

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

3/2023

ročník/volume 33

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

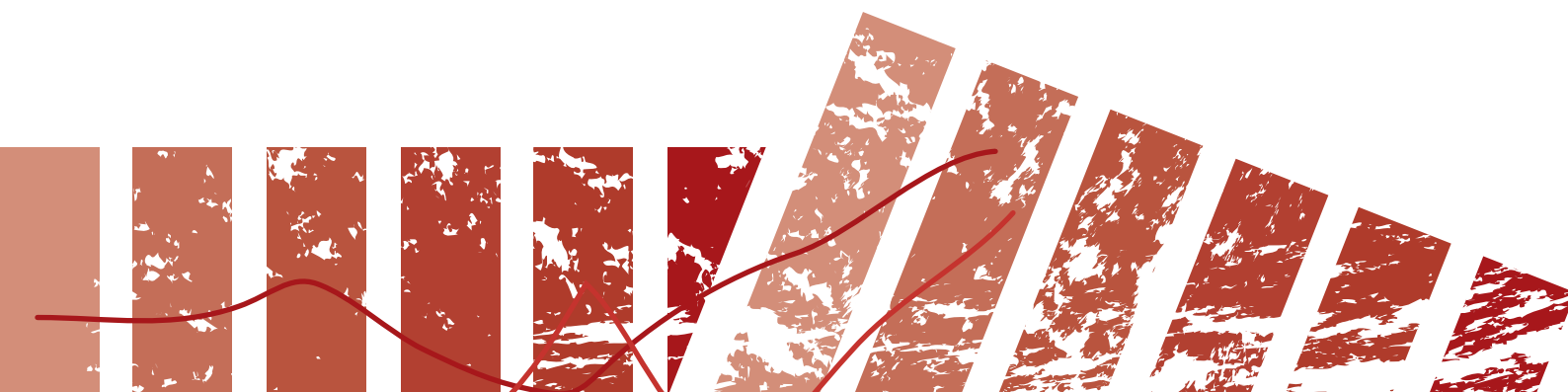
Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 6

Typ článku/Type of article: informatívny článok/informative article

Strany/Pages: 69 – 74

Dátum vydania/Publication date: 15. júl 2023/July 15, 2023



Informatívny článok/Informative article

POTENCIÁL VYUŽÍVANIA BIG DATA V ŠTATISTIKE (NESMIE NÁM UJSŤ VLAK)

THE POTENTIAL OF USING BIG DATA IN STATISTICS (WE CAN'T MISS THE TRAIN)

ABSTRAKT

Príspevok sa zaoberá iniciatívami Eurostatu v oblasti využívania Big Data v oficiálnych štatistikách a transformáciou týchto iniciatív na podmienky Štatistického úradu Slovenskej republiky. Tieto iniciatívy sa v súčasnosti realizujú prostredníctvom dvoch projektov financovaných zo zdrojov Európskej únie, pričom ich cieľom je preukázať vhodnosť využívania Big Data pri štatistickej produkcii v rôznych sférach štatistiky.

ABSTRACT

The paper deals with Eurostat's initiatives in the field of using Big Data in official statistics and the transformation of these initiatives into the conditions of the Statistical Office of the Slovak Republic. These initiatives are currently implemented through two projects funded by the European Union, while their goal is to demonstrate the appropriateness of using Big Data in statistical production in various spheres of statistics.

KLÚČOVÉ SLOVÁ

Big Data, oficiálna štatistika, cenové indexy, index sociálneho napätia, pohyb obyvateľstva, výkonnosť ekonomiky

KEY WORDS

Big Data, official statistics, price indices, social tension index, population movement, performance of the economy

1. ÚVOD

Big Data prinášajú nesmierne dôležitý rozmer do všetkých sfér života. Britský matematik Clive Humby už v roku 2006 slávne povedal, že „údaje sú nová ropa“. Odvtedy za posledných 16 rokov spoločnosti všetkých veľkostí a vo všetkých odvetviach vyhľadávali a ukladali čoraz viac údajov o svojich zákazníkoch, obchodných operáciách, aby vedeli podporiť svoj výkon a dosiahli stanovené ciele [1].

