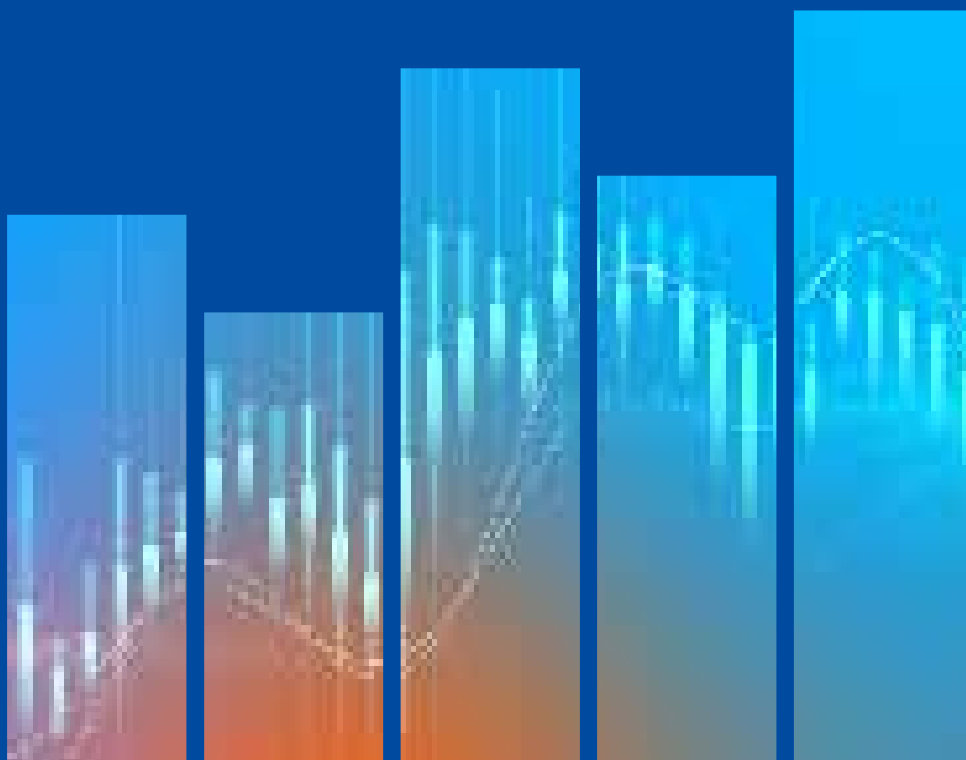
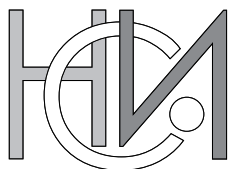


СТАТИСТИКА STATISTICS

1/2025



РЕПУБЛИКА БЪЛГАРИЯ
REPUBLIC OF BULGARIA



НАЦИОНАЛЕН СТАТИСТИЧЕСКИ ИНСТИТУТ
NATIONAL STATISTICAL INSTITUTE

СТАТИСТИКА

STATISTICS

1/2025

СОФИЯ, 2025
SOFIA, 2025

РЕДАКЦИОННА КОЛЕГИЯ

Главен редактор: проф. д-р Михаил Кончев

Зам.-председател на Националния статистически институт - София, България

Заместник-главен редактор: д-р Свилен Колев

Зам.-председател на Националния статистически институт - София, България

Членове:

проф. д-р Елефтериос Таласинос

Асоцииран професор, Университет на Малта - Мсида, Малта, гост-професор в 8 европейски университета

проф. д-р по икономика Никша Алфирович

Университет на Сплит - Сплит, Хърватия

доц. д-р Мариана Коцева

Университет за национално и световно стопанство - София, България

Генерален директор на Евростат - Люксембург, Люксембург

доц. д-р инж. Александър Ефремов

Технически университет - София, България

доц. д-р Ангел Марчев

Софийски университет „Св. Климент Охридски“ - София, България

доц. д-р Боряна Богданова

Софийски университет „Св. Климент Охридски“ - София, България

доц. д-р Екатерина Тошева

Университет за национално и световно стопанство - София, България

доц. д-р Калоян Харалампиев

Софийски университет „Св. Климент Охридски“ - София, България

доц. д-р Любомир Иванов Тодоров

Стопанска академия „Димитър А. Ценов“ - Свищов, България

доц. д-р Маргарита Ламбова

Икономически университет - Варна, България

доц. д-р Пламен Иванов Петков

Стопанска академия „Димитър А. Ценов“ - Свищов, България

доц. д-р Александър Найденов

Университет за национално и световно стопанство - София, България

доц. д-р Пламен Т. Ненов

Икономист-изследовател, Банка на Норвегия, изследовател I, Жилищна лаборатория, Метрополитен университет в Осло, доц. по икономика, Норвежко бизнес училище - Осло, Норвегия

чл.-кор. Тудорел Андрей

Председател на Националния статистически институт на Румъния - Букурещ, Румъния

проф. д.ик.н. Диана Иванова Георгиева

Лесотехнически университет - София, България

Отговорен редактор: Антоанета Илкова

директор на дирекция „Многосекторна статистика, методология и регистри“,
Национален статистически институт - София, България

Стилов редактор: Веселина Попова

Преводачи на английски и руски език: Радина Лъондева, Веселина Попова

Адрес на редакцията:

1038, София, ул. „П. Волов” № 2

тел. +359 2 9857 704;

e-mail: stat_journal@nsi.bg

ISSN 2367-5489

EDITORIAL TEAM

Editor-in-Chief: Prof. Mihail Konchev, PhD

Deputy President of the National Statistical Institute - Sofia, Bulgaria

Deputy Editor-in-Chief: Svilen Kolev, PhD

Deputy President of the National Statistical Institute - Sofia, Bulgaria

Members:

Prof. Eleftherios Thalassinos, PhD

Affiliate Professor, University of Malta - Msida, Malta, Visiting Professor in 8 European Universities

Prof. Nikša Alfrević, PhD in Economics

University of Split - Split, Croatia

Assoc. Prof. Maryana Kotzeva, PhD

University of National and World Economy - Sofia, Bulgaria

Director General of Eurostat - Luxembourg City, Luxembourg

Assoc. Prof. Alexander Efremov, PhD Eng.

Technical University - Sofia, Bulgaria

Assoc. Prof. Angel Marchev, PhD

Sofia University 'St. Kliment Ohridski' - Sofia, Bulgaria

Assoc. Prof. Boryana Bogdanova, PhD

Sofia University 'St. Kliment Ohridski' - Sofia, Bulgaria

Assoc. Prof. Ekaterina Tosheva, PhD

University of National and World Economy - Sofia, Bulgaria

Assoc. Prof. Kaloyan Haralampiev, PhD

Sofia University 'St. Kliment Ohridski' - Sofia, Bulgaria

Assoc. Prof. Lyubomir Ivanov Todorov, PhD

Academy of Economics 'Dimitar A. Tsenov' - Svishtov, Bulgaria

Assoc. Prof. Margarita Lambova, PhD

University of Economics - Varna, Bulgaria

Assoc. Prof. Plamen Ivanov Petkov, PhD

Academy of Economics 'Dimitar A. Tsenov' - Svishtov, Bulgaria

Assoc. Prof. Alexander Naydenov, PhD

University of National and World Economy - Sofia, Bulgaria

Assoc. Prof. Plamen T. Nenov, PhD

Research Economist, Norges Bank, Researcher I, Housing Lab, Oslo Metropolitan University,

Assoc. Prof. of Economics, Norwegian Business School - Oslo, Norway
Corr. Member, Tudorel Andrei
President of the National Statistical Institute of Romania - Bucharest, Romania

Prof. Diana Ivanova Georgieva, DSc in Economics
University of Forestry - Sofia, Bulgaria

Responsible editor: Antoaneta Ilkova
Director of Multi-Domain Statistics, Methodology and Registers Directorate at the National Statistical
Institute - Sofia, Bulgaria

Style editor: Vesselina Popova

English and Russian Translation: Radina Lyondeva, Vesselina Popova

Editorial address:

2, P. Volov St., Sofia 1038, Bulgaria
Phone: +359 2 9857 704;
e-mail: stat_journal@nsi.bg

ISSN 2367-5489

СЪДЪРЖАНИЕ

Стр.

РАЗВИТИЕ НА СЪВРЕМЕННАТА СТАТИСТИЧЕСКА ТЕОРИЯ И ПРАКТИКА

Румен Антонов	Развитие на европейската правна рамка в областта на статистиката 13
---------------	---

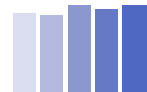
СТАТИСТИЧЕСКИ ИЗСЛЕДВАНИЯ И АНАЛИЗ НА ДАННИ

Соня Златанова	Репрезентативни изследвания по време на преброяването на населението и жилищния фонд на 4 декември 1985 година 30
----------------	---

Гергана Чешмеджиева	Методика за управление на статистическия процес в областта на околната среда Подход на НСИ за интегрирана методологична рамка 49
---------------------	---

БИЗНЕС ИНТЕЛИГЕНТНИ СИСТЕМИ И ИЗКУСТВЕН ИНТЕЛЕКТ

Мариана Ангелова Костадин Георгиев	Интегриране на данни от интернет в официална статистика: проектът „Доверена умна статистика (TSS) - мрежа за изучаване на уебпространството (WIN)“ 66
---------------------------------------	---

**СОДЕРЖАНИЕ**

Стр.

РАЗВИТИЕ СОВРЕМЕННОЙ СТАТИСТИЧЕСКОЙ ТЕОРИИ И ПРАКТИКИ

Румен Антонов	Развитие европейской правовой базы в области статистики 13
---------------	--

СТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ И АНАЛИЗ ДАННЫХ

Соня Златанова	Репрезентативные исследования во время переписи населения и жилищного фонда 4 декабря 1985 года 30
Гергана Чешмеджиева	Методика управления статистическим процессом в области окружающей среды Подход НСИ к интегрированной методологической рамке49

БИЗНЕС ИНТЕЛЛИГЕНТНЫЕ СИСТЕМЫ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Мариана Ангелова Костадин Георгиев	Интеграция интернет-данных в официальную статистику: проект «Надёжная интеллектуальная статистика (TSS) - вебпространственный интеллект (WIN)»..... 66
---------------------------------------	--

CONTENTS

Page

DEVELOPMENT OF MODERN STATISTICAL THEORY AND PRACTICE

Rumen Antonov	Development of the european legal framework in the field of statistics	13
---------------	---	----

STATISTICAL RESEARCH AND DATA ANALYSIS

Sonya Zlatanova	Representative surveys during the population and housing census of December 4, 1985.....	30
Gergana Cheshmedjieva	Guidelines for management of the statistical process in the field of environmental statistics Approach of NSI for integrated methodological framework	49

BUSINESS INTELLIGENT SYSTEMS AND ARTIFICIAL INTELLIGENCE

Mariana Angelova Kostadin Georgiev	Integrating web data into official statistics: the ‘Trusted smart statistics (TSS) - web intelligence network (WIN)’ project	66
---------------------------------------	--	----



Уважаеми читатели,

Настоящият брой на списание Статистика представя широк спектър от теми, отразяващи както историческия опит на българската статистика, така и съвременните предизвикателства пред европейската и националната статистическа практика.

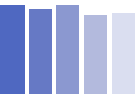
Проследяват се резултатите от репрезентативните изследвания, проведени по време и непосредствено след преброяването на населението и жилищния фонд през 1985 г., които разширяват аналитичния обхват на демографските и социалните процеси. В европейски контекст е представена еволюцията на правната рамка в областта на статистиката - от първите стъпки на интеграцията до последните изменения на Регламент (ЕО) 223/2009.

В същото време се разглеждат и актуални методологични иновации: интегрирането на данни от уеб в официалната статистика чрез проекта *Trusted Smart Statistics - Web Intelligence Network*, както и внедряването на интегрирана методологична рамка за управление на статистическите процеси в областта на околната среда от Националния статистически институт.

Чрез съчетаването на историческия опит, правните основи и съвременните технологични решения настоящият брой подчертава ключовата роля на статистиката за развитието на обществото и насочва вниманието на читателя към възможностите за нейното бъдещо развитие.

Проф. д-р Михаил Кончев

Главен редактор



Уважаемые читатели,

В текущем выпуске журнала «Статистика» представлен широкий спектр тем, отражающих как исторический опыт болгарской статистики, так и современные вызовы, стоящие перед европейской и национальной статистической практикой.

Отслеживаются результаты репрезентативных обследований, проведенных во время и сразу после переписи населения и жилищного фонда 1985 года, что расширяет возможности анализа демографических и социальных процессов. В европейском контексте представлена эволюция правовой базы в области статистики - от первых шагов интеграции до последних поправок к Регламенту (ЕС) 223/2009.

В то же время рассматриваются и современные методологические инновации: интеграция веб-данных в официальную статистику в рамках проекта *Trusted Smart Statistics - Web Intelligence Network*, а также внедрение Национальным статистическим институтом комплексной методологической основы для управления статистическими процессами в области окружающей среды.

Объединяя исторический опыт, правовые основы и современные технологические решения, этот номер подчеркивает ключевую роль статистики в развитии общества и обращает внимание читателя на возможности ее будущего развития.

Проф. д-р Михаил Кончев

Главный редактор



Dear readers,

The latest edition 1/2025 of 'Statistics' journal presents broad spectrum of topics, reflecting both the historical experience of the Bulgarian statistics and the modern challenges to be faced by the European and the national statistical practice.

The results from the representative surveys of the population and the housing fund, held at the time and immediately after the 1985 Census, are being monitored, which broaden the analytical approach of the demographic and social processes. In European context is presented the evolution of the legal framework in the field of statistics - from the first steps of the integration to the latest amendments of Regulation (EC) No 223/2009.

At the same time, the newest methodological innovations are being covered: the integration of data from web in official statistics through the *Trusted Smart Statistics - Web Intelligence Network project*, as well as the implementation of the integrated methodological framework for management of the statistical processes in the field of environment by the National statistical institute.

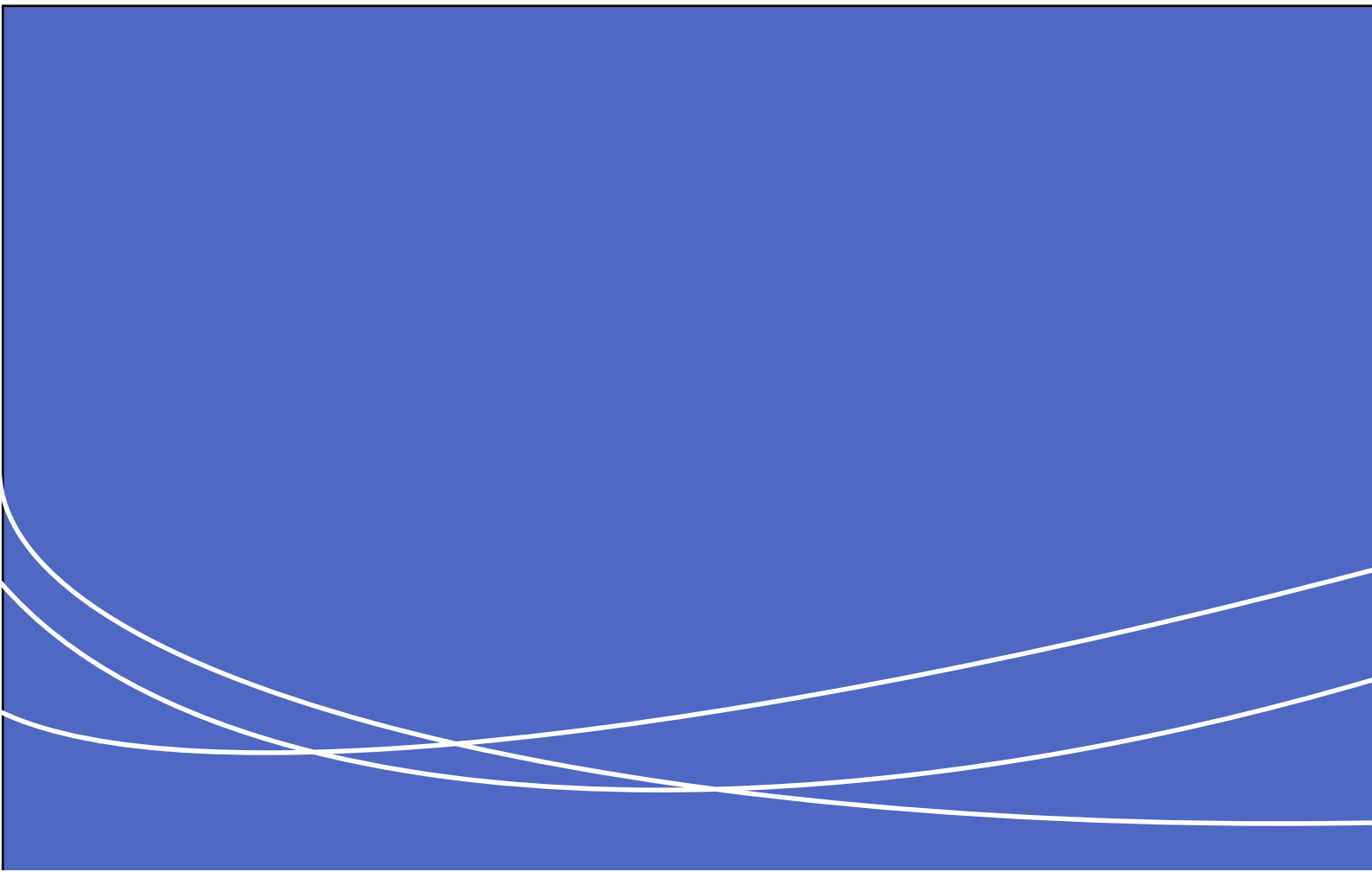
By combining the historical experience, the legal basis and the modern technological solutions, this edition highlights the key role of statistics for the development of society and draws the attention of the reader to the possibilities for its future development.

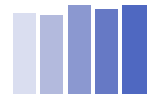
Prof. Mihail Konchev, PhD

Editor-in-Chief

**РАЗВИТИЕ НА СЪВРЕМЕННАТА
СТАТИСТИЧЕСКА ТЕОРИЯ И ПРАКТИКА**

**DEVELOPMENT OF MODERN
STATISTICAL THEORY AND PRACTICE**





РАЗВИТИЕ НА ЕВРОПЕЙСКАТА ПРАВНА РАМКА В ОБЛАСТТА НА СТАТИСТИКАТА

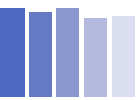
*Румен Антонов**

РЕЗЮМЕ С подписването на договора за учредяване на Европейската общност за въглища и стомана (1952 г.) първите шест държави членки започват един невиджан дотогава процес на икономическа и политическа интеграция. Данните и статистиката бързо започват да играят ключова роля в процеса по изпълнение на Договора, предоставяйки необходимите данни в сектора за производство на въглища и стомана за държавите членки на Общността. Със задълбочаването на европейската интеграция и разширяващия се брой области, които се обхващат от правото на ЕС, се появява и необходимостта от производство на европейски статистически данни в различни нови направления. Това налага наличието на стабилна правна рамка, която да позволи на Евростат и на националните статистически системи на държавите-членки възможността да се адаптират и да се справят с настоящи и бъдещи предизвикателства.

Целта на настоящата статия е да представи на читателите по един систематичен начин прехода на европейската правна рамка в областта на статистиката от началото на европейската интеграция до последното изменение на Регламент (ЕО) 223/2009 относно европейската статистика от 2024 година. Статията поставя основен фокус върху предвидените разпоредби в Регламент (ЕО) 223/2009 относно европейската статистика и двете последвали изменения съответно от 2015 и 2024 г., както и причините, довели до приемането им. По този начин читателят добива представа не само за основния европейски акт в областта на статистиката, но и за историческия контекст, в който е създаден и приет.

