravidelnej produkcie oficiálnej štatistiky.

5. ZÁVER

Využívanie údajov zo skenerov obchodných reťazcov, online transakčných údajov a webovo získavaných údajov predstavuje jeden z najvýznamnejších krokov v modernizácii cenových štatistík za posledné desaťročia. Nové zdroje údajov umožňujú presnejšie zachytiť spotrebiteľské správanie, rozširujú pokrytie sledovaných produktov a vytvárajú predpoklady na vyššiu mieru automatizácie produkčného procesu. Súčasne však prinášajú nové metodologické, technické a organizačné výzvy, ktoré si vyžadujú zásadnú transformáciu existujúcich procesov tvorby oficiálnych štatistík.

Príspevok predstavil metodologické aspekty manažérstva kvality pri implementácii nových zdrojov údajov v Štatistickom úrade Slovenskej republiky. Pozornosť bola venovaná celému produkčnému procesu – od nadviazania spolupráce s poskytovateľmi údajov, vytvorenia právneho a technického rámca ich poskytovania cez bezpečný prenos, validáciu, spracovanie a automatizáciu údajov až po výpočet cenových indexov a zabezpečenie kvality výsledných štatistických výstupov. Osobitná pozornosť bola venovaná výberu vhodných metodických postupov, používateľom definovaných kritérií a indexových metód, ktoré významne ovplyvňujú kvalitu a reprezentatívnosť výsledných indexov spotrebiteľských cien.

Skúsenosti získané počas implementácie potvrdzujú, že úspešné zavedenie nových zdrojov údajov nezávisí iba od ich dostupnosti alebo od vybudovania technickej infraštruktúry. Rozhodujúcim predpokladom je vytvorenie komplexného systému manažérstva kvality, ktorý zabezpečuje spoľahlivosť údajov počas celého produkčného procesu. Kľúčovými prvkami takéhoto systému sú dlhodobá spolupráca s poskytovateľmi údajov, automatizované validačné mechanizmy, bezpečné spracovanie obchodne citlivých údajov, priebežné monitorovanie kvality, odborný dohľad expertov a dôsledná dokumentácia metodických postupov.

Významným poznatkom je aj skutočnosť, že automatizácia produkčného procesu nepredstavuje jednorazové riešenie, ale kontinuálny proces vývoja. Produkčné aplikácie sa musia priebežne rozširovať o nové produktové skupiny, ďalších poskytovateľov údajov, nové validačné pravidlá a metodické požiadavky vyplývajúce z vývoja medzinárodných odporúčaní aj praktických skúseností z produkcie.

Zachovanie vysokej úrovne kvality preto vyžaduje neustálu spoluprácu odborníkov na cenovú štatistiku, informačné technológie a spracovanie údajov.

Výsledky prezentované v článku zároveň potvrdzujú, že vhodne nastavené procesy validácie, filtrovania údajov a výberu indexových metód významne prispievajú k stabilite a ekonomickej interpretovateľnosti výsledných cenových indexov. Výber metodických postupov musí vychádzať z dôsledných analýz a empirického testovania, pričom univerzálne riešenie vhodné pre všetky produktové skupiny neexistuje. Preto je nevyhnutné metodické postupy priebežne overovať a prispôbovať charakteru jednotlivých zdrojov údajov a sledovaných produktov.

Možno konštatovať, že implementácia nových zdrojov údajov predstavuje nielen technologickú inováciu, ale predovšetkým komplexnú transformáciu produkcie cenových štatistik. Jej úspech závisí od schopnosti prepojiť moderné informačné technológie, metodologické inovácie a princípy manažérstva kvality do jednotného produkčného systému. Skúsenosti Štatistického úradu Slovenskej republiky ukazujú, že takýto prístup umožňuje vytvoriť udržateľné a efektívne produkčné prostredie, ktoré je pripravené reagovať na ďalší rozvoj alternatívnych zdrojov údajov a budúce požiadavky na modernizáciu oficiálnej štatistiky. Modernizácia oficiálnej štatistiky preto nie je jednorazovým projektom, ale dlhodobým procesom kontinuálneho zlepšovania, ktorý musí pružne reagovať na meniace sa podmienky a nové požiadavky používateľov štatistických informácií.

So zámerom vylúčiť konflikt záujmov Helena Glaser-Opitzová ako spoluautorka článku neparticipovala na per rollam komunikácii a rokovaníach redakčnej rady časopisu Slovenská štatistika a demografia vo veciach týkajúcich sa tohto článku.

