

# SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS  
and DEMOGRAPHY

1/2024

ročník/volume 34

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: editoriál/editorial

Strany/Pages: 5 – 6

Dátum vydania/Publication date: 15. január 2024/January 15, 2024



## EDITORIÁL

Vážení čitatelia,

špeciálne číslo časopisu Slovenská štatistika a demografia je venované projektu s názvom **Socioekonomické aspekty Big Data**, ktorý Štatistický úrad SR realizoval a ukončil v roku 2023, ako aj jeho výstupom. Príspevky opisujú prácu s 3 novými typmi zdrojov – údajmi od mobilných operátorov, údajmi mýtného systému a sociálnymi sieťami. Pri práci s nimi sa zároveň používali nové nástroje, predovšetkým machine learning (strojové učenie) a umelá inteligencia. Práve práca s uvedenými zdrojmi, nastavenie prístupu k nim, otázka ochrany údajov a už spomenuté nástroje, vrátane ukladania Big Data a udržateľnosť prístupu k nim boli v centre pozornosti projektu. Dôraz na samotné výstupy je sekundárny, keďže tie majú charakter experimentálnej štatistiky.



PhDr. Ľudmila Ivančíková, PhD.

Samotný projekt je stručne opísaný v úvodnom príspevku J. Bardyho, kde autor zdôrazňuje, že práca s Big datami je *„meniaca sa oblasť poskytujúca nové zdroje dát, nové možnosti v oblasti hardvérovej a softvérovej infraštruktúry, nové oblasti využitia a potrebu výmeny skúseností doma aj v zahraničí“*.

Existencia Big Data vytvára predpoklad ich využitia v štatistickej produkcii. Často sa spomína najmä aspekt rýchlejšieho prístupu k údajom, ktoré opisujú reálny stav v spoločnosti. Treba však zdôrazniť a projekt sám to potvrdil, že na to, aby sa tento predpoklad naplnil, je potrebné nastaviť jasný a obojstranne prospešný dlhodobý vzťah s vlastními zdrojov a údajov; vytvoriť použiteľnú a robustnú infraštruktúru; mať štatistikov so znalosťami, ktorí budú pracovať aj so „surovými“ údajmi, t. j. údajmi, ktoré nie sú „predpripravené“ vlastními. Len tak bude možné nastaviť rámec hodnotenia kvality zdroja a uviesť zdroj postupne do procesu štatistickej produkcie.

V experimentálnej štatistike je potrebné zohľadniť aj viacročné opakovanie navrhnutých postupov a metód a niektoré jej výstupy nikdy do produkcie zaradené nebudú, resp. navrhnuté metódy nebudú v praxi použiteľné. To však neznamená, že uvedené zdroje by sme mali ignorovať. Súčasná znalosť poznania nám ponúka témy, ktoré nevieme štandardnými zdrojmi pokryť a len pri konkrétnej práci vieme identifikovať ich reálny potenciál. A v tom je prínos spomínaného projektu, ako aj jednotlivých článkov.

**PhDr. Ľudmila IVANČÍKOVÁ, PhD.**

*Autorka je generálnou riaditeľkou sekcie sociálnych štatistík a demografie Štatistického úradu SR.*

## EDITORIAL

Dear readers,

A special issue of the journal *Slovak Statistics and Demography* is dedicated to the project entitled ***Socioeconomic Aspects of Big Data***, which was implemented and completed together with its outputs by the Statistical Office of the SR. The articles describe working with 3 new types of sources - data from mobile operators, toll system data and social networks. At the same time, new tools were used while working with them, especially machine learning and artificial intelligence. The project focused on working with the mentioned resources, access adjustment to them, the issue of data protection and the already mentioned tools, including the storage of Big Data and the access sustainability to them. The emphasis on the outputs is secondary, as they have the character of experimental statistics.

The project itself is briefly described in the introductory contribution of J. Bardy, where the author emphasizes that work with Big Data is *"a changing field providing new data sources, new possibilities in hardware and software infrastructure, novel applications and the need to exchange experience at home and abroad"*.

The existence of Big Data creates a presumption for their use in the statistical production. The aspect of faster access to data that describes the real situation in society is often mentioned. However, it must be emphasized, and it was confirmed by the project, that in order to fulfill this assumption, it is necessary to set up a clear and mutually beneficial long-term relationship with the data and resource owners; create an applicable and robust infrastructure; to have knowledgeable statisticians who will also work with "raw" data, i.e. data not "pre-prepared" by the owners. Only then will it be possible to set the framework for evaluating the quality of the source and introduce the source gradually into the process of statistical production.

As part of experimental statistics, it is also necessary to take into account multi-annual repetition of the proposed procedures and methods, and some of its outputs will never be included in the production, or the proposed methods will not be applicable in practice. Although this does not mean that the mentioned sources should be ignored. The current state of knowledge gives us topics that cannot be covered with standard sources, and their real potential can only be identified through specific work, which is the benefit of the mentioned project, as well as of the individual articles.

***PhDr. Ľudmila IVANČÍKOVÁ, PhD.***

*The author is the Director of the Social Statistics and Demography Directorate of the Statistical Office of the SR.*