Ключови думи: Европейска статистическа система, Комитет на европейската статистическа система, административни източници, държатели на частни данни, поверителни данни

*Старши експерт в отдел „Международни отношения и проекти“, НСИ, София, България, e-mail: raantonov@nsi.bg



РАЗВИТИЕ ЕВРОПЕЙСКОЙ ПРАВОВОЙ БАЗЫ В ОБЛАСТИ СТАТИСТИКИ

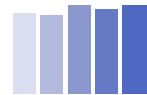
*Румен Антонов**

РЕЗЮМЕ С подписанием Договора об учреждении Европейского объединения угля и стали (1952 г.) первые шесть государства члены начали беспрецедентный процесс экономической и политической интеграции. Данные и статистика быстро стали играть ключевую роль в реализации Договора, предоставляя государствам - членам Общности необходимые данные о секторе производства угля и стали. С углублением европейской интеграции и расширением сферы действия права ЕС возникла также необходимость в разработке европейской статистики в различных новых областях. Это требует стабильной правовой базы, которая позволит Евростату и национальным статистическим системам государств членов адаптироваться и справляться с текущими и будущими вызовами.

Цель данной статьи - систематически представить читателям эволюцию европейской правовой базы в области статистики с начала европейской интеграции до последней поправки к Регламенту (ЕС) 223/2009 о европейской статистике в 2024 году. Основное внимание в статье уделяется положениям, предусмотренным Регламентом (ЕС) 223/2009 о европейской статистике, и двум последующим поправкам, внесенным в 2015 и 2024 годах соответственно, а также причинам, повлиявшим на их принятие. Таким образом читатель получает представление не только об основном европейском акте в области статистики, но и об историческом контексте его создания и принятия.

Ключевые слова: Европейская статистическая система, Комитет Европейской статистической системы, административные источники, владельцы частных данных, конфиденциальные данные.

*Старший эксперт департамента «Международные отношения и проекты», НСИ, София, Болгария,
e-mail: raantonov@nsi.bg



DEVELOPMENT OF THE EUROPEAN LEGAL FRAMEWORK IN THE FIELD OF STATISTICS

*Rumen Antonov**

SUMMARY With the signing of the Treaty establishing the European Coal and Steel Community (1952), the first six Member States began an unprecedented process of economic and political integration. Data and statistics quickly began to play a key role in the implementation of the Treaty, providing the necessary data in the coal and steel production sector for the Member States of the Community. As European integration deepened and the number of areas covered by EU law increased, the need to produce European statistics in various new areas arose. This necessitated a stable legal framework that would allow Eurostat and the national statistical systems of the Member States to adapt and cope with current and future challenges.

The aim of this article is to present to the readers in a systematic way the transition of the European legal framework in the field of statistics from the beginning of the European integration to the last amendment of Regulation (EC) 223/2009 on European statistics in 2024. The article focuses mainly on the provisions foreseen in Regulation (EC) 223/2009 on European statistics and the two subsequent amendments of 2015 and 2024 respectively, as well as the reasons that led to their adoption. In this way, the reader gets an idea not only of the main European act in the field of statistics, but also of the historical context in which it was created and adopted.

Key words: European Statistical System, European Statistical System Committee, administrative sources, private data holders, confidential data

*Senior Expert in 'International Relations and Projects' Department, NSI, Sofia, Bulgaria, e-mail: raantonov@nsi.bg

РАЗВИТИЕ НА ЕВРОПЕЙСКАТА ПРАВНА РАМКА В ОБЛАСТТА НА СТАТИСТИКАТА

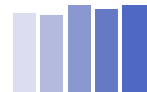
*Румен Антонов**

Въведение

Наличието на навременна, точна и съпоставима статистика в рамките на Европейския съюз (ЕС), както на национално, така и на европейско равнище, е основен инструмент при вземането на информирани политически решения. Европейската статистика също така се е превърнала в ежедневен средство за осведомяване на гражданите на ЕС. Въпреки голямата значимост на официалната европейска статистика, процесът по хармонизиране на отделните статистически системи на държавите членки и създаването на ефективни законодателни мерки в областта на статистиката отнема десетилетия. От самото начало на европейската интеграция, с подписването на Договора за създаването на Европейската общност за въглища и стомана (ЕОВС) през 1952 г., се предвижда създаването на независим изпълнителен орган, който да следи и подпомага процеса на изпълнение на договора. Към този орган, наречен Върховен орган на ЕОВС (ECSC High Authority), който е предшественикът на Европейската комисия, е създадено специализирано звено, отговорно за производството на статистика на Общностно ниво и по-специално в сектора за производство на въглища и стомана. През 1953 г. статистическото звено организира първата среща между статистическите служби на тогавашните държави - членки на Общността (Франция, Федерална република Германия, Италия, Белгия, Холандия и Люксембург). Почти веднага се отличават мащабните несъответствия между статистическите системи на държавите членки, които се оказват на съвсем различни нива на развитие, работещи по различни методи и използващи несъпоставими методологии. Оказва се почти невъзможно да се произведе адекватна и сравнима статистика на Общностно ниво. През първите десетилетия от съществуването си статистическата служба (след 1959 г. вече преименувана с настоящото си име Евростат) посвещава дейността си на създаването на хармонизирани статистически концепции и методи, които да бъдат използвани от държавите-членки на Общността, с цел изготвянето на съпоставими данни. Като резултат от тези усилия се създават някои от най-съществените инструменти на европейско ниво в областта на статистиката, а именно: „Система от интегрирани икономически сметки“ (ESA), „Статистическа класификация на икономическите дейности“ (NACE) и други.

През последното десетилетие на XX в. три събития изиграват ключова роля за признаването и затвърждаването на ключовата роля на Евростат в рамките на Европейската общност, а след 1993 г. - Европейския съюз. Първото събитие е подписването на договора от Маастрихт през 1992 г. (в сила от 1993 г.), който предвижда много по-голямо задълбочаване на европейската интеграция в политически, икономически и социален аспект. През същата 1993 г. започва да функционира и Единният пазар на ЕС, който предвижда свободното движение на хора, стоки, услуги и капитали в рамките на ЕС. В този контекст дейността на Евростат започва да се разширява, като е създадена специализирана система „Интрастат“, съдържаща данни за търговския стокообмен между държавите-членки на ЕС. В същото време Евростат започва да публикува редовни публикации и прессъобщения предвидени за широката общественост. Третото ключово събитие, което се отразява върху дейността на Евростат, е създаването на единната европейска валута - еврото. Тъй като критериите за конвергенция, които една държава членка трябва да покрие преди да приеме еврото за своя национална валута, се базират

* Старши експерт в отдел „Международни отношения и проекти“, НСИ, София, България, e-mail: raantonov@nsi.bg



върху статистически показатели, Евростат се оказва ключовият европейски орган в този процес. Освен стандарти и методи за хармонизация, Евростат започва да разработва и прилага и показатели, които да бъдат използвани от статистическите системи на държавите членки.

В резултат на засилващата се роля на Евростат, за първи път статистиката е включена в първичното право на ЕС с подписването на Договора от Амстердам през 1997 г. (в сила от 1999 година). Почти веднага след това се приема и първият европейски регламент в областта на статистиката - Регламент (ЕО) № 322/97 на Съвета от 17 февруари 1997 година относно статистиката на Общността. В него за първи път са заложили основните принципи на европейската статистика, принципа на статистическата поверителност и опазване на статистическата тайна, както и статистическата програма на ЕС и нейното изпълнение.

В началото на XXI в. европейската статистика се намира на съвсем различно равнище от това през 1952 година. Статистическите служби на държавите членки вече работят по хармонизирани европейски стандарти и методологии, а Евростат започва да публикува и разпространява свободно и безплатно статистически данни до обществеността. През 2005 г. се приема и Кодексът на европейската статистическа практика¹, който наброява 16 принципа за качеството на европейската статистика, които да бъдат прилагани и спазвани от държавите членки. Заедно с него се създава Европейският консултативен съвет за статистическо управление (ESGAB), който осигурява независим преглед на Европейската статистическа система (ЕСС) по отношение на прилагането на Кодекса на европейската статистическа практика. Създаден е и Европейският статистически консултативен комитет (ESAC) към Европейския парламент, имащ за цел да гарантира, че изискванията на потребителите, както и натоварването на респондентите и производителите на информация, се вземат предвид при разработването на статистическите програми. Въпреки постигнатия напредък започва да се очертава проблемът за преразглеждането на европейското право в областта на статистиката, предвид нарастващата роля на ЕС на световната сцена и мащабното увеличение на броя на държавите членки на ЕС, първо през 2004 г., а след това през 2007 година. В тази връзка през 2007 г. Европейската комисия публикува за обсъждане предложение за Регламент на Европейския парламент и на Съвета относно Европейската статистика, който ще стане известен като основният законодателен акт на ниво ЕС в областта на статистиката - Регламент (ЕО) № 223/2009.

1. Регламент (ЕО) № 223/2009 относно европейската статистика

а) Предпоставки за приемане на акта

Планове и концепции за бъдещото развитие на законовата основа на процеса по производство и разпространение на европейската статистика започват да се обсъждат още през 2003 г., когато се провеждат няколко заседания на специализирани работни групи с участието на държавите-членки. В рамките на тези работни групи се разглеждат два подхода, които могат да бъдат предприети на ниво ЕС. Първият предвижда правната рамка в областта на статистиката да бъде запазена в тогавашния ѝ вид, като многократно е оценявана като адекватна за текущите цели на европейската статистика и се състои в следните законодателни актове:

- Решение (ЕИО, Евратом) № 89/382 на Съвета от 19 юни 1989 г. за създаването на Статистическия програмен комитет на Европейските общности²;

¹ Пълният текст на Кодекса може да бъде намерен на следния адрес: https://www.nsi.bg/sites/default/files/files/pages/LegalBasis/Code-of-Practice_2017.pdf

² Пълен текст достъпен на: <https://eur-lex.europa.eu/legal-content/BG/ALL/?uri=CELEX%3A31989D0382>

- Регламент (ЕИО, Евратом) № 1588/90 на Съвета от 11 юни 1990 г. за предоставянето на поверителна статистическа информация на Статистическата служба на Европейските общности¹;
- Регламент (ЕО) № 322/97 на Съвета от 17 февруари 1997 г. относно статистиката на Общността²;
- Решение № 2367/2002/ЕС на Европейския парламент и на Съвета от 16 декември 2002 г. относно статистическата програма на Общността от 2003 г. до 2007 г.³;
- Регламент (ЕО, Евратом) № 1101/2008 на Европейския парламент и на Съвета от 22 октомври 2008 година за предоставянето на поверителна статистическа информация на Статистическата служба на Европейските общности⁴.

От цитираното е видно, че законодателната основа на европейската статистика е разделена в отделни взаимосвързани актове от вторичното право на ЕС. С разширяването на броя на държавите - членки на ЕС, продължаващото задълбочаване на европейската интеграция в контекста и на планираните изменения в първичното право на ЕС с договора от Лисабон (подписан 2007 г. в сила от 1 декември 2009 г.), както и със засилващата се роля на Европейския съюз като отделен субект в международните отношения, многократно се увеличава тежестта върху Евростат за предоставяне на статистически данни, което налага наличието на стабилна законова основа. Като естествен резултат от разширяващия се кръг от области, които се обхващат от правото на ЕС, започва да се появява и необходимост от съставянето на европейски статистически данни в изцяло нови области. В този контекст се появяват опасения, че в дългосрочен план, тогавашната правна рамка на европейската статистика би довела до задълбочаване на несъответствието между действащите общи разпоредби и разпоредбите касаещи специфични случаи. Това би довело до уронване на престижа на Европейската статистическа система и до загуба на общественото доверие в европейската статистика.

Дискусиите започват да се насочват към втория подход за бъдещото развитие, а именно цялостна ревизия, която има за основна цел да преразгледа и кодифицира съществуващата европейска правна рамка в областта на статистиката. Акцент се поставя и върху необходимостта от по-ясно сътрудничество, както в рамките на Европейската статистическа система, така и с европейски и международни партньори. След множество проведени срещи и обсъждания с европейските институции и държавите членки, през октомври 2007 г. Европейската комисия внася за разглеждане „Предложение за регламент на Европейския парламент и на Съвета относно европейската статистика“.

б) Цели и мотиви на предложението за регламент

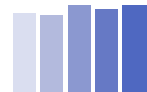
В предложението си за регламент, Европейската комисия признава значимата роля, която европейската статистика изпълнява в обществото и акцентира върху нуждата от кодификация на общностното право касаещо статистиката. На процеса „кодификация“ може да се гледа като на едно опростяване на европейската статистическа правна рамка, но не става въпрос за коренно и цялостно изменение на действащото право, а по-скоро за систематично надграждане, което да рефлектира на промените в обществото и да отговори на необходимостта от по-ясно регламентиране на ролята, функциите и структурата на Европейската статистическа система, като единствен механизъм за производство и разпространение на официална европейска статистика. Процесът на кодификация

¹ Пълен текст достъпен на: <https://eur-lex.europa.eu/legal-content/BG/ALL/?uri=CELEX%3A31990R1588>

² Пълен текст достъпен на: <https://eur-lex.europa.eu/legal-content/BG/ALL/?uri=CELEX%3A31997R0322>

³ Пълен текст достъпен на: <https://eur-lex.europa.eu/eli/dec/2002/2367/oj/eng>

⁴ Пълен текст достъпен на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=CELEX:32008R1101>



има за цел да опрости действащото право в областта на статистиката, като консолидира в един законодателен акт текстовете на гореизложените правни инструменти.

Наред с кодификацията се предлагат и редица изменения и допълнения към съществуващото право, сред които подобряване на статистическото управление чрез допълнително изясняване на ролята на националните статистически институти на държавите членки и на Евростат като неразделна и основна част от ЕСС. Подчертава се и нуждата от засилване на сътрудничеството, както в рамките на ЕСС, така и с други партньори и по-конкретно с Европейската централна банка (ЕЦБ) и Европейската система от централни банки (ЕСЦБ).

Разширява се приложното поле на Кодекса на европейската статистическа практика, за който е предвидено да се превърне в неизменна част от основното статистическо право на ЕС и да бъде използван като наръчник, подпомагащ дейността на настоящи и бъдещи статистици.

Заговаря се за „общ европейски подход към статистиката“, който поставя Евростат в центъра на процеса по производство на европейска статистика и цели намаляване на тежестта върху респондентите, статистическите институти на държавите членки и другите национални органи, които произвеждат и разпространяват европейска статистика, както и цялостна оптимизация на производствения процес, навременността и качеството на данните.

Доразвива се правната рамка за опазване на статистическата поверителност, чиято защита е от съществено значение за укрепване на доверието на обществото в Евростат и статистическите институти на държавите членки. Обсъжда се възможността за проявяване на известна гъвкавост по отношение на предоставянето на достъп до поверителни данни на различни изследователски организации с цел провеждане на статистически анализи за научни цели.

Въвеждането на новия акт има за цел да даде нов тласък в развитието на европейската статистика и ЕСС, да засили доверието на потребителите в работата на Евростат и статистическите служби на държавите членки и да подготви статистиката на ЕС за адаптация и справяне с настоящи и бъдещи предизвикателства.

в) Регламент (ЕО) № 223/2009 от 11 март 2009 г. относно европейската статистика¹

Окончателният вариант на предложението за регламент е публикуван в Официален вестник на Европейския съюз на 31 март 2009 г. под наименованието „Регламент (ЕО) № 223/2009 на Европейския парламент и на Съвета от 11 март 2009 година относно европейската статистика и за отмяна на Регламент (ЕО, Евратом) № 1101/2008 за предоставянето на поверителна статистическа информация на Статистическата служба на Европейските общности, на Регламент (ЕО) № 322/97 на Съвета относно статистиката на Общността и на Решение 89/382/ЕИО, Евратом на Съвета за създаване на Статистически програмен комитет на Европейските общности“. Регламентът съдържа 29 члена, разделени по смислови области в шест глави. В следващото изложение ще се спрем на основните и ключови нововъведения в Регламент (ЕО) № 223/2009 в съответните глави.

Глава I - общи разпоредби

В Глава I са заложили предмета, приложното поле и статистическите принципи, от които се ръководи процесът по производство на европейска статистика, както и дефиниции на използваните термини. За обхват на регламента е посочено създаването на правна рамка за разработването, изготвянето и разпространението на европейска статистика (чл. 1). Статистическите принципи (чл. 2),

¹ Пълен текст достъпен на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=CELEX:32009R0223>

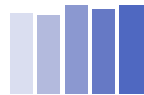
заложи в регламента, са аналогични на тези от Регламент (ЕС) № 322/97 относно статистиката на Общността, а именно безпристрастност, обективност, надеждност, статистическа поверителност и ефективност на разходите. Ключовият момент е добавянето на нов принцип, който в днешно време е признат за неизменна част от статистическия процес на статистическите институти на държавите членки и Евростат, а именно принципът за професионалната независимост. Под „професионална независимост“ трябва да се разбира, че статистиката *„трябва да се разработва, изготвя и разпространява по независим начин, особено що се отнася до подбора на техники, дефиниции, методологии и източници, които да бъдат използвани, и времето и съдържанието на всички форми на разпространение, без всякакъв натиск от политически или заинтересовани групи или от общностни и национални органи, без да се засяга институционалната структура, като например общностните или националните институционални или бюджетни разпоредби или определянето на статистическите потребности“*. Признавайки този принцип като част от правото на ЕС, европейската и националните статистики на държавите членки се поставят на много високо ниво в обществените отношения. Въведена е и препратка към Кодекса на европейската статистическа практика, където тези принципи са допълнително разработени.

Глава II - управление на статистиката

В тази глава изчерпателно са регламентирани ролите и задълженията, които членовете на Европейската статистическа системата изпълняват, а самата ЕСС е дефинирана като партньорство между статистически орган на ЕС - Евростат, националните статистически институти на държавите членки и другите национални органи, отговарящи за разработването, изготвянето и разпространението на европейска статистика (чл. 4). За „национален статистически институт на държава членка“, по смисъла на регламента се разбира организация, определена на национално ниво за отговорна за координацията на процеса по разработването, изготвянето и разпространението на европейска статистика (чл. 5), а Евростат е определен за структура на ниво ЕС, която е отговорна за вземането на решения относно процесите, статистическите методи, стандарти и процедури, както и времевата линия за публикуване на статистическите издания (чл. 6).

Съществена новост е създаването на Комитет на Европейската статистическа система (КЕСС), който се председателства от Евростат и е съставен от представители на статистическите институти на държавите членки, за които изрично е посочено, че са експерти по статистика (чл. 7). В днешно време участници в него обикновено са ръководителите на статистическите служби, придружени от началниците на респективните международни отдели. Заседанията на КЕСС се провеждат три пъти годишно. КЕСС представлява професионалното ръководство на ЕСС при разработването, изготвянето и разпространението на европейска статистика в съответствие с признатите статистически принципи. Евростат е длъжен да се консултира с КЕСС относно всеки въпрос, който касае предприемане на мерки на ниво ЕС за разработването, изготвянето и разпространението на европейска статистика. Към КЕСС могат да бъдат отнесени и редици други въпроси произтичащи от статистическия процес. Общите позиции на ЕСС по въпроси от особено значение за европейската статистика на международно равнище се подготвят и съгласуват в рамките на КЕСС под ръководството на Евростат (чл. 10).

В чл. 12 са изброени критериите за качество, които ръководят дейността по производство на европейска статистика: относимост, точност, актуалност, навременност, достъпност, съпоставимост и съгласуваност. За статистическите институти на държавите-членки се създава задължението да



изготвят и предоставят на Евростат доклади относно качеството на предадените данни, въз основа на които Евростат извършва цялостна оценка на качеството на европейската статистика.