Treba povedať, že: „Potenciál Big Data pre oficiálnu štatistiku spočíva práve v obrovskom množstve informácií, ktoré obsahujú. Môžu tiež pokrývať oblasti spoločnosti, pre ktoré ešte neexistujú oficiálne štatistiky. Ich veľkosť a objem môže viesť k vyššej presnosti a získaniu väčších detailov na tvorbu štatistík. Vysoká rýchlosť ich produkcie môže viesť k častejším a včasnejším štatistickým odhadom a ich veľká rozmanitosť zároveň k tvorbe štatistík v nových oblastiach.“ [2]

Z toho vyplýva, že ak Štatistický úrad SR nevyužije v dobe Big Data ich potenciál, môže prísť komerčný sektor a efektívne tieto údaje využiť a vytvoriť tak suplement k štatistickej produkcii, ktorý sa časom môže stať presnejším, efektívnejším a zrozumiteľnejším. A preto nie je na čo čakať a je potrebné, aby sa Štatistický úrad SR vydal smerom k Big Data.

2. VÝCHODISKÁ NA VYUŽITIE BIG DATA V ŠTATISTIKE V PODMIENKACH SR

Štatistický úrad SR má povinnosť a zodpovednosť za hľadanie a využívanie vhodných údajov na štatistickú tvorbu.

Štatistický úrad Slovenskej republiky je v zmysle §5 ods. 1 zákona č. 540/2001 Z. z. zákona o štátnej štatistike ústredným orgánom štátnej správy pre oblasť štátnej štatistiky, je kontaktným miestom pre Európsku komisiu v rámci Európskeho štatistického systému a koordinátorom Národného štatistického systému.

Jednou zo základných úloh Štatistického úradu SR je získavanie štatistických údajov a štatistických informácií potrebných na posudzovanie sociálno-ekonomického vývoja spoločnosti, poskytovanie a zverejňovanie štatistických informácií, koordinácia Národného štatistického systému Slovenskej republiky a plnenie záväzkov vyplývajúcich z medzinárodných zmlúv v oblasti štátnej štatistiky, ktorými je Slovenská republika viazaná.

Realizácie projektov v súlade s iniciatívami Európskej komisie v oblasti Big Data pomôžu zviditeľniť a spropagovať Štatistický úrad SR.

Práve v súvislosti s vyššie uvedeným a s možnosťou získavania dátových zdrojov sa dlhodobo rozvíja iniciatíva Európskej komisie prostredníctvom Eurostatu na využívanie Big Data v kontexte oficiálnych štatistík. Táto iniciatíva sa transformovala do projektu ESSnet Big Data (Big data: from exploration to exploitation). Cieľom projektu je integrácia Big Data do pravidelnej tvorby oficiálnych štatistík prostredníctvom pilotných prieskumov potenciálu vybraných Big Data a budovania konkrétnych aplikácií. Slovensko sa do tejto iniciatívy zapojilo ako jeden z 28 partnerov (https://cros-legacy.ec.europa.eu/content/essnet-big-data-1_en).

Záveru iniciatívy podľa Eurostatu sú:

- údaje obsiahnuté na webových stránkach môžu byť prínosné pre oficiálnu štatistiku,
- tradičné spôsoby zberu môžu byť obohatené týmito údajmi,
- údaje môžu pomôcť získavať ďalšie indikátory, ktoré nie je možné produkovať z tradičných zdrojov údajov,
- **nevyužívaním alternatívnych zdrojov údajov v oblasti oficiálnych štatistík sa môže oslabiť postavenie národných štatistických úradov v prospech podnikateľského komerčného sektora.**

Preto sa aj Štatistický úrad SR rozhodol prispieť k vytváraniu pilotných riešení na transformáciu Big Data do štatistických produktov, či už vo forme experimentálnej štatistiky, alebo aj do oficiálnych štatistík po overení stability a vhodnosti využitia príslušných údajov.