LITERATÚRA

- Böttcher, I., & Sergeev, S. (2014). *Austrian scanner data project – Report: Multipurpose consumer price statistics: The use of scanner data* [Paper presentation]. ECE/ILO Meeting of the Group of Experts on Consumer Price Indices, Geneva, Switzerland.
https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2014/WS4/WS4_08_Multipurpose_Consumer_Price_Statistics_The_use_of_scanner_data.pdf
- Chessa, A. G. (2026). *Index methodology for transaction data in the Dutch CPI* [Discussion Paper]. Statistics Netherlands.
https://www.cbs.nl/en-gb/background/2026/06/index-methodology-for-transaction-data-in-the-dutch-cpi?utm_source=chatgpt.com
- Chessa, A. G., Verburg, J., & Willenborg, L. (2017). *A comparison of price index methods for scanner data* [Paper presentation]. 15th Meeting of the Ottawa Group, Eltville, Germany. <https://stats.unece.org/ottawagroup/download/f460.pdf>
- Eurostat. (2017). *Harmonised Index of Consumer Prices: Practical Guide for Processing Supermarket Scanner Data*. European Commission.
<https://circabc.europa.eu/ui/group/7b031f10-ac19-4da3-a36f-58708a70133d/library/8e1333df-ca16-40fc-bc6a-1ce1be37247c/details/1.2>
- Eurostat. (2020). *Practical guidelines on web scraping for the HICP*. European Commission.
<https://ec.europa.eu/eurostat/documents/272892/12032198/Guidelines-web-scraping-HICP-11-2020.pdf>

- Eurostat. (2022). *Guide on Multilateral Methods in the Harmonised Index of Consumer Prices* (2022 ed.). Publication Office of the European Union.
<https://doi.org/10.2785/873932>
- Eurostat. (2024). *Harmonised Index of Consumer Prices (HICP): Methodological Manual* (2024 ed.). Publication Office of the European Union.
<https://ec.europa.eu/eurostat/documents/3859598/18594110/KS-GQ-24-003-EN-N.pdf>
- Glaser-Opitzová, H. (2019). Nové zdroje údajov pre cenovú štatistiku a metódy ich spracovania. *Slovenská štatistika a demografia/Slovak Statistics and Demography*, 29(4), 49 – 66. <https://ssad.statistics.sk/SSaD/index.php/nove-zdroje-udajov-pre-cenovu-statistiku-a-metody-ich-spracovania/>
- Glaser-Opitzová, H. (2022). Use of scanner data in measuring the consumer price index in the conditions of the Slovak Republic. In D. Kuběnková, J. Zeman, & P. Puškárová (Eds.), *EDAMBA 2021: Conference proceedings* (pp. 92 – 102). University of Economics in Bratislava.
<https://doi.org/10.53465/EDAMBA.2021.9788022549301.92-102>
- Glaser-Opitzová, H., & Mazureková, P. (2023). Vplyv parciálneho časového pokrytia údajov zo skenerov na presnosť cenových indexov. *Slovenská štatistika a demografia/Slovak Statistics and Demography*, 33(3), 39 – 54.
<https://ssad.statistics.sk/SSaD/index.php/vplyv-parcialneho-casoveho-pokrytia-udajov-zo-skenerov-na-presnost-cenovych-indexov/>
- Greenhough, L. (2019). *New data sources in consumer price statistics: July 2019*. Office for National Statistics.
<https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/compendium/economicreview/july2019/newdatasourcesinconsumerpricestatisticsjuly2019>
- IMF, ILO, Eurostat, UNECE, OECD, & The World Bank. (2020). *Consumer price index manual: Concepts and methods*. <https://www.imf.org/-/media/files/data/cpi/cpi-manual-concepts-and-methods.pdf>
- Knížat, P. (2023). Web scraped data in consumer price indices. *Statistical Journal of the IAOS*, 39(1), 203 – 212. <https://doi.org/10.3233/SJI-220115>
- Knížat, P., Furková, A., Glaser-Opitzová, H., & Peťko, P. (2025). Nowcasting industrial production index with high-frequency toll data. *Statistical Journal of the IAOS*, 41(1), 102 – 111. <https://doi.org/10.1177/18747655241307540>
- Knížat, P., & Glaser-Opitzová, H. (2023). Index spotrebiteľských cien z webscrapovaných údajov: Analýza vybranej produktovej skupiny. *Slovenská štatistika a demografia/Slovak Statistics and Demography*, 33(1), 37 – 49.
<https://ssad.statistics.sk/SSaD/index.php/index-spotrebiteľskych-cien-z-webscrapovanych-udajov-analyza-vybranej-produktovej-skupiny/>
- Knížat, P., Glaser-Opitzová, H., Furková, A., & Vojtková, M. (2024). Multilateral indices in official price statistics and a new additive splicing method. *Quality & Quantity*, 58(5), 4207 – 4222. <https://doi.org/10.1007/s11135-024-01848-3>
- Krsinich, F. (2015). *Implementation of consumer electronics scanner data in the New Zealand CPI* [Paper presentation]. New Zealand Association of Economists Conference, Wellington, New Zealand.
<https://www.nzae.org.nz/wp-content/uploads/2017/01/Frances.pdf>
- Štatistický úrad Slovenskej republiky (2026). *Spotrebiteľské ceny a ceny produkčných štatistik*. Retrieved April 17, 2026, from <https://lnk.sk/qtwfm>