Глава III - изготвяне на европейска статистика

В рамките на тази глава се предвижда приемането на двата водещи стратегически документа в областта на европейската статистика - Европейската статистическа програма (чл. 13) и годишната работна програма на Евростат (чл. 17). Европейската статистическа програма (ЕСП) е дългосрочен документ (настоящата програма е в сила за периода 2021 - 2027 г.), който създава рамка за разработването, изготвянето и разпространението на европейска статистика, както и основните приоритети за нейното развитие. Годишната работна програма на Евростат се съставя въз основа на ЕСП и съдържа конкретни дейности, които Евростат счита за приоритетни през съответната референтна година. Важно е да се отбележи ключовата роля, която държавите членки изпълняват в процеса по изготвянето на тези стратегически документи, тъй като и двата се разглеждат и съгласуват в рамките на КЕСС. Евростат също така представя пред държавите членки съответните доклади за изпълнението на ЕСП и годишната си програма.

Глава IV - разпространение на европейската статистика

Съгласно предвидените разпоредби в Глава IV разпространението на европейска статистика се извършва от Евростат, статистическите институти на държавите членки и другите национални органи, които произвеждат европейска статистика. Този процес се осъществява при пълно спазване на статистическите принципи и принципите за защита на статистическата поверителност, както и гарантиране на равен достъп до статистическа информация за всички потребители.

Глава V - статистическа поверителност

Въпросът за защитата на поверителните данни¹ е от изключително значение за гарантирането на доверието в обществото. По тази причина е предвиден строго регламентиран механизъм за събирането на този тип данни от статистическите институти на държавите членки и предаването им на Евростат. Общият принцип е, че тези данни могат да бъдат събирани, обработвани и предавани на Евростат единствено за статистически цели и в случай, когато без тяхната употреба не може да бъде изготвена надеждна и точна европейска статистика. В рамките на Евростат достъп до получените поверителни данни от държавите членки имат точно определен кръг длъжностни лица, в зависимост от конкретната работна област (чл. 22). Много малко са случаите, когато достъп до поверителни данни може да се предостави на допълнителен кръг от длъжностни лица, но дори когато това стане факт, отново може да бъде извършено единствено за статистически цели и даденото лице/лица остават обвързани със задълженията произтичащи от защитата на поверителните данни, дори след преустановяване на респективните им служебни функции. В допълнение е предвидено правомощие на държавите членки и Евростат за санкциониране на всякакви установени злоупотреби с поверителни данни.

Същественото нововъведение в процеса по защитата на поверителни данни е възможността за предоставяне на достъп до този тип данни на изследователски организации, извършващи

¹ По смисъла на чл. 3 ал. 7 от Регламент (ЕО) № 223/2009 под „поверителни данни“ се разбира данни, които позволяват разпознаването на статистическите единици, независимо дали пряко или непряко, като по този начин се оповестява индивидуална информация. За да се определи дали дадена статистическа единица е разпознаваема, трябва да се вземат под внимание всички съответни средства, които биха могли в разумна степен да бъдат използвани от трета страна за разпознаването на ѝ.

статистически анализи за научни цели (чл. 23). Предпоставката е поверителните данни да позволяват единствено непряко разпознаване на статистическите единици. Правото за предоставяне на достъп до този тип данни на изследователски организации е признато на всички членове на ЕСС, но с добавеното изискване към Евростат за получаване на изрично съгласие от определена държава-членка за предоставяне на предадени от нея поверителни данни на трети страни.

В Глава V се признава изключителното право на достъп на Евростат, статистическите институти на държавите членки и другите национални органи до източници на административни данни, поддържани от другите органи на публичната администрация в респективните им държавни (чл. 24). Въвеждането на тази разпоредба в правото на ЕС показва стремежа на Европейската комисия за цялостно намаляване на тежестта върху респондентите и статистическите служби на държавите членки, съответно и другите национални органи, които произвеждат и разпространяват европейска статистика. В допълнение е регламентирана възможността за набавянето на данни от публични източници, които са обществено достъпни съгласно националното законодателство на респективната държава членка (чл. 25), като този тип данни не могат да бъдат счестени за поверителни.

2. Изменения и допълнения на Регламент (ЕО) № 223/2009

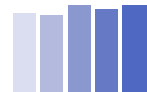
Регламент (ЕО) № 223/2009 е изменян два пъти до момента, съответно през 2015 и през 2024 година. Всяко едно от двете изменения разширява и надгражда действащата правна рамка в различни направления. Описването на целия мащаб на измененията ще надхвърли целите на настоящия труд, затова ще се спрем на най-съществените нововъведения в основния европейски законодателен акт в областта на статистиката, започвайки от първото изменение от 2015 година.

а) Регламент (ЕС) № 2015/759 на Европейския парламент и на Съвета от 29 април 2015 година за изменение на Регламент (ЕО) № 223/2009 относно европейската статистика¹

Разговори за преразглеждане на предвидената правна рамка в Регламент (ЕО) № 223/2009 започват да се водят още през 2011 г. в контекста на Световната рецесия (2008 - 2009 г.), когато отново се откроява ключовата роля на статистиката за предприемане на информирани политически решения. Появява се силен стремеж от страна на Евростат и държавите членки за засилване на професионалното управление в рамките на ЕСС за да се отговори по адекватен начин на нуждата от данни, както в рамките на политическите среди, така и в обществото. Отново се очертава проблемът за безусловното осигуряване на професионална независимост на европейско и национално равнище на служителите на Евростат, националните статистики на държавите членки и другите национални органи, произвеждащи европейска статистика. Дискусиите обхващат и константния стремеж на ниво ЕС за намаляване на тежестта върху респондентите и статистическите институти. В резултат от множеството проведени срещи и дискусии, Европейската комисия публикува предложението за изменение през месец април 2012 година, което се стреми да намери решения в споменатите области.

По отношение на засилването на професионалното ръководство в рамките на ЕСС с изменението се цели допълване и разширяване на съществуващите разпоредби относно ролята на националните статистики в рамките на съответните национални статистически системи. Националните статистически институти са определени за единствените отговорници на национално равнище за координация на статистическата дейност в респективната държава членка (чл. 5, ал. 1). Има се предвид координирането на статистическите програми, докладите за наблюдение на качеството,

¹ Пълен текст достъпен на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=celex%3A32015R0759>



методологията и предаването на данни. Статистическите институти на държавите членки са длъжни да оказват съдействие и да предоставят насоки на другите национални органи, които произвеждат и разпространяват европейска статистика. Когато се касае за статистически данни събирани от национални централни банки (НЦБ) за статистическите служби се предвижда провеждането на тясно сътрудничество с НЦБ, в съответствие с национални договорености, с цел изготвянето на навременна и съгласувана европейска статистика.

Друг аспект на този процес е изричното задължение за националните правителства на държавите членки безусловно да гарантират професионалната независимост на всички служители на статистическите институти и други национални органи, включително и на ръководителите им. За целта в изменението се акцентира върху водещата роля на ръководителите на статистическите институти на държавите членки, чиито отговорности и правомощия са изчерпателно изложени (чл. 5а). Ръководителите на статистическите служби са еднолично отговорни за вземането на решения относно процесите, статистическите методи, статистическите стандарти и процедури, както и за съдържанието и времето за оповестяване на статистически съобщения и публикации на изготвената от съответната статистическа служба европейска статистика. В допълнение ръководителите на статистически институти са оправомощени да вземат решения по всички въпроси свързани с тяхното вътрешното управление и по-специално относно изпълнението на бюджета на съответната институция. Относно координацията на съответната национална статистическа система, ръководителите на статистически институти са компетентни относно изготвянето на насоки, които са приложими и задължителни, както за съответната статистическа служба, така и за другите национални органи, които произвеждат и разпространяват европейска статистика. Във външен план ръководителите на статистически институти представляват съответните национални статистически системи в рамките на ЕСС и други международни организации и форуми (напр. Статистическа комисия на ООН).

За да могат да изпълняват предвидените задължения на ръководителите на статистически институти се признава абсолютна независимост в действията, като те не могат да търсят или получават указания от съответните национални правителства или други институции на национално, европейско и международно равнище. За държавите членки се създава задължение за определяне на прозрачни механизми за назначаването, прехвърлянето и освобождаването на ръководителите на статистически институти, които се базират единствено на професионални критерии. Ръководителите имат възможността свободно и публично да изказват мнения по въпроси свързани със статистическата дейност на съответната статистическа служба, както и по отношение на разпределения бюджет за институцията.

По аналог на установените права и задължения на ръководителите на статистически институти са регламентирани и правомощията на генералния директор на Евростат (чл. 6а). Предвиден е същият безусловен гарант за професионална независимост на генералния директор при изпълнение на предвидените задължения, както и прозрачна процедура за назначаване и освобождаване от длъжност.

Второто съществено изменение обхваща разширяването на предвиденото в Регламент (ЕО) № 223/2009 право на достъп до данни от административни източници за статистически цели, с цел намаляване на тежестта върху респондентите, статистическите служби и другите национални органи. Съгласно изменението статистическите институти следва да бъдат включени в дискусиите и в процеса по взимане на решение на национално ниво относно създаването, разработването и прекратяването на административни източници на данни, които биха могли да се използват

за производството на статистически данни (чл. 17а). Статистическите институти координират дейностите по стандартизация и получават съответните метаданни относно извлечените данни от административни източници.

С изменението се въвежда изричното задължение на Европейската комисия и националните правителствата на държавите членки за предприемането на всички необходими действия за запазване на доверието в европейската статистика (чл. 11, ал. 3). В тази връзка правителствата на държавите членки изготвят и приемат на най-високо ниво така наречените „Ангажменти във връзка с доверието в статистиката“, които съдържат конкретни политически отговорности към респективните статистически служби и националните статистически системи. Ангажментът на правителството на Република България е приет и подписан през месец декември 2021 година¹. Тези документи са обект на постоянна проверка от страна на Европейската комисия, като държавите членки подготвят и изпращат годишни доклади за тяхното приложение.

б) Регламент (ЕС) № 2024/3018 на Европейския парламент и на Съвета от 27 ноември 2024 година за изменение на Регламент (ЕО) № 223/2009 относно европейската статистика²

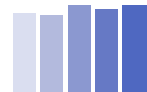
Второто и най-мощно досега изменение на Регламент (ЕО) № 223/2009 се представя за обсъждане от Европейската комисия през лятото на 2023 година. Основната цел на изменението е да адаптира и подсили готовността на европейската статистическа правна рамка за адекватен отговор на настоящи и бъдещи предизвикателства, както и да подобри капацитета на ЕСС за предоставянето на по-актуална и точна европейска статистика. До голяма степен, необходимостта от повторно изменение на статистическата правна рамка на ниво ЕС се налага от динамичните събития, настъпили в началото на третото десетилетие на XXI век. Пандемията от COVID-19, последвалата след нея икономическа и социална криза, както и руската военна агресия срещу Украйна, крайно изострят нуждите и очакванията за по-навременна и детайлна европейска статистика, необходима за вземането на информирани политически решения на европейско ниво при кризисни ситуации. В този контекст ЕС разработва и установява няколко механизма за справяне с потенциални кризисни ситуации, сред които е създаденият Европейския орган за готовност и реакция при извънредни здравни ситуации (HERA)³, в рамките на който европейската статистика отново изиграва ключова роля. Европейска комисия акцентира върху нуждата от навременна реакция в рамките на ЕСС при бъдещи кризисни ситуации.

Скоростните темпове на развитие на новите технологии и на цифровата трансформация създават както нови възможности за модернизирание на процеса по производство на статистика, така и много неизвестни. На преден план възниква въпросът за мястото на статистическите институти на държавите членки в рамките на цифровата трансформация, където частният сектор взема превес. В резултат на цифровизацията се появяват множество нови цифрови източници на данни, които биха могли да бъдат използвани за производство на европейска статистика. Проблемът е установяването на тяхната надеждност, тъй като новите технологии позволяват наличието на значителен брой неофициални канали за разпространение на статистически данни. Същото се отнася и за прекомерната употреба на изкуствен интелект (ИИ), като отраженията от това върху процеса по производство на статистика предстои тепърва да бъдат оценени.

¹ Пълен текст достъпен на: https://www.nsi.bg/file/25265/CoC_BG_2021.pdf

² Пълен текст достъпен на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=celex%3A32024R3018>

³ За повече информация относно HERA виж: https://commission.europa.eu/about/departments-and-executive-agencies/health-emergency-preparedness-and-response-authority_en?prefLang=bg&etrans=bg



В този контекст става ясно, че настоящите методи за събиране и обработка на статистически данни не могат да отговорят на растящите нужди и очаквания на ползвателите. Затова Комисията си поставя за цел да засили ролята на Евростат, на статистическите институти на държавите членки и на другите национални органи в качеството им на единствени официални производители на европейска статистика, като насърчи свободното споделяне на данни и позволи на ЕСС напълно да се възползва от възможностите, произтичащи от цифровата трансформация. Тези процеси трябва да бъдат изпълнени изрично в съответствие с регламентираните мерки за опазване на личните данни в Общия регламент относно защитата на данните (GDPR) и принципа за защита на поверителните данни.

В изменението е предвидена актуализация на съществуващите дефиниции в Регламент (ЕО) № 223/2009, която има за цел да отрази цифровата трансформация (чл. 1). Добавени са дефиниции за данни, метаданни и държател на данни, който може да бъде всяко юридическо или физическо лице, имащо правото, в съответствие с законодателството на ЕС и респективното национално право, да управлява и предоставя данни, получени в резултат на съответната негова дейност. В допълнение към определението за „използване за статистически цели“ са включени научни и изследователски дейности. Новият подход към източниците на данни е, че те могат да бъдат използвани еднолично или в комбинация с данни от други източници за производство на европейска статистика. По този начин за статистическите служби се създава възможността да търсят и използват най-подходящия и ефективен източник или група от източници при разработването на статистически данни.

В тази връзка, с цел подобряване на ефикасността и ефективността на ЕСС, в изменението се предвижда разширяването на предвидения достъп до административни данни, но също така се регламентира правото на достъп до данни от частни държатели. Както е посочено по-горе, с изменението от 2015 г. в Регламент (ЕО) № 223/2009 се регламентира изричният достъп на Евростат, статистическите институти и другите национални органи до административни източници на данни за статистически цели. На статистическите институти се предоставя значителна роля и участие в дискусиите по създаването, разработването и прекратяването на подобни източници на данни. С настоящото изменение се създава задължение за всички национални публични и смесени публично-частни органи, които поддържат източници на административни данни или всякакъв друг тип данни, които биха могли да бъдат необходими за разработването на европейска статистика, да предоставят навременен и безвъзмезден достъп на статистическите служби и другите национални органи (чл. 17а). В допълнение статистическите институти имат право не само да използват тези данни, но и да ги интегрират, заедно с релевантните метаданни. Евростат се ползва от същите привилегии за достъп по отношение на поддържаните бази данни от органите и агенциите на ЕС.

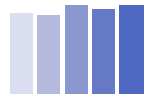
Най-същественото изменение е въвеждането на задължение за частните държатели на данни да предоставят данни за разработването и разпространението на европейска статистика (чл. 17б). Процесът е ясно регламентиран, като всяка статистическа служба на държава членка може да подаде искане за предоставяне на данни до частен държател на данни единствено за статистически цели. Условието е, че тези данни са строго необходими за разработването на европейска статистика и не могат да бъдат набавени по друг начин. Статистическите служби на държавите членки са длъжни да определят в исканията си типа данни и метаданни, които се изискват, честота и краен срок за предаването им, както и съответната необходимост от това предаване. След това се предвижда стартирането на диалог между двете страни с цел сключване на споразумение. В качеството си на координатор на респективната национална статистическа система, статистически институт на

държава членка може да подаде искане за предоставяне на данни от частен държател от името на друг национален орган, спазвайки същите процедурни изисквания. В по-редки случаи и Евростат може да изиска достъп до данни от частни държатели, най-вече когато се касае за предприятия, опериращи в рамките на целия ЕС. В рамките на този процес статистическите институти и Евростат работят в тясно сътрудничество за да избегнат прекомерна тежест върху частните държатели на данни. Предвижда се възможността Евростат да създаде сигурна инфраструктура за улесняване на трансфера на исканите данни от частните държатели, с изричното съгласие на съответната държава-членка. Статистическите институти и Евростат също така са длъжни да предприемат всички подходящи мерки за защита на статистическата поверителност и опазване на търговската тайна при получаване на данни от частни държатели. Когато за предоставянето на поисканите данни се изискват допълнителни дейности за обработка, статистическите институти и Евростат могат да предоставят компенсация на частния държател на данни. Важно е да се отбележи, че компенсацията може да бъде предоставена единствено за конкретните допълнителни дейности, а не за цялостното предоставяне. При неизпълнение от страна на частния държател на данни на договорените отговорности се предвижда налагането на съответната административна глоба.

Изменението цели да засили капацитета на ЕСС за адекватен и бърз отговор на неотложни нужди от предоставяне на данни, необходими за вземането на информирани политически решения в случай на криза (чл. 16а). Евростат има правото да предприеме спешни статистически дейности извън съответната Европейска статистическа програма. Дейностите се изпълняват в сътрудничество със статистическите служби на държавите членки и другите национални органи и могат да включват разработването на нови типове статистически данни, предоставянето на нови статистически показатели и разработването на методологически насоки. Преди да предприеме подобни мерки Евростат е задължен да се консултира с КЕСС и да вземе предвид резултатите от проведените дискусии. Участието на държавите членки в тези извънредни действия се извършва на доброволна основа. При участие държавата членка е длъжна да спазва установените от Евростат изисквания, срокове, честота и критерии за качество на предадените национални данни.

С изменението се въвежда и изцяло нова разпоредба, която касае въпроса за статистиката, която е в процес на разработване или т.нар. „експериментална статистика“ (чл. 17ж). Целта е да се насърчи непрекъснатото въвеждане на иновации и употребата на съвременните технологии в процеса по производство на европейска статистика, както и разработването на нови статистически продукти. При изрично съгласие от национален статистически институт или друг национален орган, Евростат може да публикува и разпространява експериментална статистика, но с уточнението, че тя е в процес на разработване. Евростат може да започне процес по разработване на нови статистически продукти, приложими за цялата ЕСС. Това действие може да бъде извършено единствено в тясно сътрудничество с КЕСС.

Освен въвеждането на иновации изменението цели и да подобри процеса по споделяне на данни в рамките на ЕСС и между ЕСС и Европейската система от централни банки за статистически цели (чл. 17е). Споделянето на неповерителни данни е предвидено да се извършва при поискване, а за поверителните - единствено, когато това е необходимо за разработването и разпространението на европейска статистика (чл. 21). Евростат трябва да получи изричното одобрение от държавата - членка, предоставила поверителните данни, преди да може те да бъдат споделени. По този начин се цели разширяването на набора от източници на данни в рамките на ЕСС.