3. TRANSFORMÁCIA INICIATÍVY EURÓPSKEJ KOMISIE V PODMIENKACH ŠTATISTICKÉHO ÚRADU SR

Realizácia experimentálnej štatistiky nie je jednoduchá a vyžaduje si relatívne vysoké finančné a odborné kapacity. V období rozvíjania myšlienok o využití Big Data v experimentálnej štatistike Ministerstvo investícií, regionálneho rozvoja a informatizácie SR vyhlásilo výzvu, zameranú na zlepšenie využívania údajov vo verejnej správe (<https://www.mirri.gov.sk/projekty/projekty-esif/operacny-program->

[integrovana-infrastruktura/prioritna-os-7-informacna-spolocnost/vyzvania-a-vyzvy/vyzva-c-opii-2019-7-10-dop/index.html](https://www.itms2014.sk/schvalena-zonfp?id=51997001-4fc1-42d9-8191-c2b15abbaa74)).

Cieľom tejto dopytovo orientovanej výzvy je realizovať projekty zamerané na analytické využitie údajov v definovanej problémovej oblasti s cieľom poskytnúť riešenia na zlepšenie rozhodovania a zabezpečenia fungovania verejnej správy.

Štatistický úrad sa zapojil do tejto výzvy predložením dvoch projektov, na ktorých realizáciu prostredníctvom žiadosti o nenávratný finančný príspevok získal zdroje práve prostredníctvom predmetnej výzvy. Ide o nasledovné projekty:

- Dynamický cenový model – <https://www.itms2014.sk/schvalena-zonfp?id=51997001-4fc1-42d9-8191-c2b15abbaa74>
- Socioekonomické aspekty Big Data v štatistike – <https://www.itms2014.sk/schvalena-zonfp?id=c13d0406-9036-44bb-8b76-51f5f9b53193>

V nasledujúcom texte uvádzame rámcové zameranie projektov, ktoré bolo definované na začiatku kreovania projektov:

- Socioekonomické aspekty Big Data v štatistike – účelom realizácie projektu je spracovanie veľkých objemov dát od mobilných operátorov a zo sociálnych sietí na vytvorenie modelov pohybu populácie a modelu nálad populácie s dôrazom na:
 - spracovanie údajov a prepojenie do GIS na identifikovanie sociálnych vplyvov pohybu populácie,
 - spracovanie analýz na vypracovanie odhadov tzv. dennej populácie, ktoré poskytujú prehľad o počte ľudí umiestnených v konkrétnych oblastiach v konkrétnom čase,
 - spracovanie informácií na sociálnych sieťach a definovanie závislosti medzi náladou v spoločnosti a súčasnou socioekonomickou problematikou v jednotlivých regiónoch,
 - spracovanie modelu dopravy z údajov z Národnej diaľničnej spoločnosti NDS a definovanie odhadov HDP, resp. indexu priemyselnej produkcie.
- Dynamický cenový model – účelom realizácie projektu je otestovanie možností využitia iných ako v súčasnosti využívaných údajov pre cenové štatistiky, ktoré Štatistický úrad SR realizuje. Ide o využitie moderných metód alternatívneho zberu údajov ako napr. web scrapingu, data feeding (alebo obdobných) na získavanie údajov o cenách a porovnávanie výsledkov s tradičnými formami sledovania cien.
- Výstupmi projektu budú cenové modely, ktoré budú testované v porovnaní s existujúcim štandardným postupom v rámci štatistickej produkcie.

4. PROCES NASTAVENIA REALIZÁCIE PROJEKTOV A ICH REALIZÁCIA

Realizácia projektov nie je samoučelná a na ich obhajobu bolo potrebné definovať hypotézy a benefity tak, aby mohli získať podporu

V procese prípravy projektov bola vypracovaná štandardná projektová dokumentácia v zmysle výzvy. Súčasťou dokumentácie bolo aj definovanie prínosov, ktoré projekty spoločnosti prinesú. Tak ako uvádza projektová dokumentácia, dosah

realizácie oboch projektov bol vysoko pozitívny už v čase ich prípravy, na čo poukazuje aj analýza prínosov a nákladov (CBA analýza) projektov.