van der Grient, H., & de Haan, J. (2010). *The use of supermarket scanner data in the Dutch CPI*. Statistics Netherlands. Statistics Netherlands. <https://www.cbs.nl/en-gb/our-services/methods/surveys/comprehensive-description/the-use-of-supermarket-scanner-data-in-the-dutch-cpi>

RESUMÉ

Príspevok sa zaoberá metodologickými aspektmi manažérstva kvality pri implementácii nových zdrojov údajov do produkcie indexov spotrebiteľských cien v Štatistickom úrade Slovenskej republiky. Pozornosť je venovaná celému produkčnému procesu od nadviazania spolupráce s poskytovateľmi údajov cez bezpečný prenos, validáciu, spracovanie a automatizáciu údajov až po výpočet cenových indexov a zabezpečenie kvality výsledných štatistických výstupov. Príspevok poukazuje na význam vytvorenia robustného produkčného prostredia, ktoré kombinuje moderné informačné technológie, automatizované spracovanie veľkoobjemových dát a odborný dohľad expertov na cenovú štatistiku. Zároveň zdôrazňuje, že úspešná implementácia údajov zo skenerov obchodných reťazcov, online transakčných údajov a webovo získavaných údajov si vyžaduje nielen metodologické inovácie, ale aj systematické uplatňovanie princípov manažérstva kvality počas celého životného cyklu údajov. Skúsenosti Štatistického úradu Slovenskej republiky potvrdzujú, že takýto prístup vytvára predpoklady na zvyšovanie kvality, spoľahlivosti a udržateľnosti moderných cenových štatistík.

RESUME

This paper examines the methodological aspects of quality management in the implementation of new data sources for the compilation of the Consumer Price Index at the Statistical Office of the Slovak Republic. It covers the entire production process, from establishing cooperation with data providers through secure data transmission, validation, processing, and automation to the compilation of price indices and quality assurance of statistical outputs. The paper highlights the importance of developing a robust production environment that combines modern information technologies, automated processing of large volumes of data, and expert supervision by price statistics specialists. It further emphasizes that the successful integration of scanner data, online transaction data, and web-scraped data requires not only methodological innovation but also the systematic application of quality management principles throughout the entire data life cycle. The experience of the Statistical Office of the Slovak Republic confirms that such an approach provides a basis for improving the quality, reliability, and sustainability of modern price statistics.

PROFESIJNÝ ŽIVOTOPIS

RNDr. Petra Mazureková, PhD., pracuje ako metodička v sekcii všeobecnej metodiky, registrov a koordinácie národného štatistického systému v odbore metód štatistických zisťovaní Štatistického úradu Slovenskej republiky ako členka tímu zodpovedného za efektívne štatistické metódy, analýzy a výpočty s dôrazom na šandardizáciu štatistických procesov, systém monitorovania, vykazovania a hodnotenia kvality štatistických zisťovaní a ich produktov a využívanie administratívnych zdrojov údajov na štatistické účely. Od mája 2018 je projektovou manažérkou projektu Scanner data.

Ing. Helena Glaser-Opitzová, PhD., je generálna riaditeľka sekcie všeobecnej metodiky, registrov a koordinácie národného štatistického systému Štatistického úradu Slovenskej republiky a členka riaditeľskej skupiny Eurostatu pre metodológiu (DIME), ktorá poskytuje poradenstvo Európskemu štatistickému výboru (ESSC) v strategických otázkach. Riadila a podieľala sa na mnohých modernizačných aktivitách úradu. V súčasnosti vedie interný projekt úradu zameraný na modernizáciu cenových štatistík.

KONTAKT

petra.mazurekova@statistics.sk

helena.glaser-opitzova@statistics.sk