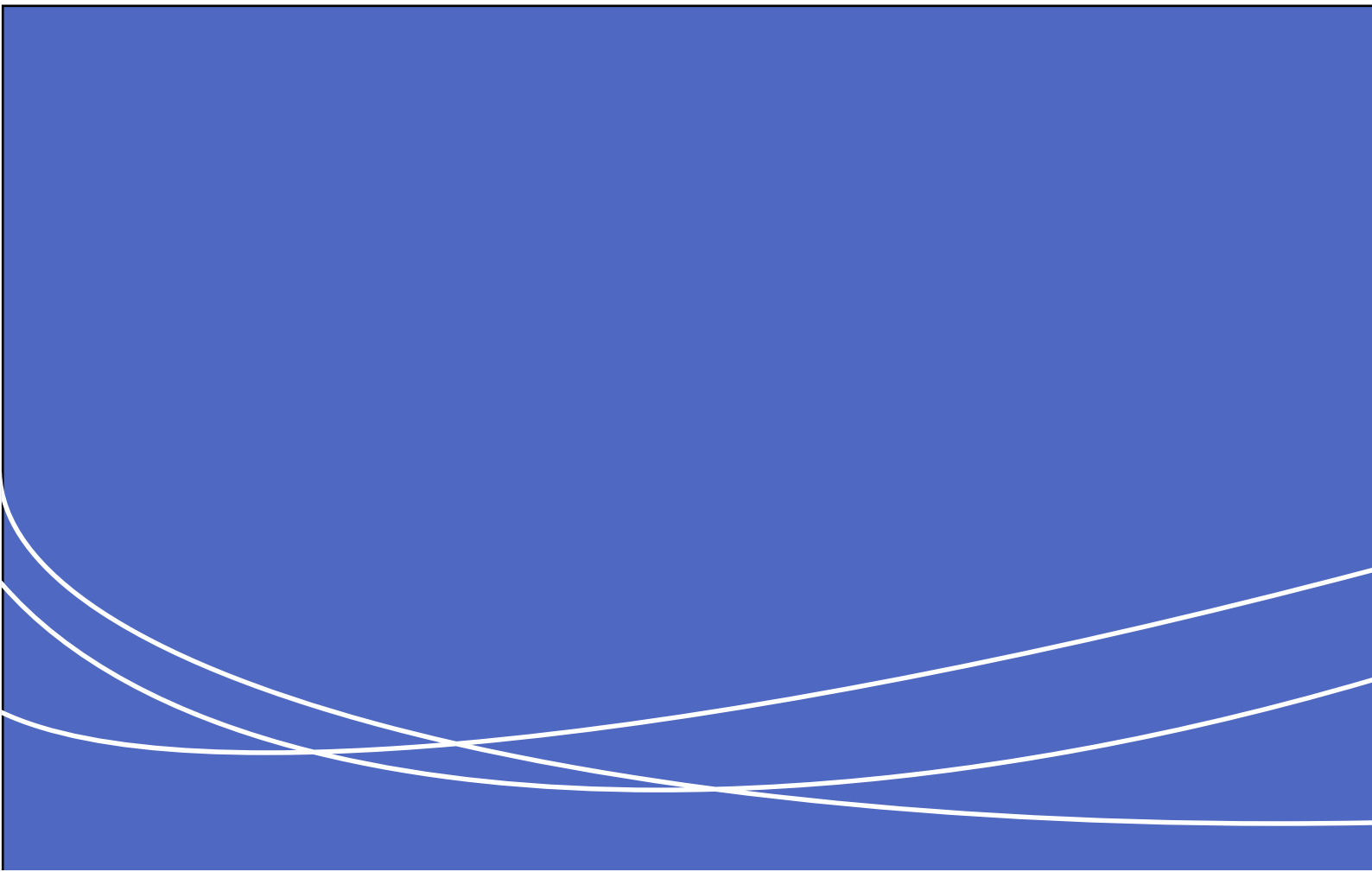
Заклучение

На този етап не е ясна бъдещата посока на развитие на европейската правна рамка в областта на статистиката. Бързите темпове на развитие на цифровите технологии превръщат данните и информацията в най-ценния ресурс на XXI век. Последното изменение на Регламент (ЕО) № 223/2009 създава и регламентира множество нови възможности за производство и разпространение на европейска статистика, в лицето на предвидения достъп за членовете на ЕСС до данни от частни държатели. В същото време, чрез насърчаването на внедряването на нови технологии в процеса по производство на европейска статистика, както и разработването на експериментална статистика, цели едно непрестанно модернизиране на националните статистически системи на държавите членки. Теоретично възможностите изглеждат наистина безкрайни. Въпросът, който предстои да бъде решен е практическото приложение на предвидените промени. В момента на писане на настоящата статия, въпросът все още продължава да се разисква, както на национално, така и на европейско ниво, особено в Комитета на ЕСС. Необходимо е да бъде съставен и съгласуван един общ европейски подход, в името на запазването на релевантността и значимостта на официалната европейска статистика.

ИЗПОЛЗВАНИ ИЗТОЧНИЦИ:

1. Alastalo, Marja. **The History of Eurostat**, 2022 г. Достъпно на: <https://ehne.fr/en/encyclopedia/themes/material-civilization/digital-europe/history-eurostat> (достъп 21.07.2025 г.)
2. Eurostat. **Our history REVAMP**. Достъпно на: <https://ec.europa.eu/eurostat/about-us/history-re-vamp> (достъп 21.07.2025 г.)
3. Eurostat. **History**. Достъпно на: <https://ec.europa.eu/eurostat/about-us/history> (достъп 21.07.2025 г.)
4. **Решение на Съвета от 19 юни 1989 година за създаване на Статистически програмен комитет на Европейските общности**. Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/ALL/?uri=CELEX%3A31989D0382> (достъп 21.07.2025 г.)
5. **Регламент (ЕВРАТОМ, ЕИО) № 1588/90** на Съвета от 11 юни 1990 година за предоставянето на поверителна статистическа информация на Статистическата служба на Европейските общности. Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/ALL/?uri=CELEX-%3A31990R1588> (достъп 21.07.2025 г.)
6. **Регламент (ЕО) № 322/97** на Съвета от 17 февруари 1997 година относно статистиката на Общността. Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/ALL/?uri=CELEX-%3A31997R0322> (достъп 21.07.2025 г.)
7. **Решение № 2367/2002/ЕС** на Европейския парламент и на Съвета от 16 декември 2002 г. относно статистическата програма на Общността от 2003 г. до 2007 г. Достъпно на: <https://eur-lex.europa.eu/eli/dec/2002/2367/oj/eng> (достъп 21.07.2025 г.)
8. **Регламент (ЕО, Евратом) № 1101/2008** на Европейския парламент и на Съвета от 22 октомври 2008 година за предоставянето на поверителна статистическа информация на Статистическата служба на Европейските общности. Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/TX-T/?uri=CELEX:32008R1101> (достъп 21.07.2025 г.)
9. **Предложение за регламент** на европейския Парламент и на Съвета относно европейската статистика. Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=CELEX-%3A52007PC0625> (достъп 21.07.2025 г.)
10. **Регламент (ЕО) № 223/2009** на Европейския парламент и на Съвета от 11 март 2009 година относно европейската статистика и за отмяна на Регламент (ЕО, Евратом) № 1101/2008 за предоставянето на поверителна статистическа информация на Статистическата служба на Европейските общности, на Регламент (ЕО) № 322/97 на Съвета относно статистиката на Общността и на Решение 89/382/ЕИО, Евратом на Съвета за създаване на Статистически програмен комитет на Европейските общности. Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=CELEX:32009R0223> (достъп 21.07.2025 г.)
11. **Предложение за регламент** на Европейския парламент и на Съвета за изменение на Регламент (ЕО) № 223/2009 относно европейската статистика (17.04.2012 г.). Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=CELEX%3A52012PC0167> (достъп 21.07.2025 г.)
12. **Регламент (ЕС) 2015/759** на Европейския парламент и на Съвета от 29 април 2015 година за изменение на Регламент (ЕО) № 223/2009 относно европейската статистика. Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=celex%3A32015R0759> (достъп 21.07.2025 г.)
13. **Предложение за регламент** на Европейския парламент и на Съвета за изменение на Регламент (ЕО) № 223/2009 относно европейската статистика (10.07.2023 г.). Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=CELEX%3A52023PC0402> (достъп 21.07.2025 г.)
14. **Регламент (ЕС) 2024/3018** на Европейския парламент и на Съвета от 27 ноември 2024 година за изменение на Регламент (ЕО) № 223/2009 относно европейската статистика. Достъпно на: <https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=celex%3A32024R3018> (достъп 21.07.2025 г.)
15. **Евростат. Кодекс на европейската статистическа практика**, 2005 г. (ревизиран 2011 г. и 2017 г.). Достъпно на: https://www.nsi.bg/sites/default/files/files/pages/LegalBasis/Code-of-Practice_2017.pdf (достъп 21.07.2025 г.)

СТАТИСТИЧЕСКИ ИЗСЛЕДВАНИЯ И АНАЛИЗ НА ДАННИ
STATISTICAL RESEARCH AND DATA ANALYSIS



РЕПРЕЗЕНТАТИВНИ ИЗСЛЕДВАНИЯ ПО ВРЕМЕ НА ПРЕБРОЯВАНЕТО НА НАСЕЛЕНИЕТО И ЖИЛИЩНИЯ ФОНД НА 4 ДЕКЕМВРИ 1985 ГОДИНА

*Соня Златанова**

РЕЗЮМЕ В периода от 4 до 14 декември 1985 г. се провежда четиринадесетото преброяване на населението и жилищния фонд в България. В своята почти стогодишна история Преброяването има за цел да отговори на съвременните потребности от информация за човешките ресурси в страната. То достоверно и пълно отразява както настъпилите промени в демографските, образователно-квалификационните и социално-икономическите структури на българското население, така и процесите и социалните механизми, които ги обуславят.

Обект на настоящата статия са три социологически изследвания проведени по време на преброяването:

I. Раждаемост и репродуктивно поведение на жените

Събрани са ценни данни, които дават възможност да се анализира репродуктивното поведение на жените, както и факторите, които го обуславят в съответния етап на общественно-икономическото развитие на България.

II. Вътрешна миграция и миграционно поведение на населението

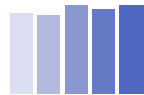
За първи път, наред с количествените измерения на този социално-демографски процес, се акцентира на широкото изследване на поведенческите аспекти на миграцията. В този смисъл проучването предоставя много полезна информация на органите за социално управление.

III. Трудови възможности и реализация на заетото население

Ценностната ориентация към труда е определена от мястото на труда в системата от житейски ценности, в която са включени осем категории: общественно положение, материално благополучие, усъвършенстване на способностите чрез труд, да бъдеш полезен на хората с труда си, семейството, лек и безгрижен живот, живот без рискове и конфликти и други ценности.

Ключови думи: Преброяване 1985, репрезентативни изследвания, раждаемост, миграция, репродуктивно поведение на жените, вътрешна миграция, миграционно поведение на населението, трудови възможности, реализация на заетото население

* Младши експерт в отдел „Публикации, библиотека и дигитални продукти“ на НСИ, e-mail: SZlatanova@NSI.bg.



РЕПРЕЗЕНТАТИВНЫЕ ИССЛЕДОВАНИЯ ВО ВРЕМЯ ПЕРЕПИСИ НАСЕЛЕНИЯ И ЖИЛИЩНОГО ФОНДА 4 ДЕКАБРЯ 1985 ГОДА

*Соня Златанова**

РЕЗЮМЕ С 4 по 14 декабря 1985 года в Болгарии была проведена четырнадцатая перепись населения и жилищного фонда. За почти столетнюю историю перепись стремилась удовлетворить актуальные потребности в информации о человеческих ресурсах страны. Она достоверно и всесторонне отражает как изменения в демографической, образовательной и социально-экономической структурах населения Болгарии, так и процессы и социальные механизмы, обуславливающие эти изменения.

Объектом настоящей статьи являются три социологических исследования, проведённые в рамках переписи:

I. Рождаемость и репродуктивное поведение женщин

Были собраны ценные данные, позволяющие проанализировать репродуктивное поведение женщин и факторы, влияющие на него на данном этапе социально-экономического развития Болгарии.

II. Внутренняя миграция и миграционное поведение населения

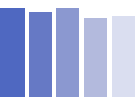
Впервые, наряду с количественными характеристиками этого социально-демографического процесса, особое внимание было уделено поведенческим аспектам миграции. В этом смысле исследование предоставляет важную информацию для органов социального управления.

III. Трудовой потенциал и реализация занятых лиц

Ценностные ориентации в отношении труда рассмотрены через призму его места в системе жизненных ценностей, включающей восемь категорий: социальный статус, материальное благополучие, самореализация через труд, польза обществу, семья, лёгкая и беззаботная жизнь, жизнь без рисков и конфликтов, а также другие личные ценности.

Ключевые слова: Перепись 1985 года, репрезентативные исследования, рождаемость, миграция, репродуктивное поведение женщин, внутренняя миграция, миграционное поведение населения, трудовые возможности, реализация занятости

* Младший эксперт отдела «Публикации, библиотека и цифровые продукты» Национального статистического института, e-mail: SZlatanova@NSI.bg.



REPRESENTATIVE SURVEYS DURING THE POPULATION AND HOUSING CENSUS OF DECEMBER 4, 1985

*Sonya Zlatanova**

SUMMARY In the period from 4 to 14 December 1985, the fourteenth Population and Housing Census in Bulgaria was conducted. In its nearly one-hundred-year history, the Census aimed to meet the contemporary needs for information on the country's human resources. It reliably and comprehensively reflects both the demographic, educational, and socio-economic transformations of the Bulgarian population, as well as the processes and social mechanisms that determine them.

This article focuses on three sociological surveys conducted during the Census:

I. Fertility and Reproductive Behaviour of Women

Valuable data were collected, enabling analysis of women's reproductive behaviour and the factors influencing it at the given stage of Bulgaria's socio-economic development.

II. Internal Migration and Migration Behavior of the Population

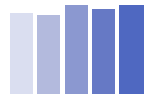
For the first time, alongside the quantitative dimensions of this socio-demographic process, significant attention was given to the behavioural aspects of migration. The study thus provides highly useful information to the social management bodies.

III. Labour Potential and Employment Fulfillment

Labour-related values were examined through the role of work within a broader system of life values, including eight categories: social status, material well-being, self-development through work, to be useful for society through your work, family, easy and carefree life, life without risks and conflicts and other values.

Keywords: Census 1985, representative surveys, birth rate, migration, reproductive behaviour of women, internal migration, migration behaviour of the population, labour opportunities, employment realization

* Junior Expert in Activities for Improvement of the Statistical Literacy, Web Content and Digitalization Department, National Statistical Institute, e-mail: SZlatanova@NSI.bg.



РЕПРЕЗЕНТАТИВНИ ИЗСЛЕДВАНИЯ ПО ВРЕМЕ НА ПРЕБРОЯВАНЕТО НА НАСЕЛЕНИЕТО И ЖИЛИЩНИЯ ФОНД НА 4 ДЕКЕМВРИ 1985 ГОДИНА И НЕПОСРЕДСТВЕНО СЛЕД НЕГО

*Соня Златанова**

Въведение

Удовлетворяването на растящите потребности от информация за измерване на реалното проявление на демографските и социални процеси и на влиянието им за формиране на различни демографски структури води до включване в програмата на преброяването на населението и жилищния фонд през 1985 г. на система от репрезентативни изследвания. С помощта на тези изследвания се разширява и обогатява информацията, съдържаща се в картите за изчерпателно преброяване и се изучава мнението на населението по редица важни за нашата страна проблеми като раждаемостта, миграцията, трудовата и социалната активност, образованието и културата.

Да се включи всичко това в една програма се счита за нецелесъобразно поради претоварването, което ще се получи върху преброителите и върху населението. Ето защо се взема решение допълнителните изследвания да се оформят в шест самостоятелни програми, но обвързани помежду си и имащи за ядро общото преброяване на населението и жилищния фонд. Тези програми са: раждаемост и репродуктивно поведение, миграция и миграционно поведение; трудови възможности и реализация; промени, проблеми и житейски планове на личността; град - село 86 и здравно състояние на населението.

В процеса на разработването на проектите за тези изследвания се оформят две становища относно най-подходящото време за провеждането им - по време на общото преброяване или след него.

Първият начин има предимства преди всичко от организационен характер - без допълнителни средства, а само със създадената организация за общото преброяване и с едно посещение на преброителя в дома на наблюдаваните лица, се попълват анкетните карти за изчерпателната и извадковата форма на наблюдение. Не е необходимо лицата да се разпитват допълнително за демографските, социално-икономическите характеристики и жилищните условия, тъй като информацията за тях се съдържа в картите за изчерпателното наблюдение. Така се намалява продължителността на интервюто в сравнение с едно самостоятелно изследване. Това становище обаче има и недостатъци - например, че не може и не бива да се отива към голяма детайлност на заложените признаци в програмата за наблюдение, което би утежило традиционното преброяване и би влошило качеството му, както и ограниченото време за провеждане на изследването - времето за провеждане на преброяването.

Второто становище има важното предимство, че дава възможност да се разгърнат програмите на изследванията, без това да влияе на качеството на преброяването. Разполага се с повече време за провеждането им, което дава възможност за повторни посещения и уточнения, което естествено влияе положително върху качеството и обхвата на наблюдението.

След обстойни обсъждания се стига до решението четири от изследванията - за раждаемостта; за миграцията, за трудовите възможности и реализация, за проблемите, промените и житейските

* Младши експерт в отдел „Публикации, библиотека и дигитални продукти“ на НСИ; e-mail: SZlatanova@NSI.bg.

планове на личността - да се проведат по време на преброяването на населението и жилищния фонд, а останалите две - „Градът и селото - 86“ и „Здравното състояние на населението“ - да се проведат в един кратък срок след преброяването.

Четиринадесетото преброяване на населението и жилищния фонд се провежда в периода 4 - 14 декември 1985 година. Със своя над стогодишен опит българската статистика провежда едно нестандартно преброяване, което отговаря на съвременните потребности от информация за човешките ресурси в страната. То достоверно и пълно отразява както настъпилите промени в демографските, образователно-квалификационните и социално-икономическите структури на българското население, така и процесите и социалните механизми, които ги обуславят.

Обект на настоящата статия са три социологически изследвания проведени по време на преброяването и непосредствено след него:

I. „Личността - промени, проблеми и житейски планове” - проведено паралелно с преброяването;

II. „Градът и селото - 86” - проведено непосредствено след традиционното преброяване, но обвързано с него;

III. „Здравно състояние на населението” - с това изследване се изчерпват социологическите изследвания свързани с преброяването.

Широко е приложен социологическият подход в българската статистика. Успешно проведеното преброяване сполучливо съчетава статистическата информация от класическото преброяване и информацията от социологическите изследвания. Това дава възможност за задълбочен научен анализ на основните явления и процеси, които протичат в обществото. Събрана е комплексна информация за човешките ресурси в обществото, за разкриване на процесите и явленията, за тенденцията и закономерностите в демографското и културното развитие на народа, както и за социално-икономическото развитие на страната.

Основно изложение

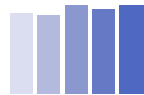
Личността - промени, проблеми и житейски планове

Цел и задачи

Целта на социологическото изследване е да се съдейства за изясняване на проблемите, потребностите и интересите на личността.

За постигане на основната цел изследването си поставя за разрешаване следните конкретни задачи:

- да се установят основните промени в живота на личността през последните десет години;
- да се изяснят проблемите и трудностите, които има личността в настоящия етап от развитието на обществото;
- да се набележат целите и задачите, които си поставя съвременната личност за реализиране в следващия десетгодишен период;
- да се изследва удовлетвореността както от отделни страни на живота, така и общата оценка за живота през последните десет години;
- да се изследва влиянието на основните социално-икономически характеристики на личността



както върху промените, проблемите и житейските планове, така и върху удовлетвореността от отделни страни на живота и общата оценка за живота;

- да се проучи влиянието на удовлетвореността от отделните страни на живота за общата оценка за живота.

Обект и предмет

За обект на проучването е възприета съвкупността на пълнолетното население. Това е населението, което към критичния момент на преброяването на населението от 04.12.1985 г. е навършило 18 години т.е. родените преди 04.12.1967 година. Това разграничаване е прието, за да се обхване и най-младото население, сред което за проучвания десетгодишен период (1976 - 1985 г.) са настъпили промени в отделни страни от живота - образователна подготовка, работа, семейно положение и други.