Medzi najvýznamnejšie benefity patria:

- Socioekonomické aspekty Big Data v štatistike:
 - zvyšovanie transparentnosti a adresnosti rozhodovania, najmä v oblasti investícií,
 - znižovanie prevádzkových nákladov dopravných spoločností s ohľadom na poznatky o pohybe obyvateľstva,
 - adresné rozhodovanie pri socioekonomických otázkach (napr. sociálne balíčky) v závislosti od správania populácie,
 - zvýšenie využitia potenciálu dátového trhu poskytnutím otvorených údajov.
- Dynamický cenový model:
 - znižovanie záťaže podnikateľov zaradených do procesu zisťovania spotrebiteľských cien,
 - znižovanie nákladov na spracovanie cenových štatistík na úrade,
 - zefektívnenie rozhodovania v oblasti investícií Slovenskej republiky,
 - spresnenie údajov o inflačnom vývoji Slovenskej republiky.

Počas realizácie oboch projektov sa neustále prehodnocujú benefity a objavujú nové potenciálne prínosy, ktoré sa ich realizáciou postupne overujú.

V súčasnosti prebieha realizácia projektov, ktorá iteratívne poukazuje na dôležitosť hľadania spôsobov, ako využívať Big Data v oblasti štatistiky.

Oba projekty sa dostali do realizačnej fázy, ktorá prebieha práve v súčasnosti. Na projektoch pracujú projektové tímy zložené z interných expertov, ako aj externých expertov, ktorých úlohou je práve tvorba dátových modelov, na ktorých sú oba projekty založené.

Vzhľadom na fakt, že ide do veľkej miery o experimentálne projekty, boli spresnené očakávané výstupy projektov, a to takto:

- Dynamický cenový model:
 - Dátový model na výpočet indexov položiek spotrebného koša, ktorý je založený na využití online transakčných údajov spoločnosti Heureka. Spolupráca bola aj medializovaná.
 - Potenciálom tohto modelu je nahradiť tradičné procesy získavania a spracovania údajov o cenách jednotlivých položiek spotrebného koša.
- Socioekonomické aspekty Big Data v štatistike – tento projekt rozvinul svoje zadanie na 3 základné modely: model pohybu obyvateľstva; model indexu sociálneho napätia a model výkonu ekonomiky na základe transakčných údajov z mýtného systému:
 - Model pohybu obyvateľstva – cieľom tohto modelu je vytvoriť mapu dennej a nočnej populácie založenej na údajoch od poskytovateľov mobilných telekomunikačných služieb. Údaje budú spracovávané anonymizovane. Potenciál modelu do budúcnosti je predovšetkým v oblasti tvorby funkčných regiónov, resp. intervenčných opatrení, ktoré budú mať vyššiu adresnosť.

- Model indexu sociálneho napätia – tento model sa realizuje podľa vzoru Holandského štatistického úradu. Cieľom modelu je hľadať korelácie medzi významnými udalosťami, ktoré sa v spoločnosti udiali a zmenou nálad, ktorá bude vyhodnocovaná na základe algoritmov umelej inteligencie.
- Model výkonu ekonomiky – cieľom tohto modelu je na základe transakčných údajov z mýtného systému identifikovať tzv. flash odhad ekonomických indikátorov krajiny. Potenciálom modelu je disponovať časovým predstihom informácie o budúcom ekonomickom vývoji krajiny, čo je potrebné na makroekonomické rozhodovanie a tvorbu politík.