Следва се принципът при провеждането на преброяването на населението по местоживеење в проучването да бъдат включени лицата, попаднали в гнездата на извадката, които са преброени като постоянно население на съответното населено място.

Предмет на изследването са промените, проблемите, житейските планове и удовлетвореността на пълнолетните лица по отношение на отделни страни от живота. В проучването са обхванати следните седем основни страни от начина на живот на съвременната личност:

- образователна подготовка;
- работа;
- семейно положение;
- обществено-политическо положение;
- бит и свободно време;
- материално положение;
- здраве и физически сили.

Проучват се промените, които са настъпили през последните десет години (1976 - 1985 г.), проблемите и трудностите понастоящем и житейските планове за следващите десет години (1986 - 1995 година). Изследвана е и общата оценка на личността за живота от 1975 година досега, както и оценката за изпълнението на нейните планове и за намеренията и очакванията за близкото бъдеще.

Интервалите, в които се проучват промените и житейските планове, съвпадат с десетгодишния период между две преброявания на населението и жилищния фонд, с което се създава възможност за обвързване на обективните промени, установени от преброяването, с тяхното субективно възприемане, както и с промените в съзнанието и поведението на личността.

Промените, проблемите, житейските планове и общата оценка за живота са свързани с множество фактори, но безспорно е влиянието на обективните характеристики на личността - местоживеење, пол, възраст, образование, семейно положение, социална група, членство в обществено-политически организации, средно месечен доход и жилищна площ на едно лице, които са включени в изследването. Личностните характеристики на непълнолетните се разглеждат като фактори за изследването и на втория аспект от предмета на проучването - удовлетвореността от всяка от седемте страни от начина на живот. Всяка от седемте страни от начина на живот обаче се разглежда и като фактор относно общата оценка за живота.

Модел и обем на извадката

При определяне на модела на извадката се изхожда от организационната структура на преброяването на населението и жилищния фонд. Най-малката техническа единица на преброяването е жилището с постоянното население, живеещо в него. Това обстоятелство (липса на поименен списък на лицата) прави невъзможно формирането на индивидуална извадка, при която желаната точност се достига с най-малък обем на извадката. Практически не е възможно да се проведе изследването и чрез едностепенна гнездова извадка с гнездо - жилище. Такава извадка би изисквала обучение на изключително голям брой преброители, които да анкетира само по няколко души. Това налага да се оптимизира съотношението между броя на преброителите, участващи в изследването, и броя на лицата, анкетирани от тях.

Всичко това води до решението анкетното изследване да се концентрира в определен брой преброителни участъци, в които да се анкетира всички пълнолетни лица в десет жилища или средно по 22 души.

Моделът на извадката е двустепенен гнездови с гнездо на първа степен - преброителният участък, а на втора степен - жилището. Този модел обаче води до значителни стохастически грешки, с които ще са обременени оценките, получени от тази извадка. Това налага при определяне на обема на извадката да се компенсира влошаващото влияние на вътрешногнездовите корелации (от гнездовия характер на гнездата на първа и от гнездата на втора степен) върху оценките на изследваните показатели. Тъй като не се разполага с точните значения на вътрешногнездовите корелации от двете степени гнезда, се приема, че обем от 22 хиляди души осигурява резултати, които са обременени с достатъчно малки стохастични грешки - оценки от около 0.05 (или пет на сто) да бъдат обременени с максимални относителни грешки не по-големи от 15 на сто.

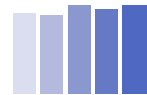
Въз основа на приетия обем на извадката (22 хил. души) и средния брой лица, подлежащи на анкетиране в един преброителен участък (22 души), се определя броят на преброителните участъци (1000), в които се провежда допълнително социологическо изследване „Промени, проблеми и житейски планове на личността“. За преодоляване на необходимостта от пресмятане на вътрешногнездовите корелации, като условие за изчисляване на стохастичните грешки, с които са обременени изследваните показатели, извадката се формира по метода на взаимно проникващите подизвадки чрез пет независими подизвадки. Въз основа на разсейването на значението на показателя между подизвадките се осигурява възможност за изчисляване на максималните стохастични грешки едновременно с електронната обработка на данните, получени от изследването. Извадката е районирана по признаците „окръг“ и „местоживеење“ (град - село).

Изискването в преброителен участък, попаднал в извадката, да се обхванат по еднакъв брой жилища при различен брой жилища в преброителните участъци, налага извадката да се формира, като включването на преброителните участъци се извърши с вероятност, пропорционална на размера на гнездата (брой жилища в преброителните участъци).

Инструментариум и метод на регистрация

Инструментариумът, чрез който се извършва социологическото изследване, включва:

- списъци на преброителните участъци с броя на жилищата (гнезда от втора степен) за всеки окръг поотделно за градските и селските населени места;



- списък ПШ - 10 за жилищата (с имената на главата на домакинството), в които се анкетираат пълнолетните лица в преброителните участъци, попаднали в извадката;
- анкетна карта Л, чрез която се набира необходимата информация;
- инструкция за преброителя (анкетъора);
- инструкция за контролора;
- инструкция за шифриране;
- техническо задание за получаване на таблици чрез електронна разработка.

Карта Л се състои от обща част и два раздела. В общата част са включени идентификационни признаци за преброителните участъци - окръг, населено място, контролен район, и за анкетираното лице - номер на жилището и номер на лицето в преброителния лист.

Първият раздел, който се попълва от анкетираното лице, включва по три въпроса за всяка от изследваните седем страни от начина на живот - съответно за настъпилите промени, за трудностите и проблемите и за плановите. Освен това за същите седем страни е включен по един въпрос - за удовлетвореността на анкетирания от тях. В този раздел са включени и три обобщаващи въпроса - за общата преценка за живота, за изпълнението на плановите и за очакваното изпълнение на плановите в близко бъдеще. В края на раздела са включени и два въпроса от обективните характеристики на личността - членство в общественно-политически организации и средно месечен брутен доход, които не се наблюдават при преброяването на населението.

Вторият раздел (попълва се от преброителя-анкетъор) включва въпроси, които се преписват от преброителния лист, с който е преброено анкетираното лице. Това са въпросите за пол, възраст, семейно положение, образование, социална група, жилищна площ и други.

Изследването е проведено под ръководството на Програмно-целевия колектив по преброяване на населението и жилищния фонд и на окръжните преброителни комисии, респ. окръжните бюра по преброяването. Самото наблюдение е извършено от преброителите, в чиито участъци се провежда допълнително и социологическото изследване.

Анкетирането се извършва чрез самонаблюдение, като преброителят разнася анкетните карти в дома на подлежащите за наблюдение лица.

„Градът и селото - 86“

През 70-те и първата половина на 80-те години в България настъпват значителни качествени изменения във всички сфери и области на обществения живот. Преди всичко това е период на интензивен растеж, на обогатяване и диференциране на обществените структури, на дълбоко преустройство в личностната мотивация, активност и реализация. Улавянето на основните проблеми на функционирането и развитието на обществото, както и анализирането на отделни параметри на състоянието на основните компоненти на обществения живот, налагат необходимостта повторно да се набира информация за цялостната социологическа система на обществото. Възможността тази информация да се сравнява с информацията от предходното изследване на града и селото внася нов елемент в социологическата теория и практика по отношение на съчетаването на синхронна и диахронна информация за състоянието и процесите в обществото като социологическа система. Целта на проведеното изследване включва набирането и анализирането на информация за състоянието, процесите и тенденциите на цялостния обществен организъм в тяхната диалектическа връзка.

Наименованието му има условен характер и преди всичко отразява приемствеността с изследването от 1968 година. По същество „Градът и селото - 86“ е цялостно изследване на характерните черти на българското общество и българския народ, на зрелостта на социалния организъм и тенденциите в развитието му.

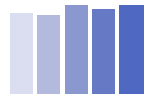
Изследователската му програма се отличава с изключителна всеобхватност. От една страна тя е екстензивно по-богата от тази на изследването от 1968 г., тъй като в изследователското поле сега попадат проблемите на езика, съобщенията, транспорта - тези компоненти на обществото, за които преди не е специално набирана информация. От друга страна тя е интензивно обогатена и доразвита, доколкото дава възможност да се събере по-всеобхватна информация в сравнение с първото за редица компоненти на обществото - физкултура, морал, средства за масова комуникация и други. Картината на социалните процеси се очертава и от функционално-териториалното устройство на страната - тип селище, функционален тип селищна система и т.н., но това е само един от разрезите, в които те се изследват и съпоставят.

Какви нови възможности за анализ се разкриват чрез провеждането на „Градът и селото - 86“?

Можем да ги разгледаме в две главни насоки:

- на първо място се дава простор за периодична повторемост на изследването на обществото в неговата цялост. Оттук произтичат възможностите за анализ на диахронна информация за социалните процеси. Тя съдържа значителен научно-познавателен заряд поради възможностите, които дава за сравнителен анализ на състоянието на обществения организъм и на отделните негови подсистеми и явления в различни периоди. Върху тази основа могат да се правят изводи за тенденциите и темповете на общественото развитие, за скоростта и посоката, за устойчивостта и изменяемостта на социалните процеси и явления. Това е солидна информационна база за разкриване на необходимите, съществени, повтарящите се социологически взаимодействия, за оценка на миналото и настоящето и за прогностично виждане на бъдещето;
- на второ място е новият момент с широки перспективи за обогатяване на теорията и практиката на емпиричното социологическо изследване (ЕСИ) - включването и провеждането на изследването „Градът и селото - 86“ като неразделна част на изследователската програма на преброяването на населението и жилищния фонд в България, проведено през декември 1985 година.

От една страна подобно обвързване на изследователските програми на социологическо и статистическо изследване на обществото като цяло разкрива допълнителни възможности за по-пълно обхващане и анализиране на социалните процеси, тъй като на практика това е експериментиране с възможностите за обвързване на информацията за „акмулирания аспект“ на социологическата структура на обществото с тази за нейния „реален аспект“. Оттук произтичат по-широките параметри за задълбочено изучаване на сложните механизми на функционирането и развитието на обществото чрез по-богатата информация относно всички елементи на социологическата система, съобразно с тяхната тежест и роля в обществения живот. Още повече, че в един период от десет години могат да се забележат промените в състоянието и да се „уловят“ тенденциите на общественото развитие. От друга страна това обвързване внася нови положителни аспекти в самата организация на емпиричните социологически изследвания чрез възможността да се синхронизират включените въпроси, с което едновременно се обогатяват параметрите на социологическото изследване, намалява се броят на



въпросите, разширяват се възможностите за различни комбинации.

Вземайки предвид факта, че в рамките на преброяването през 1985 г. са осъществени и други самостоятелни извадкови изследвания - за репродуктивното поведение, миграцията, трудовите възможности на заетото население, жизнените планове на личността, става ясно, че на практика се пристъпва решително към изграждане на действителна система за социална информация.

В съответствие с целта на изследването за набиране на информация за състоянието и процесите в обществото като социологическа система към 1986 г. и използването на тази информация за сравнителен анализ, се очертават няколко изследователски центъра на изследването „Градът и селото - 86“. Техен обект са основните сфери на обществото като социологическа система - материалното производство, сферата на демографското развитие, духовното производство, социалното управление, сферата на комуникациите. Главните социални процеси и изменения в нашата страна се обхващат в различни разрези: по различни равнища на административно-селищното устройство, по отделни социални общности, по отрасли и области на труда и съответно по трудова заетост по социално-групова принадлежност, по водещи демографски признаци.

За изясняване на тенденциите на общественото развитие, главните задачи на изследването, както и заложените хипотези и очакваните информационни потоци, се концентрират в няколко изследователски полета, чрез които се обхваща степента на развитост, както и изградеността и пропорционалността на основните компоненти на обществения живот.

Изследването „Градът и селото - 86“ има за задача да осъществи приемственост в методиката за набиране на индивидуалната информация. Това означава да се набира информация за акумулирания аспект на социологическата структура на обществото. Доколкото базовата изследователска единица е отделната личност, която е съвкупност от всички обществени отношения и представлява специфична социологическа система, носеща в смет вид цялостната структура на обществото, дотолкова може при прехода от индивидуалната информация за изследваното лице към съвкупната информация, да се твърди, че е събрана информация за цялото общество (при спазване на определени процедури за представителност на извадката спрямо генералната съвкупност).

Генералната съвкупност на социологическото изследване „Градът и селото - 86“ обхваща цялото население на страната над 16 навършени години, като се изключват срочно военнослужещите и лишените от свобода, живеещите в чужбина, както и колективните домакинства (манастири, пансион и прочие). Изследователската единица е отделното лице, а цялата допълнителна информация за заобикалящата го социална (вкл. вещна) среда се набира дотолкова, доколкото тя има някакво отношение към лицето - пряко или косвено.

Прилагането на стохастичния репрезентативен метод позволява изследването да се извърши чрез наблюдаване на представителна извадка от единиците на генералната съвкупност.

С оглед богатата програма на изследването и задачите, поставени пред него, необходимият обем на извадката е определен на около 10 000 лица (изследвани са 10 380 лица). Този обем осигурява постигането на достатъчна научна представителност на резултатите от изследването по отношение на генералната съвкупност.

Като модел за извадката е използван тристепенен гнездови подбор. На първата степен са избрани преброителни участъци, на втора - домакинства, а на трета - лица на възраст над 16 навършени години.

При всяко емпирично изследване, особено при подобно социологическо изследване с многостранна и богата програма, изключително значение има въпросът за осигуряване на научна достоверност

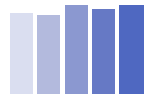
на набраните първични данни. Установяването на обективната истина за всяко изследвано лице е труден и деликатен процес. Той преминава през разработката на теоретичния модел на изследвания социологически обект, през избора на методиката на регистрация, през умелото разработване на инструментариума на изследването, през подбора и обучението на анкетаторите, които ще набират тази информация и така нататък.

И в това отношение е ползван опитът от първото изследване „Градът и селото - 86“. Индивидуалната информация е набрана чрез интервю и косвена анкета успоредно с него. Характерът на методите за набиране на първичната информация определя и източниците ѝ - самите изследвани лица, техният антураж по местоживеење и по месторабота, личните впечатления и наблюдения на прекия изследовател (анкетатор), проучването на документи, свободната беседа с изследваните лица и така нататък.

Индивидуалната информация за всяко изследвано лице е регистрирана от анкетатора в два независими въпросника - „Въпросник за интервю“ и „Въпросник за косвена анкета“.

Залогът за набирането на адекватна информация (която вярно и точно да отразява фактите, интересувачи изследователя) е и добрият инструментариум на изследването. В случая стремежът за постигане на тази цел се изразява в разработването на въпросници, към които са дадени пълни и точни указания за попълването на отделните въпроси. Непосредствено свързани с тях са и подробни инструкции за организацията на изследването, за работата с въпросниците - „Инструкция за организацията на провеждането на ЕССИ“, „Градът и селото - 86“, „Указания за работа с въпросниците“, „Инструкция за кодиране и логически контрол на въпросниците“. Особено силно е подчертана необходимостта от точно спазване на дадените пояснения за попълване на отделните въпроси, критериите за разграничаване на техните отговори, посочените източници за набиране на данни по всеки конкретен въпрос. Със същата цел - да се осигури достоверна информация - е изградена йерархична организационна структура. Централният щаб на изследването, с помощта на окръжните социологически дружества, формира във всеки окръг окръжен щаб на изследването - с ръководител, негов заместник, районни ръководители. Окръжните ръководители и техните заместници участват в специален курс - семинар по въпросите на методологията, методиката и организацията на изследването. Окръжните щабове от своя страна подбират и подготвят непосредствените изследователи (анкетаторите), като провеждат с тях срещи и специални сбирки - инструктажи. Основен принцип при подбора на анкетаторите е да познават поне частично изследваните лица или членове на техните семейства, да бъдат интелигентни, добросъвестни, честни и уважавани граждани. Всички изследователски окръжни екипи са в непосредствена и непрекъсната връзка с Централния щаб на изследването. Участват 1 037 анкетатори, 84 районни ръководители, 28 окръжни ръководители и 28 техни заместници. Приключвайки своята дейност по анкетирането на лицата и по логическия контрол над събраната информация по определени в инструкцията правила, всеки изследовател попълва „Методически въпросник“, в който отразява своето виждане за проведеното изследване, положителните страни, срещнатите трудности, отправя своите препоръки към авторския му колектив.

С особено важно значение за правилното протичане на изследването е изградената система за контрол на работата на всички равнища. Всеки окръг е неколkokратно посетен от членове на Централния щаб на изследването, включително и за участие в окръжните инструктажи. Своевременно са проведени консултации по възникнали проблемни ситуации. За осигуряване на пълна и логически непротиворечива първична информация всеки въпросник преминава през неколkokратен контрол - от страна на районния ръководител, от окръжния ръководител или неговия заместник (които кодират въпросниците), от



специално създадено звено за „суперконтрол“ в Централния щаб на изследването в гр. София.

Проектът за обработка на информацията като преход от индивидуалната към съвкупната информация се изразява главно в изчисляването на процентни съотношения в табличен вид, както за целия обем на извадката, така и за формирани конкретни подмасиви от нея - на живеещите в градовете, в селата, в отделните типове селища - в столицата, в окръжните градове, в останалите градове, в селата - центрове на общини, в останалите села, на живеещите в различните функционални типове на селищните системи - нулев, първи, втори и т.н., на заетите в общественото производство и на незаетите; на заетите, пенсионерите, учащите се, незаетите и така нататък. Към всяка групировъчна таблица се изчисляват различни коефициенти на взаимовръзка и средни величини. Особено внимание е отделено на възможностите за съпоставимост на резултатите от проведените изследвания на града и селото през 1968 и 1986 година. За да се постигнат целите и да се изпълнят задачите на изследването, както и да се проверят включените в програмата му хипотези, е необходимо да се събере стабилна статична и динамична информация. Това се постига чрез използване на основните въпроси от предишните изследвания и чрез включване на редица отделни индивидуални въпроси, даващи възможност да се улови динамиката на развитието.

Изследователският подход налага формирането на нови въпроси (съставни индикатори) чрез използването на различни процедури за групиране на първичната индивидуална информация. Замяната на няколко отделни емпирични индикатора е реализирана върху тяхната основа - съставен индикатор - логически конструкт, който позволява на изследователите да осъществят пълноценен анализ, разглеждайки изучаваното свойство на обекта на социологическото изследване в неговата цялост - например трудовата активност на личността, социалната активност на личността, образователно-квалификационното ѝ равнище, степен на активност при изразяване на духовни ценности, степента на изграденост на материалнобитовите условия на живот и така нататък.

В други случаи емпиричните резултати за отделни признаци, разглеждани сами за себе си, не носят необходимата за социолога информация (например наличието поотделно на радио, грамофон, телевизор, библиотека и пр. в домакинството) или са трудни за интерпретация (например когато полученото разпределение е твърде неравномерно или някои интересувачи социолога характеристики са представени с много малки относителни дялове). С обобщаването на няколко такива признака (близки по съдържание и характер) в съставен индикатор, в значителна степен се подобрява възможността за включването им в социологическия анализ.

Какви основни изводи могат да се направят за ролята и значението на ЕСИ „Градът и селото - 86“?

Проведеното изследване не е регулярно, не е едно от стотиците, които се провеждат ежегодно у нас. По обема на генералната съвкупност (цялото население на страната над 16-годишна възраст) то може да бъде съпоставено единствено с изследването на религиозността (1962 г.) и с „Градът и селото - 68“. Всички други големи изследвания имат ограничения на някои параметри - отнасят се или само за заетото население, или за семействата в ограничен възрастов диапазон, или са тематично ограничени по отделни проблеми. Това налага да се направят анализи от съдържателна, методологична и методическа гледна точка на изследването „Градът и селото - 86“, тъй като изводите от него могат да имат твърде широк периметър на валидност.

Изследването от 1986 г. е своеобразна равностметка на изминатия в това отношение път. То проправя път към нов опит в социологическата работа - повтарянето на базисни изследвания. Заедно с това то е не просто „ехо“ на проведеното през 1968 г. изследване, а отражение и динамично претворяване

на натрупания опит, на самото развитие на социологическата наука и на практическите умения на социолозите. То е фокус на приемствеността и развитието. „Градът и селото - 86“ проверява как уменията и опитът на социолозите се проявяват в „работна обстановка“ и доколко формулираните класификационни критерии действат.

Емпиричното представяне на всеки отделен въпрос във въпросника се нуждае от здрава теоретична „подплата“ - например въпросите за сферата на заетост на лицето и за специфичността на занятието, съотнесени с компонентите на обществения живот, изискват задълбочено структуриране на компонентите и свързването им с конкретните дейности и съответстващите им институции. Нужна е стройна теоретична постановка, за да се формулират въпросите (и особено възможните отговори) за социално груповата принадлежност на лицето. Емпиричното представяне на някои възлови социологически категории (социална, трудова и естетическа активност, социална мобилност между поколенията и т.н.) изисква теоретично определяне, съответно операционализиране за свеждане на емпирично равнище, с цел включване на необходим и достатъчен брой емпирични индикатори, чрез които по-късно да бъде извървян обратният път за изграждане на обобщения индикатор, чрез който да се отрази съответната социологическа категория.

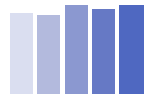
„Градът и селото - 86“ е практически резултат преди всичко от теоретичното израстване на социологията. Ползвайки базата на изследването от 1968 г., това през 1986 г. отразява както развитието, така и практиката на реализацията и организацията на социологическите изследвания. Приемствеността и различието цялостно се „оглеждат“ в инструментариума.

В организационно отношение „Градът и селото - 86“ строго спазва теоретичните изисквания в тази насока, пречупвайки ги през спецификата на едно голямо представително изследване, което е приемник на друго, вече проведено. Самата организация на изследването на всички етапи изисква специално описание, тъй като съдържа много интересни моменти, които биха могли да бъдат образци.

Естествено в най-голяма степен в изследването трябва да се търси съдържателната гледна точка. Неговата главна цел от една страна е да установи чрез средствата на национално представително изследване състоянието на българското общество като социологическа система в средата на 80-те години, като разкрие необходимите, съществени, повтарящите се взаимодействия и въз основа на тях да направи опит за прогнозиране на развитието, а от друга е практико-приложната цел - да се разработят за органите на социалното управление съответни предложения и решения, които могат да се решат:

- чрез правилно подбиране на въпросите, включени в изследването, така, че те точно, вярно и адекватно да отразят функционирането на обществото чрез всички негови компоненти и техни подразделения;
- чрез добре организирана на всички етапи на изследването и точно разработена методика да се събере достоверна и надеждна информация;
- чрез сериозно обмислена програма за обработка набраната индивидуална информация да се превърне в обозрим за изследователя масив на обобщена информация;
- чрез правилен анализ на включените въпроси да се разкрият собствените евристични възможности на набраната информация за отразяване на социалните процеси и в динамичен аспект - например социална и образователна мобилност, социална еднородност между поколенията (респ. разнородност), миграционно поведение и така нататък.

Богатството на съдържанието на набраната информация е практически неизчерпаемо поради огромния брой комбинации, които могат да се получат при съответните връзки на над 300 базови,



60 групирани, 40 съкратени и 220 нови (по съставни индикатори) въпроси. Тези съдържателни възможности неколkokратно се увеличават, като се вземат предвид базовите изследвания на религиозността (1962 г.) и „Градът и селото - 86“, както и резултатите от преброяването през 1985 г. и свързаните с него представителни изследвания. Изключително важно е тази информация да се използва пълно и интензивно. Фактически набраната информация е една отворена система.

Названието „Градът и селото“ действително до голяма степен е условно. То отразява приемствеността с изследването от 1968 г., което от своя страна с това име е обединило двата основни етапа на подготовката му, замислено е като изследване на българското село, а след това при известно обогатяване на програмата е включен и градът като подсистема на обществото. По този начин научно-практически се доказва теоретичната постановка, че градът и селото са основни подсистеми на обществото, изоморфни на обществото като цяло, в голяма степен изоморфни и помежду си, поради способността си да се самовъзпроизвеждат. Всички основни масиви от набраната информация са разпределени по град и село и показват, че всички компоненти имат своите проекции и в града, и в селото. В това отношение „Градът и селото - 86“ не само спазва приемствеността, но и прави крачка напред с допълнителни разработки, даващи и разпределенията по тип селище (София, окръжен град, друг град, село - център на община, село - кметство, село - без статут на кметство). Може да се каже, че проведеното изследване съдържа в себе си две самостоятелни изследвания - на българския град и на българското село, подрабтиращи и сравнимостта между тях.

Избраната информация и получените вече разработки показват, че данните от изследването дават действителна възможност да бъдат направени такива социологически анализи, чрез които да се изпълнят основните задачи на изследването:

- да се изяснят основните тенденции на развитие в нашето общество през 80-те години на XX в. чрез използване на синхронна и диахронна информация, като се разкрие степента на развитост, изграденост и пропорционалност на основните компоненти на обществения живот;
- да се формулират съответни мнения и препоръки към органите за социално управление относно най-съществените фактори на социалното развитие чрез тяхното диференциране и очертаване на основните тенденции в перспективното им развитие;
- да се решат собствено научни проблеми свързани както с теорията на социологията като наука, така и с теорията и практиката на ЕСИ.

Провеждането на социологическото изследване „Градът и селото - 86“ има и положително влияние върху самата социологическа общност - за нейното обединяване, за актуализиране на връзките между нейните членове, за изграждане и на неформални отношения. „Градът и селото - 86“ от своя страна поставя като бъдеща задача да се продължи работата по изграждането на единна система за социологическа информация, за традиционност при повторемостта на изследванията, за възможностите за сравнение и така нататък.

Здравно състояние на населението

Състоянието на семейното и общественото здраве, здравните потребности на населението и степента на тяхното задоволяване са въпроси от изключително значение за социалната политика в страната. Изследването им, съвместно с проучването на професионалния състав и квалификация на населението, неговото репродуктивно, миграционно и трудово поведение, както и със социологическото изследване на града и селото, създава реална възможност за получаване на по-

цялостна качествена характеристика на основната производителна сила - населението.

Здравното състояние на населението се характеризира най-добре чрез показателите на заболяемостта, инвалидността, физическото и нервно-психическото развитие. То се определя от условията на живот и по-конкретно от трудовите и културно-битовите условия. В значителна степен оказват влияние също така и принципите на здравеопазването, методите и формите на здравното обслужване. Здравната култура и здравното поведение са съществен фактор, който влияе върху семейното и общественото здраве.

От казаното по-горе следва, че здравето на населението трябва да се изследва комплексно и във връзка с условията и начина на живот. Възможност за осъществяване на това изискване дава свързването на здравното проучване с преброяването на населението, по време на което пълно и точно могат да се разкрият условията на живот.

Цел и задачи на изследването

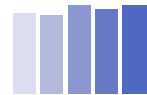
Основната цел на изследването е да се получи цялостна оценка на общественото здраве, чрез точно определяне и задълбочено изследване на неговите компоненти: индивидуално здраве; трудова годност; здравна култура; здравно поведение; здравни потребности на населението.

За изпълнение на поставената цел е необходимо да се изследват задълбочено:

1. Заболеваемостта на населението.
2. Травматизмът - общ, трудов, битов, транспортен, спортен.
3. Вродените заболявания и физически недъзи.
4. Инвалидността и трудоустрояването.
5. Медицинските и социалните потребности на населението и степента на тяхното удовлетворяване.
6. Здравното поведение и здравословният начин на живот.
7. Коефициент на детност, плодовитостта, стерилитетът, детската смъртност, абортите, мъртворождаемостта и контрацепцията (противозачатъчните средства).
8. Здравната култура на населението.
9. Физическото развитие и физическата дееспособност.
10. Нервно-психическото развитие.
11. Трудовата годност и работоспособността на населението.
12. Влиянието на жилищнобитовите и професионално-трудовите условия на живот върху здравето на лицата и тяхната трудова годност.

На тази основа е възможно да се извърши:

1. Комплексна оценка на здравето, трудоспособността и работоспособността на населението.
2. Комплексна оценка на здравните потребности на населението в национален мащаб и планирането им на национално равнище.
3. Комплексна оценка на здравето, трудоспособността, работоспособността, здравните нужди и здравното поведение на населението в регионален аспект - по икономикогеографски райони (области) и проблеми на конкретната социална и здравна политика на управлението, планирането, икономиката и организацията на здравеопазването и на другите сфери на



обществения живот.

Метод и единици на наблюдение

Изследването на здравното състояние на населението се организира и провежда съвместно с Научния институт по социална медицина при Медицинска академия. Наблюдението се осъществява с три анкетни карти по предварително съставени списъци на жилищата и всички лица, живеещи в тях.

Обект на наблюдение са:

1. домакинството;
2. семейството;
3. отделното лице - дете или възрастен.

Изследването е репрезентативно за цялата страна.

Една от задачите на проучването на здравето на населението е събирането на най-точна информация за трудовата годност на икономически активните лица, която е един от компонентите на трудовите възможности, наред със способността и готовността на хората за извършване на обществено значима трудова дейност. Поради тази причина за най-целесъобразно е прието извадката на изследването да се формира като подизвадка на проучването за трудовите възможности и реализация на лицата в трудоспособна възраст. Здравно проучване се провежда в 1/3 от жилищата, в които по време на преброяването е попълвана карта Т. От преброителните участъци, попаднали в извадката, са наблюдавани по 15 жилища с всички живеещи в тях лица. На всяко лице, подлежащо на изследване, са направени прегледи от участъков терапевт и девет лекари специалисти, а за икономически активните лица се сформират трудово-експертни лекарски комисии (ТЕЛК) за определяне на трудовата годност и работоспособността. Или за всяко гнездо (преброителен участък) по време на изследването са ангажирани 15 души анкетъорски апарат, състоящ се от: медицински сестри, участъкови терапевти и педиатри, лекари специалисти и членове на ТЕЛК. Общо за страната техният брой възлиза на 5 250 души медицински персонал.

Обемът на извадката е определен на около 23 хиляди лица. От тях по време на изследването по различни причини отпадат над 1 200 души. В масива за обработка влизат общо 21 792 индивидуални карти, от които 17 073 на лица на възраст над 15 г. и 4 719 на деца от 0 до 14 години.

Признаци

Първичната информация е събрана с помощта на инструментариум, състоящ се от индивидуална карта на лицето, карта за дете до навършена 14-годишна възраст, карта на домакинството и инструкция за попълване на картите и тяхното шифриране.

Индивидуалната карта на лицето се състои от четири раздела:

- Първият раздел включва интервю на медицинската сестра в дома на лицето, по въпроси, свързани със социално-демографските и икономическите характеристики на лицата, условията за труд, заболяемостта, травматизма, здравната просвета, свободното време, наличието на някои вредни навици (употреба на алкохол и тютюнопушене). Тук е изразено мнението на лицето за неговото здраве и е характеризирен неговият начин на живот.
- Във втория раздел се записват данни от здравната документация на лицето, преди всичко от личната амбулаторна карта. В него се отбелязват броят и времетраенето на регистрираните заболявания, временната нетрудоспособност, хоспитализацията и диспансеризацията на лицето. Всички тези данни се отнасят само за боледуванията през 1985 година. В края на

този раздел медицинската сестра записва и своята преценка за здравната култура на лицето.

- Третият раздел се попълва от осем до десет лекари специалисти след обстоен преглед на лицата. Заключителната оценка за общото здравно състояние на лицето, начина му на живот и мероприятията, необходими за подобряване на неговото здраве, се определят от лекарска комисия в състав: завеждащ терапевтично отделение, участъков лекар-терапевт и завеждащ СЛПЗ (селски лечебно-профилактични заведения).
- Четвъртият раздел се попълва само за икономически активните лица. Чрез провеждането на определени измервания и тестове в него се дава агрегирана оценка на физическото развитие, физическата дееспособност и нервно-психическото развитие на трудещите се. Експертизите на трудовата годност и работоспособността на тези лица се извършват от ТЕЛК.

Информацията, събрана в картата на лицето, дава възможност да се направи по-цялостна характеристика за здравето на населението на 15 и повече години, да се съпоставят мненията на отделните лица и здравните работници по въпросите, свързани със здравната култура, здравната просвета и здравословния начин на живот на населението. Тук се разкриват най-точно здравните потребности на индивидите и степента на тяхното задоволяване.

Картата на децата до навършена 14-годишна възраст съдържа три раздела.

- В първия се записват по документи данни за раждането и общото здравно състояние на детето.
- Вторият раздел включва въпроси, свързани с физическото развитие на детето и естествените промени при растежа му, а също така и срещането на някои вродени и придобити аномалии.
- Третият раздел се попълва след комплексен преглед от лекари специалисти и дава най-вярна представа за действителното здравно състояние на подрастващите, за техните здравни потребности и проблеми във връзка с условията на живот.

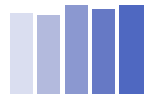
Третата карта осъществява връзката между индивидуалните карти на всички членове от домакинството. Тя се състои от два раздела. Първият разглежда въпросите, свързани с жилищнобитовите условия, при които живее домакинството и съставлящите го семейства.

Вторият раздел обобщава данните от индивидуалните карти на всички членове на домакинството. В него се събира информация за количествения състав на домакинството - тип на семействата в него, брой на членовете му и брой на децата. Тук се прави социална и образователна характеристика на домакинството като цяло. Разглеждат се някои въпроси, свързани с доходите и разходите за храна на членовете му, жилищната площ и жилищните условия. Подробно се проучват възможностите за наличие на рискови семейства. Определят се медицинските и социалните мероприятия, психоклиматът и здравното състояние на домакинството. Събраната информация създава възможност за първи път в нашата страна да се направи цялостна характеристика на семейното здраве, да се разкрият здравните потребности, психоклиматът и съществуващите рискови фактори в българското семейство.

Резултати от изследването

Резултатите от изследването на здравното състояние на населението следва да бъдат обобщени в пет основни публикации, които обхващат до голяма степен проблемите, потребностите и състоянието на семейното и общественото здраве на населението в България.

Събраният фактологически материал дава възможност да се установи какво влияние оказват върху здравното състояние жилищнобитовите и материалните условия, социално-демографските



фактори и здравното поведение на населението. Прави се съпоставка между субективното и обективното мнение на здравното състояние на лицата, тяхната здравна култура и здравните им проблеми. На базата на тези резултати може да се определи действителният интерес на индивидите към собственото им здраве и да се направят някои конкретни изводи за годността на българското население да води здравословен начин на живот.

От много важно обществено значение са резултатите, свързани с трудоспособността и работоспособността на икономически активните лица. Комплексното проучване на признаците, характеризиращи физическото развитие, дееспособността и нервно-психическото развитие, а също така и единният подход при тяхното оценяване (използва се методиката, възприета от Международната организация на труда при ООН), допринасят за възможно най-точното определяне на годността на лицата за полагане на определен вид трудова дейност. На преценка подлежи също така готовността и възможността на хората да извършват определена професионална дейност.

Разкриването на семейното и общественото здраве, както и степента на задоволяване на здравните потребности и тяхното равнище в дадения момент позволяват по-целенасочено да се планират редица мероприятия, да се определят по-точно водещите насоки на здравната и социалната политика на страната.

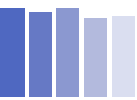
Изводи

Проведеното изследване на здравното състояние на населението доказва че е възможно при създаване на единна методология и еднакъв подход да се получи точна представа за здравето на населението в цялата страна. Това става възможно благодарение на предварителното уеднаквяване на критериите за работа на създадените специално за проучването лекарски екипи, лекарски консултативни комисии (ЛКК) и ТЕЛК. Така до голяма степен се премахва субективизма при оценяване на семейното и общественото здраве и трудовата годност. Съвестното отношение на медицинския персонал и населението към изследването спомагат за качествено му провеждане. Получените резултати създават възможност за постигане на основната цел и задачите поставени пред проучването.

Заклучение

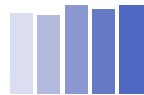
В заключение може да се каже, че методологията и програмата на преброяването на жилищния фонд, проведено в края на 1985 година, се основават на значителния опит и постиженията на преброяванията на жилищния фонд в миналото, на изискванията на съвременната статистическа и демографска наука и практика и на изискванията на обществената практика в областта на социалната политика.

Изчерпателното представяне на репрезентативните наблюдения, проведени по време на преброяването и непосредствено след него, както и историческият аспект, чрез който е представена информацията в статията, ще обогати съдържанието на списание „Статистика“ и ще запознае читателската аудитория с теорията и практиката на едно нестандартно по своята същност преброяване на населението и жилищния фонд проведено в края на 1985 година.



ЦИТИРАНА ЛИТЕРАТУРА:

1. **Балевски, Д. (1989). Актуални проблеми на преброяванията на населението в НР България за периода 1900 - 1985 година. София, ЦСУ, 136 с.**
2. **Аналитични доклади. Програмно целеви колектив по изграждане на единна система за социално-демографска статистика. (1988), София, ЦСУ, ДП „Георги Кирков“ - Бургас, 255 с.**
3. **Материали за извършената работа по организирането и подготовката на преброяване на населението и жилищния фонд в края на 1985 г. през II полугодие на 1984 година. (1985). КЕССИ при МС, печатна база при КЕССИ при МС, 109 с.**
4. **Програма на преброяването на населението и на жилищния фонд на 4 декември 1985 г. (1984). София, КЕССИ при МС, Издателство и печатница при Комитета по ЕССИ , 304 с.**



МЕТОДИКА ЗА УПРАВЛЕНИЕ НА СТАТИСТИЧЕСКИЯ ПРОЦЕС В ОБЛАСТТА НА ОКОЛНАТА СРЕДА ПОДХОД НА НСИ ЗА ИНТЕГРИРАНА МЕТОДОЛОГИЧНА РАМКА

Гергана Чешмеджиева*

РЕЗЮМЕ За ефективното прилагане и мониторинг на политиките в областта на околната среда е необходима достоверна, надеждна и навременна статистическа информация. С нарастващото търсене на качествени статистически данни, необходимостта от ефективно управление на статистическия процес става особено значима.

В отговор на нарастващите предизвикателства и установените след задълбочен анализ нужди, касаещи производството на статистика на околната среда, НСИ разработва и внедрява „Методика за статистиката и сметките за околната среда“. Чрез тази сериозна крачка напред за първи път всички процеси, свързани с планиране, събиране, валидиране, анализ и докладване на данни за околната среда, са обединени в един последователен и логически структуриран документ, съдържащ както общите принципи на статистическото производство, така и спецификата на всяко отделно изследване за околната среда, провеждано от НСИ.

Практическото прилагане на Методиката води до подобряване на координацията, намаляване на грешките при събирането и обработката на данни, по-бързо обучение на нови служители и по-добра проследимост на извършените дейности. Методиката изпълнява и стратегическа функция, служейки като инструмент за обучение и за поддържане на институционалната памет. Този иновативен подход в областта на статистиката на околната среда е от ключово значение за гарантирането на устойчивост и последователност в работата и непрекъснато повишаване на качеството на статистическите данни.