5. ZÁVER

Realizáciou týchto projektov Štatistický úrad SR potvrdzuje, že záujem o využívanie Big Data v štatistických procesoch nie je len platonický, ale predstavuje reálnu a dôležitú súčasť budúcnosti štatistiky. Výstupy projektov budú prezentované po ukončení ich realizácie (cca september – október 2023), avšak už dnes sa ukazuje, že to je ten správny smer, ktorým sa Štatistický úrad SR dostáva viac do pozitívneho povedomia spoločnosti a využije potenciál, ktorý doba údajov prináša.

Skúmanie a experimentovanie v oblasti využívania takéhoto typu údajov je kontinuálny proces a jedným alebo dvoma projektmi nie je možné preskúmať všetky možnosti a otestovať ich v praxi. Týmito projektmi sa otvorilo veko truhlice pokladov, ktoré Big Data pre štatistiku prinášajú.

Už v súčasnosti sa hľadajú možnosti, ako pokračovať v začatých iniciatívach a prípadne ich rozvinúť o nové možnosti a trendy. Pevne veríme, že sa tento smer zachová a výstupy z projektov budú zavedené do praxe v rámci produkcie oficiálnych štatistík.

LITERATÚRA

- [1] LAROCK, T.: „Data is the new oil“, But that also means it can be risky. [online]. [cit.20-02-2023] Dostupné na: <https://www.dbta.com/Columns/Next-Gen-Data-Management/Data-is-the-New-Oil-But-That-Also-Means-it-Can-be-Risky-155275.aspx>
- [2] BRAAKSMA, B. – ZEELLENBERG, Z.: Big data in official Statistics. [online], [cit. 30-05-2023, s. 4] Dostupné na: <https://www.cbs.nl/-/media/pdf/2020/04/dp-big-data-in-official-statistics.pdf>

RESUMÉ

Štatistika je oblasť, pre ktorú sú údaje nesmierne dôležité a bez nich by nebolo možné realizovať štatistické výstupy. Projekty, ako sú dva dopytové projekty opísané v texte, sú nevyhnutné na overovanie možností využívania aj alternatívnych zdrojov údajov pre oficiálne štatistiky. Vzhľadom na dôležitosť a striktnosť štatistického procesu je relatívne dlhá cesta od identifikovania potenciálneho nového zdroja údajov do jeho zavedenia do štatistickej produkcie a procesu diseminácie. Experimentálne štatistické projekty preto musia byť súčasťou agendy úradu, aby bolo možné zefektívňovať, modernizovať a skvalitňovať výstupy hodnototvorného štatistického procesu. Preto je potrebné neustále hľadať zdroje na podporu takýchto iniciatív, ktoré z dlhodobého hľadiska prinesú rozvoj a pokrok do sféry štatistiky. Bez investícií nie je pokrok.

RESUME

Statistics is a field for which data is extremely important and without them it would be impossible to carry out statistical outputs. Projects, such as the two projects described in the text, are necessary to verify the possibilities of using alternative data sources for official statistics. Given the importance and stringency of the statistical process, there is a relatively long way from identifying a potential new source of data to introducing this source into the statistical production and dissemination process. Experimental statistical projects must therefore be part of the Office's agenda in order to enhance, modernize and improve the quality of the outputs of the value-creating statistical process. Therefore, it is necessary to constantly look for resources to support such initiatives, which in the long term will bring development and progress to the field of statistics. There is no progress without investment.

PROFESIJNÝ ŽIVOTOPIS

Peter Ďuriš je absolventom Ekonomickej univerzity v Bratislave. Profesionálne začína ako procesný analytik v konzultačnej spoločnosti Centire, kde mal na starosti okrem riadenia analytických tímov aj riadenie projektov. Od roku 2012 je konateľom spoločnosti Go SMART, s. r. o., v ktorej okrem iného pôsobí aj ako konzultant v oblasti využívania nových zdrojov údajov a metód v oficiálnej štatistike. Rovnako sa zaoberá prípravou projektov a ich realizáciou v štátnej, verejnej ale aj súkromnej sfére.

KONTAKT

peter.duris@gosmart.consulting