JEL класификация

Q56 ‘Environment and Development; Environment and Trade; Sustainability; **Environmental Accounts and Accounting**; Environmental Equity; Population Growth’

(Източник: JEL Classification System, <https://www.aeaweb.org/econlit/jelCodes.php?view=jel#Q>)

* д-р инж., началник на отдел „Сметки в околната среда и енергетиката“, НСИ; e-mail: GCheshmedjieva@NSI.bg

МЕТОДИКА УПРАВЛЕНИЯ СТАТИСТИЧЕСКИМ ПРОЦЕССОМ В ОБЛАСТИ ОКРУЖАЮЩЕЙ СРЕДЫ ПОДХОД НСИ К ИНТЕГРИРОВАННОЙ МЕТОДОЛОГИЧЕСКОЙ РАМКЕ

Гергана Чешмеджиева*

РЕЗЮМЕ Для эффективного применения мониторинга политиках в области охраны окружающей среды необходима достоверная, надежная своевременная статистическая информация. В условиях возрастающего спроса на качественные статистические данные, необходимость эффективного управления статистического процесса становится особенно значимой.

В ответ на растущие вызовы и выявленные после тщательного анализа потребности, касающиеся производства статистики по окружающей среде, НСИ разрабатывает и внедряет «Методику по статистике и счетам окружающей среды». Этот важный шаг вперед впервые объединяет все процессы, связанные с планированием, сбором, проверкой, анализом и отчетностью данных по окружающей среде, в один последовательный и логически структурированный документ, содержащий как общие принципы статистического производства, так и специфику каждого отдельного исследования окружающей среды, проводимого НСИ.

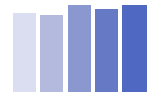
Практическое применение Методики приводит к улучшению координации, снижению ошибок при сборе и обработке данных, ускоренному обучению новых сотрудников и лучшей прослеживаемости выполненных действий. Методика также выполняет стратегическую функцию, служа инструментом для обучения и поддержания институциональной памяти. Этот инновационный подход в области статистики окружающей среды имеет ключевое значение для обеспечения устойчивости и последовательности работы, а также для непрерывного повышения качества статистических данных.

JEL классификация

Q56 'Environment and Development; Environment and Trade; Sustainability; **Environmental Accounts and Accounting**; Environmental Equity; Population Growth'

(Источник: JEL Classification System, <https://www.aeaweb.org/econlit/jelCodes.php?view=jel#Q>)

* Доктор технических наук, заведующий кафедрой «Учет в сфере охраны окружающей среды и энергетики», НСИ,
e-mail: Cheshmedjieva@NSI.bg



GUIDELINES FOR MANAGEMENT OF THE STATISTICAL PROCESS IN THE FIELD OF ENVIRONMENTAL STATISTICS APPROACH OF NSI FOR INTEGRATED METHODOLOGICAL FRAMEWORK

*Gergana Cheshmedjieva**

SUMMARY The effective implementation and monitoring of environmental policies requires reliable, trustworthy and timely statistical information. With the growing demand for quality statistical data, the need for effective management of the statistical process becomes particularly significant.

In response to the growing challenges and the needs identified after in-depth analysis regarding the production of environmental statistics, NSI has developed and implemented the 'Guidelines for Environmental Statistics and Accounts'. Through this major step forward, for the first time, all processes related to planning, collection, validation, analysis and reporting of environmental data are united in one consistent and logically structured document, containing both the general principles of statistical production and the specifics of each individual environmental survey conducted by NSI.

The practical application of the Guidelines leads to improved coordination, reduced errors in data collection and processing, faster training of new employees and better traceability of the carried out activities. The methodology also performs a strategic function, serving as a tool for training and for maintaining institutional memory. This innovative approach in the field of environmental statistics is of key importance for ensuring sustainability and consistency in the work process and continuous improvement of the quality of statistical data.

JEL classification

Q56 'Environment and Development; Environment and Trade; Sustainability; **Environmental Accounts and Accounting**; Environmental Equity; Population Growth'

(Source: JEL Classification System, <https://www.aeaweb.org/econlit/jelCodes.php?view=jel#Q>)

* Dr. Eng., Head of 'Environmental and Energy accounts' Department, NSI; e-mail: GCheshmedjieva@NSI.bg

МЕТОДИКА ЗА УПРАВЛЕНИЕ НА СТАТИСТИЧЕСКИЯ ПРОЦЕС В ОБЛАСТТА НА ОКОЛНАТА СРЕДА ПОДХОД НА НСИ ЗА ИНТЕГРИРАНА МЕТОДОЛОГИЧНА РАМКА

Гергана Чешмеджиева*

Въведение

Опазването на околната среда е комплексна и интердисциплинарна тема, залегнала в редица международни и европейски документи, включително в Европейския зелен пакт, модела за кръгова икономика и Целите за устойчиво развитие на Организацията на обединените нации (ООН). За ефективното прилагане и мониторинг на свързаните политики е необходима достоверна, надеждна и навременна статистическа информация.

Сметките за околната среда, изградени върху принципите на националните сметки, предоставят интегрирана информационна рамка за връзката между околната среда и икономиката. С нарастващото търсене на качествени екологични данни, необходимостта от ефективно управление на статистическия процес става особено значима.

Настоящият материал представя Методиката за статистиката и сметките за околната среда - вътрешен инструмент на Националния статистически институт (НСИ), разработен по проект, финансиран от Европейската комисия (Споразумение за субсидия № 101112756, РД 10-6/18.04.2023 година). Методиката цели стандартизиране и подобряване на цялостния процес по събиране, валидиране, анализ и разпространение на данни за околната среда, включително осигуряване на методологична приемственост и устойчивост в условията на кадрови промени.

Основно изложение

1. Анализ на потребностите

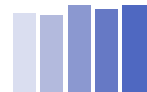
Разработването на такъв всеобхватен и същевременно конкретен документ, какъвто е Методиката за статистиката и сметките за околната среда, изисква предварително задълбочено проучване и анализ на наличните документи и прилаганата практика на национално, европейско и международно ниво. Подобен анализ позволява изграждането на общо разбиране за цялостната методологична и практическа картина на производството на статистически данни за околната среда и формулирането на изводи в аспекта на целевата методика.

1.1. Анализ на статистическата дейност за околната среда

1.1.1. Обща статистическа методология

Проучването относно общата статистическа методология, включително за статистическия процес, е проведено въз основа на ключови документи, сред които Регламент (ЕО) № 223/2009 относно европейската статистика, Закон за статистиката, Регламент (ЕС) № 691/2011 относно европейските икономически сметки за околната среда и неговите изменения, Системата за екологично-икономическо счетоводство на ООН, Кодекса на европейската статистическа практика, вътрешните правила и инструкции на НСИ по управление на статистическия процес и контрол на качеството и други. Извършен е и преглед на специализираната учебна литература в областта на статистиката.

* д-р инж., началник на отдел „Сметки в околната среда и енергетиката“, НСИ; e-mail: GCheshmedjieva@NSI.bg



Европейските и националните методологични документи осигуряват концептуалната рамка за прилагане на хармонизиран подход в статистическата дейност. Те засягат най-общо процесите по разработването на методология и планирането на статистическите изследвания, получаването, събирането, обработката, съхраняването на индивидуални и статистически данни, анализа, предоставянето и разпространението на статистическа информация. Нормативните документи съдържат общи и принципни изисквания за етапите на статистическия процес и неговия административен структурен аспект, включително отношенията между различните участници в процеса. От друга страна учебните материали предоставят теоретична информация за статистическото изследване, когнитивната същност и характеристики на статистическия метод, статистическите понятия, грешките в статистическото изследване, статистическия анализ и други. Литературата разглежда подробно методите, прилагани на всеки етап от статистическата дейност, включително математически изчисления, примери и други.

В резултат от проведения анализ е установена необходимост от обобщаване, систематизиране и адаптиране на действащата обща методологична база за конкретните нужди на статистиката на околната среда.

1.1.2. Специализирана методология

Проучването по отношение на методологията на всяко от статистическите изследвания за околната среда е проведено въз основа на специализираните методологични документи на Евростат, Организацията за икономическо сътрудничество и развитие и др., например методологичните ръководства на Евростат за сметките за емисиите във въздуха, разходите за опазване на околната среда, екологичните стоки и услуги и други. С цел осигуряване на последователност на методологичната работа са детайлно прегледани и резултатите от предходни проекти на НСИ, свързани със статистиката и сметките за околната среда.

Анализът на специализираната методологична обезпеченост показва, че прилаганите към момента писмени методологии за всяко статистическо изследване на околната среда, провеждано от НСИ, са твърде общи и липсва информация, отразяваща най-актуалната прилагана практика. Необходимо е те да бъдат актуализирани с цел включването на всички практически прилагани елементи, включително методологичните подобрения от предишни проекти.

1.2. Анализ на капацитета

Качественото и устойчиво производство на статистическа информация за околната среда изисква не само наличие на ясни методологични рамки и нормативна база, но и адекватен човешки ресурс по отношение на брой, квалификация и институционална подкрепа. В този контекст е извършен задълбочен анализ на организационния и кадровия капацитет на НСИ по отношение на дейностите, свързани със статистиката за околната среда.

1.2.1. Организация

Организацията на статистическата дейност по околна среда е изградена на две основни нива - централно (централно управление) и териториално (териториални статистически бюра). В централното управление се осъществява методологичното ръководство, координацията с Евростат и междуведомствените партньорства, докато териториалните структури отговарят за събирането, предварителната проверка и предаването на първични данни.

Анализът показва, че координацията между тези две нива е от решаващо значение за гарантиране на ефективното и навременно изпълнение на дейностите. В условията на повишени изисквания за

качество, разширяване обхвата на данните и съкращаване на срокове за докладване, всяко забавяне или неразбиране между участниците в процеса води до риск от отклонения и загуба на качество.

1.2.2. Текучество и институционална памет

През последните години се наблюдава значително обновяване на експертния състав както на централно, така и на регионално ниво. Макар това да има положителен ефект за въвеждане на нови перспективи и подходи, то води и до сериозно предизвикателство в осигуряването на приемственост. В условията на ограничена документация и преобладаваща устна комуникация, новите служители се сблъскват със затруднения при възпроизвеждане на процедурите и стандартите.

Пандемията от COVID-19 рязко изостри тези проблеми, като доведе до временна реорганизация на екипите, дистанционна работа и натрупване на напрежение в определени функционални зони. Това подчерта още веднъж уязвимостта на процеса в условия на криза и необходимостта от наличие на централизирана и достъпна методическа рамка.

1.2.3. Обучение и развитие

Понастоящем НСИ провежда годишни вътрешни обучения за статистиката на околната среда, както и профилирани/тематични такива при установена необходимост за това. Възникващите в хода на ежедневната дейност практически въпроси се изясняват чрез оперативна комуникация в хода на работата.

Анализът показва необходимостта от по-обширна, задълбочена и структурирана писмена база, както за регулярното изпълнение на задълженията на експертите по околна среда, така и за бъдещо обучение и въвеждане в работата. Методиката, като писмен документ, предоставя такава възможност и допринася за намаляване на зависимостта от индивидуални знания и опит.

1.2.4. Натоварване и производителност

Нарастващият обхват от статистически задължения, включително новите модули и изисквания по регламентите на Европейския съюз (ЕС), води до значително увеличаване на натоварването на експертите, без това да е задължително съпътствано от увеличаване на ресурсите. Чрез въвеждането на стандартизирана методика се цели оптимизация на работните процеси и улесняване на задачите чрез ясно разписани стъпки, графици и критерии за качество.

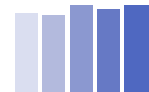
1.3. Национални и международни практики

В рамките на разработването на Методиката е извършен систематичен преглед на добри практики и модели на организация в други национални статистически институти от ЕС. С тази цел е съставен и разпространен структуриран въпросник до 26-те национални статистически служби на държавите - членки на ЕС, с цел идентифициране на практики по прилагането на писмени методологии и ръководства, както и структурата на техните статистически процеси в областта на околната среда.

Над 50% от участниците предоставят отговори, които са използвани за съставяне на сравнителен анализ. Обобщеното наблюдение от това изследване е, че държавите членки (аналогично на България) изготвят статистика и сметки за околната среда, съгласно действащите нормативни документи, въз основа на общи стандартни методи и подходи и утвърдени европейски и международни указания. Наличието на единен писмен методически документ, специално разработен за нуждите на статистиката за околната среда, не е упоменато в никой от получените отговори.

1.4. Формулирани потребности

Въз основа на извършените анализи - на нормативната рамка, на международните практики и на



наличния капацитет - са формулирани конкретни потребности, които обосновават необходимостта от създаването на Методиката за статистиката и сметките за околната среда. Тези потребности обхващат както методологични, така и организационни аспекти на статистическия процес и се групират по следния начин:

1.4.1. Систематизация и централизация на знанията

В хода на работа е установено, че съществуващите документи и указания, използвани от експертите, са разпокъсани, с различна степен на актуалност и обхват. Част от тях се базират на устно предадени практики или локални инициативи. Така се създават предпоставки за нееднакво прилагане на процедурите, което води до риск от несъответствия и разлики в качеството на данните. Това обуславя необходимостта от разработване на централен, обединяващ документ, съдържащ всички релевантни методологични и процедурни насоки, с осигурена периодична актуализация.

1.4.2. Уеднаквяване на процесите и актуализиране на методологията

В последните години НСИ въвежда подобрения в статистическия процес за околната среда от една страна чрез подобряване на организацията на процедурата и координацията между структурните звена, а от друга - чрез активно обучение на експертния персонал и въвеждане на критерии, позволяващи съвместно валидиране на данни, съобразно аналитичните връзки между различните статистически области. Съществуващите писмени методологии по отделните изследвания за околната среда обаче не отразяват тези най-актуални практики в производството на статистическите данни. Поради това е налице необходимост от дефиниране на обща рамка на статистическия процес, валидна за всички изследвания, с включени фази, отговорности и шаблони за работа, както и от актуализиране на методологиите на изследванията за околната среда.

1.4.3. Документиране на добрите практики

През последните години НСИ внедрява и редица полезни практики в работата по статистиката за околната среда. Така например е създадена и успешно функционира информационна система „Статистика на околната среда“, осигуряваща възможност за въвеждане, обработване, логическо обвързване, проверка и съхраняване на данните от изследванията. В резултат са оптимизирани информационните потоци, съкратени са сроковете за събиране на данни, подобрени са процедурите за проверка и контрол на качеството на информацията и е създадена единна база данни. Също така са подобрени процедурите по организация на статистическия процес и валидиране на данните - както първични, така и статистически. Тези практики обаче не са формализирани и документирани, което ограничава тяхната възпроизводимост. В този смисъл е налице необходимост от интегриране на добрите вътрешни практики в методологичен документ, достъпен за целия екип, с цел устойчивост и мащабируемост.

1.4.4. Осигуряване на разбираема структура за въвеждане на нови експерти

Промените в състава на екипа налагат бързо въвеждане в работата и разбиране на сложните изисквания по регламентите на ЕС. В момента това зависи основно от неформални инструкции и лични усилия на експертите. Налице е належаща нужда от изграждане на логически структуриран документ, включващ както основни статистически понятия, така и конкретни процедури, ръководства и шаблони за практическо приложение.

1.4.5. Подкрепа при въвеждане на нови модули

С измененията в европейските нормативни документи нараства натискът за въвеждане на

нови статистически модули с кратки срокове и специфични методологии (например сметките за екосистемите). Без ясно дефиниран процес по разработка, валидиране и координация, рискът от забавяне или непълнота на данните е сериозен. За да се намали този риск, е необходимо да се установи стандартизиран процес за разработка и интегриране на нови методологии и тематични направления, базиран на вече разработената структура.

2. Цели на Методиката

Разработването на Методиката за статистиката и сметките за околната среда отразява необходимостта от структурирано и устойчиво управление на статистическия процес в тази тематична област. Методиката е замислена като вътрешен инструмент за управление на процесите, който същевременно да бъде достатъчно гъвкав, за да отговаря на динамичните изисквания на международната статистическа рамка. Тя обединява на едно място концептуалната и практическата страна на производството на статистика за околната среда - от дефиниции и методи до процедури и отговорности.

Общата цел на Методиката е да осигури качествено, съгласувано и устойчиво производство на статистическа информация в областта на околната среда, съобразено с националните приоритети и ангажиментите на България като страна - членка на ЕС.

Специфичните цели включват:

- Унифициране на подхода за работа по всички изследвания в областта на екологичната статистика чрез внедряване на последователни и съгласувани методологични принципи;
- Подобряване на надеждността и съпоставимостта на данните, произвеждани от НСИ, чрез стандартизиране на процедури по събиране, валидиране, анализ и представяне;
- Повишаване на административния и експертен капацитет в НСИ чрез улесняване на въвеждането в работата, прехвърляне на знания и практическо обучение;
- Намаляване на административната тежест върху респондентите и служителите чрез ясно дефинирани процедури и оптимизирани потоци на информация;
- Осигуряване на методологична подкрепа при бъдещи промени, включително по отношение на нови тематични модули или регулаторни изисквания на Евростат.

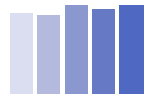
3. Обхват и съдържание на Методиката

3.1. Обхват

Методиката обхваща както общите принципи и подходи на статистическия процес, приложими в статистическата практика, така и специализираните методологии и процедури, валидни за конкретните изследвания за околната среда. Обхватът ѝ е структуриран в две основни части:

- Общи раздели - посветени на статистическата теория, процедурата за производство на данни, указания за валидиране и подходи към качеството;
- Приложения (специализирани раздели) - съдържащи методологии и критерии по всяко конкретно статистическо изследване (емисии във въздуха, статистика на отпадъците и водите, екологични стоки и услуги и така нататък).

Методиката се прилага както на централно, така и на регионално ниво в НСИ и е предназначена за всички експерти, участващи в производството, валидирането, анализа и разпространението на данни за околната среда. Тя изпълнява ролята на оперативна рамка, чрез която се постига както ефективност в ежедневната дейност, така и устойчивост при трансформации, иновации или промени в нормативната среда.



Методиката е съобразена както с националните и международните законодателни изисквания и приложимите методологии и насоки, така и със съответните метаданни и доклади за качеството на изследванията. Всяка глава съдържа препратки към използваните източници на информация по конкретната тема, както и списък на свързаните документи (законодателство, методологични документи), за да се насърчи по-нататъшното четене, като по този начин се подобряват разбирането и знанията по темата. Представени са илюстративни примери за таблиците за докладване на Евростат или на национално ниво, визуализиращи крайните резултати от всяко изследване.

Всички раздели са взаимосвързани, което означава, че за изпълнение на задълженията си по статистиката за околната среда с необходимото качество, експертите трябва да са запознати с Методиката в пълния ѝ обхват. Разделите относно общата статистическа методология са разработени в контекста на динамичната среда на човешките ресурси, като се има предвид, че може не всички експерти да имат статистическа подготовка и/или не всички от тях да имат еднакъв професионален опит. Познаването на общите основи на статистиката позволява по-добро разбиране на прилаганите методологии, включително дефинирането на техния обхват (напр. критерии за излъчване на извадка), използваните класификации, източниците на данни и техните специфики, степента, до която данните са детайлизирани и други. Познаването на методологията на едно изследване допринася за по-доброто и по-лесно събиране и валидиране на данните, което се очаква да доведе до намаляване на натоварването на експертите и респондентите. Прилагането на всички представени знания и практически насоки в ежедневната работа осигурява по-високо качество на статистическите данни, което е и общата цел на Методиката.

Цялата информация, включена в Методиката, е практически ориентирана, насочена към подобряване на статистическата дейност, към повишаване на качеството на данните и намаляване на тежестта. По този начин тя допълва цялостния инструментариум на НСИ за статистиката на околната среда.

3.2. Съдържание

3.2.1. Обща методология

Разделите относно общата методология представят основната теоретична рамка върху която се изгражда статистическият процес в областта на статистиката за околната среда. Въпреки че Методиката е насочена най-вече към практическото приложение, разбирането на основните статистически концепции е от съществено значение за поддържането на високо качество и съгласуваност в цялостната работа. Затова тук е включен необходимият минимум от понятия, принципи и подходи, съобразени с нуждите на експерти, които не разполагат с формално статистическо образование, но активно участват в производството на данни за околната среда.

3.2.1.1. Обща методология на статистическото изследване

Този раздел, разработен въз основа на приложимата нормативна уредба, Общия модел на статистическия производствен процес, внедрен в НСИ, и учебната литература съдържа обща теоретична информация за статистическото изучаване.

Общата теория на статистиката е наука за познавателната същност и общите методологични и организационни основи на статистическото изучаване. Системата от знания, които формира общата теория на статистиката, е теоретична основа на статистическите изучавания във всички области на действителността. Статистическата методология се разбира като система от техники и методи, целящи изучаване на количествени модели, проявяващи се в структурата, динамиката и

взаимовръзките на социално-икономическите явления.

Целта на раздела е да предостави основно разбиране за същността и характеристиките на статистическото изследване, необходимо за успешното осъществяване на статистическа дейност и производството на качествена статистическа информация. За нуждите на Методиката разделът е разработен като необходим минимум от информация и не изчерпва всички теоретични и практически аспекти на статистическото изследване.

Обобщението на тези теоретични, законодателни и практически указания в Методиката, допълнени от връзки към съответните документи и източници на информация, осигурява бърз, лесен и удобен за ползване начин за експертите да придобият и подобрят своите знания и експертиза и гарантира устойчиво внедряване и развитие на статистиката за околната среда.

3.2.1.2. Методологични насоки за извадковите изследвания и оценките

Извадковите изследвания са често използван метод в статистиката, особено когато е невъзможно или нецелесъобразно да се извърши пълно наблюдение върху всички единици от статистическата съвкупност - обективно съществуващото крайно множество, елементите (единиците) на което са определени за дадено място и време (момент или период). В тези случаи характеристиките на съвкупността се изследват чрез наблюдаване само на част (обикновено малка) от съвкупността, наричана проба или извадка.

В този раздел се представят основните принципи и практически насоки за прилагане на извадковия метод в рамките на статистическите изследвания за околната среда, провеждани от НСИ. Неговата цел е да предостави по-практическо разбиране по въпроса, което, в комбинация с теоретичната информация относно общата статистическа методология, служи като основа за провеждане на статистическата дейност в областта на околната среда.

Разработен въз основа на научна литература, този раздел притежава добавената стойност на подробното практическо обяснение на статистическия подход конкретно в статистиката за околната среда (с приведен пример за сметките за разходите за опазване на околната среда), включително формули, изчисления и сравнителна таблица с резултати. Това служи като отправна точка за прилагането на извадковия и оценъчния подход при производството на статистически данни за околната среда.

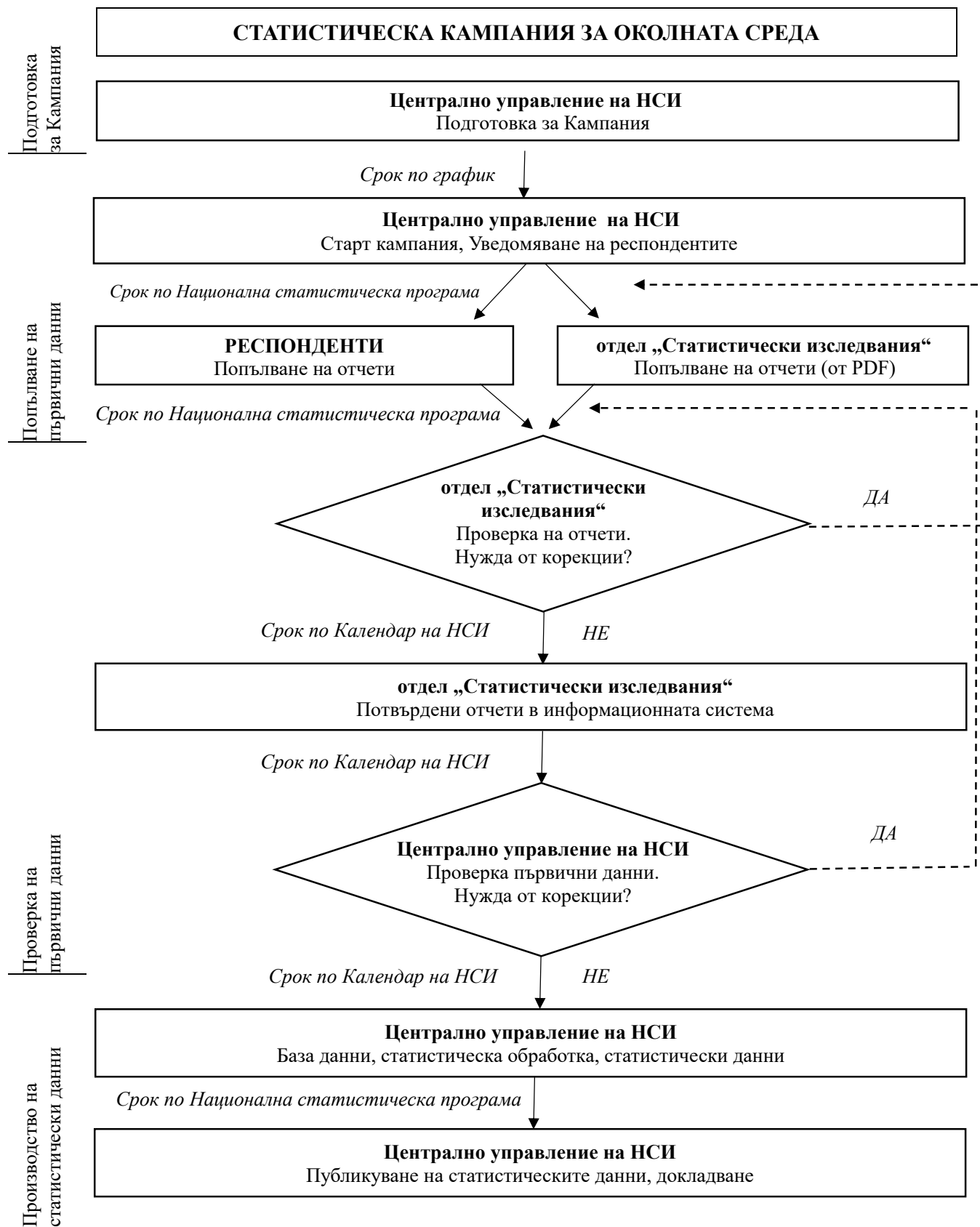
3.2.1.3. Процедура на статистическата дейност

Производството на статистическа информация за околната среда включва последователност от дейности, които са обвързани както с времеви график, така и с отговорности на различни екипи в рамките на НСИ. За да бъде процесът ефективен и устойчив, е необходимо той да бъде добре организиран, с ясно определени стъпки и координация между звената, ангажирани с конкретните изследвания. Цялата процедура е съобразена с Интегрираната система за управление, прилагана в НСИ, както и с европейските стандарти, заложи в Кодекса на европейската статистическа практика. Всеки етап от производствения процес е с ясна формулировка, времеви рамки и отговорни структури.

Целта на раздела, освен гарантиране прозрачността на процеса на производство на данни за околната среда, е да осигури познаване на процедурата, въз основа на която във всяка статистическа кампания за околната среда се създава оптимална организация на структурите на НСИ с цел навременно и качествено изпълнение на техните задължения. За онагледяване на процеса е разработена и блок-схема на процедурата, както и примерни образци за писмена кореспонденция.

Блок-схемата на статистическата кампания за околната среда е представена на фигура 1.

Фигура 1. Блок-схема на статистическата кампания за околната среда



3.2.1.4. Оценка на качеството

Качеството на статистическата информация е основен критерий за нейната полезност, надеждност и използваемост от заинтересованите страни, включително държавните органи, бизнеса, научните институции и обществеността. В областта на статистиката на околната среда това изискване е още по-значимо, тъй като данните се използват за формулиране, мониторинг и оценка на бързо развиващите се политики в контекста на Европейския зелен пакт, кръговата икономика и устойчивото развитие.

Разделът обобщава основните принципи и механизми за осигуряване на качеството в рамките на статистическата дейност за околна среда в съответствие с Кодекса на европейската статистическа практика, Политиката по качеството и на сигурността на информацията на НСИ, ръководствата на Евростат за докладване и качество, Интегрираната система за управление на качеството и сигурността на информацията на НСИ.

Конкретните указания за извършване на оценка на качеството за всяко отделно изследване са включени в съответните специализирани методологии.

3.2.2. Специализирана методология

Разделът представя методологиите и статистическия процес на изследванията в областта на околната среда, провеждани в НСИ. Всяко изследване има своя собствена специфика, но основната цел е една и съща - да се осигурят надеждни и съпоставими данни, които могат да бъдат използвани от институциите, бизнеса и обществото.

За да бъде информацията разбираема и прилагана по единен начин, за всяко изследване е създаден методологичен документ. Той съдържа описание на целта на изследването, какви данни се събират, от кого, колко често и по какъв начин. Подробно е обяснен начинът на изчисляване на статистическите данни и показатели, които България докладва на европейско, международно и национално ниво. Методологията обяснява също кои нормативни документи са в сила, какви класификации се използват и как се гарантира качеството на данните.

3.2.2.1. Методология на статистическите изследвания за околната среда

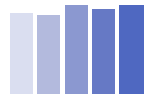
За всяко изследване за околната среда е разработена подробна методология, която се основава на приложимите нормативни разпоредби и специализирани методологични указания, като взема предвид и най-актуалната прилагана практика като част от общия процес по непрекъснато подобряване на статистиката за околната среда.

В началото на всяка методология се дава кратко представяне на изследването - цел, същност, предназначение, нормативното основание. След това се уточнява какъв е неговият обхват - статистическа съвкупност и статистическа единица, периодичност на събиране и разпространение на данните.

Внимание е отделено и на източниците на информация. В зависимост от изследването, това могат да бъдат статистическите формуляри на НСИ, организирани в информационна система „Статистика на околната среда“, и/или административни бази данни, и/или комбинация от тях.

Тук е важно да се отбележи, че формулярите по изследванията за околната среда се актуализират ежегодно, включително по отношение указанията за тяхното попълване от респондентите и методологичните бележки. Те се оповестяват на уебстраницата на НСИ и са част от инструментариума на информационната система, поради което не са включени в Методиката.

Методологията включва и описание на начина, по който се обработват и анализират данните. Обяснява се как се проверяват за грешки, как се обобщават, какви изчисления се правят и как се представят крайните резултати. Там, където се прилагат извадкови методи и оценки, се дава подробно описание.



Специално внимание се отделя на качеството, като се посочват общи насоки за неговото гарантиране. Подробните инструкции за валидиране на данните и минималните критерии за това са представени в отделен специализиран раздел на Методиката.

В края на всяка методология се включват полезни връзки - например към европейски регламенти или тематични ръководства на Евростат. Целта е всеки експерт, който работи с тази информация, да разполага не само с инструкции, но и с достъп до допълнителни източници за справка и обучение.

Този подход позволява всички участници в процеса - от новоназначени служители до опитни експерти - да прилагат единни правила и да допринасят за надеждността на статистиката за околната среда.

3.2.2.2. Инструкции и минимални критерии за валидиране на данните от статистическите изследвания за околната среда

Процесът на валидиране на данните представлява ключов етап от статистическото производство и е неразделна част от управлението на качеството. Целта му е да гарантира логическа, структурна и статистическа коректност на получените данни преди тяхното агрегиране, анализ и разпространение.

След като данните бъдат събрани, те не се използват веднага. Първо трябва да се провери дали информацията е пълна, вярна и логически последователна. Този процес се нарича валидиране и е много важен етап от статистическата работа. Той осигурява надеждност на резултатите и гарантира, че крайните стойности, които се публикуват и докладват, отразяват точно действителността.

За всяко изследване е разработена специфична и подробна информация, както и указания относно валидирането на данните (първични и статистически), което е необходимо да се извърши в рамките на статистическия процес. Чрез стриктното прилагане на посочените критерии се гарантира качеството на произведената статистическа информация. Включените критерии представляват необходимия минимум за проверка на данните и не ограничават извършването на други проверки по експертна преценка.

Валидирането се извършва чрез съчетание на автоматични и ръчни проверки. Част от грешките се улавят чрез вградени алгоритми в информационната система - например липсващи стойности в задължителни полета, несъответствия между свързани променливи и други. Останалите проверки се извършват от експерти, които анализират резултатите с професионална преценка, сравняват ги с предходни периоди или с други източници и преценяват дали има нужда от корекция или пояснение.

За всяко изследване съществува списък с така наречените „минимални критерии“, които трябва да бъдат изпълнени, за да може даден отчет да бъде приет като валиден. Те включват задължителни полета, логически връзки между стойности, приемливи диапазони за определени показатели и сравнения във времето. Например ако предприятие отчита разход за опазване на околната среда, се очаква също да има дейност, която поражда замърсяване или генерира отпадъци. Ако стойностите за даден показател се различават рязко от предходната година, се изисква обяснение.

Целта на тези минимални критерии не е да затруднят респондентите или експертите, а да създадат ясна и еднакво приложима основа, върху която да се оценяват и сравняват данните. Те са част от вътрешния контрол на качеството и представляват базова линия, под която резултатите не се приемат. Отделно от това, експертите имат възможност да прилагат и допълнителни проверки, ако открият несъответствия или имат основания да се усъмнят в коректността на данните.

Критериите за валидиране могат да бъдат актуализирани периодично. Причините могат да бъдат промени в европейските изисквания, натрупан практически опит, нови технологии за обработка или открити пропуски и неточности в предишни периоди. Затова се предвижда редовно преразглеждане и допълване на Методиката.

Този систематичен подход към проверката на данните позволява на НСИ да поддържа високи стандарти и доверие сред потребителите на информация. Качествената статистика е основа не само за научен анализ, но и за вземане на решения в реални политики - особено когато става въпрос за околната среда, където въздействията са дългосрочни и често необратими.

4. Постигнати резултати

Разработването и въвеждането на Методиката за статистиката и сметките за околната среда води до конкретни и измерими резултати както в организационен, така и в методологичен план. Тези резултати отразяват усилията на екипа, участващ в реализацията на проекта, както и създадената основа за бъдещо развитие на статистиката в тази област.

Методиката е официално утвърдена със заповед на председателя на НСИ и се използва като вътрешен инструмент за управление на статистическата дейност, свързана с околната среда. Това представлява сериозна крачка напред, тъй като за първи път всички процеси, свързани с планиране, събиране, валидиране, анализ и докладване на данни за околната среда, са обединени в един последователен и логически структуриран документ.

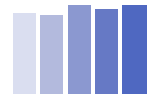
Практическото прилагане на Методиката вече показва реални ползи - подобрена координация, намалени грешки при събирането и обработката на данни, по-бързо обучение на нови служители и по-добра проследимост на извършените дейности. Наблюдава се по-ясно разграничаване на отговорностите между централното и регионалното ниво, както и между различните функционални звена.

Паралелно с утвърждаването на Методиката е извършена и цялостна актуализация на съпътстващите документи - включително методологиите на отделните изследвания, метаданните и докладите за качество, които вече са публикувани в актуализиран вид на интернет страницата на НСИ и са достъпни за всички заинтересовани страни. Това подобрява прозрачността на статистическата дейност и улеснява комуникацията с респонденти, потребители на данни и външни партньори.

Особено важно е, че чрез прилагането на Методиката се очаква подобрене в качеството на данните, които се докладват към Евростат, съгласно регламентите по статистика в областта на околната среда. Това включва както съгласуваност и последователност на данните, така и по-добро покритие и навременност. По този начин НСИ отговаря по-пълноценно на своите ангажименти в рамките на Европейската статистическа система.

Друг важен резултат от проекта е укрепването на експертния капацитет на НСИ. Не само е проведено специално обучение на териториалните експерти по прилагането на Методиката, но и служителите на централно ниво получават възможност както да участват в създаването на Методиката, така и да я използват като средство за собствено професионално развитие. Практическата насоченост на документа позволява той да се използва и за обучение на новоназначени служители, и за уеднаквяване на подходите при изпълнение на ежедневните задачи.

В обобщение, резултатите от разработването и прилагането на Методиката не се ограничават до създаването на един нов документ. Те се изразяват в съществено подобрене на целия статистически процес, повишаване на доверието към екологичната статистика и създаване на основа за устойчиво развитие на системата в следващите години. Това от своя страна намалява административната тежест за респондентите и същевременно позволява на експертите да отделят повече внимание на анализа и тълкуването на информацията.



Заклучение

Разработването и внедряването на Методиката за статистиката и сметките за околната среда бележи важен етап в развитието на екологичната статистическа дейност в България. В процеса на изпълнение на проекта са постигнати не само формално заложените цели, но и редица съпътстващи ползи, които ще окажат дългосрочно въздействие върху качеството, съгласуваността и устойчивостта на статистическите процеси.

На първо място, Методиката отговаря на реална и отдавна идентифицирана потребност - тази за уеднаквяване, систематизиране и документиране на дейностите по производство на статистически данни за околната среда. Чрез нея в единен формат са обединени както общите принципи на статистическото производство, така и спецификата на всяко отделно изследване. Това осигурява необходимата последователност, яснота и предвидимост за всички участници в процеса - от експертите в НСИ до външните потребители на данните.

На второ място, Методиката изпълнява не само краткосрочната цел за създаване на писмено ръководство, но и стратегическа функция - тя служи като инструмент за обучение, за въвеждане в работата на нови експерти и за поддържане на институционалната памет. В условията на комплексни изисквания, нарастващо международно координиране и при промяна на експертния състав, подобен инструмент е от ключово значение за гарантиране на устойчивост на натрупаното знание.

Процесът по създаване на Методиката не е лесен. Най-същественото предизвикателство се оказва съчетаването на теоретичната коректност с практическата приложимост. Необходимостта от яснота, без да се жертва прецизността, наложи внимателен подбор на формулировки, примери и ниво на детайлност. Това изисква активна работа в екип, интензивна вътрешна комуникация и често пренаписване на раздели. Въпреки това, усилията дават резултат - документът е вече част от ежедневната работа на експертите и се използва като реален инструмент, а не само като формално изискване.

Създаването на Методиката поставя солидна основа за бъдещи действия и подобрения. Сред тях могат да се открият няколко препоръки:

- Да се въведе редовен преглед на Методиката с участието на всички звена, които я прилагат на практика. Това ще позволи навременна корекция на пропуски и допълване с нови елементи;
- Да се използва Методиката като основа при разработване на нови изследвания, като се адаптира структурата към други статистически области, свързани с околната среда;
- Да се продължи изграждането на капацитет чрез регулярни обучения, базирани на съдържанието на Методиката. Така тя ще изпълни ролята си и като учебен ресурс;
- Да се запази възможността за двупосочна обратна връзка - както между експертите на централно и регионално ниво, така и от респондентите към експертите. Това ще осигури практическа релевантност на документите и ще поддържа жив процеса на подобрене.

В заключение, Методиката е не просто резултат от изпълнението на един проект. Тя е свидетелство за професионализъм, ангажираност и визионерски подход към статистиката за околната среда. Този документ има потенциала да се превърне в стандарт за работа и пример за други области, където е необходима яснота, устойчивост и качество при управление на статистическия процес.

Признателност

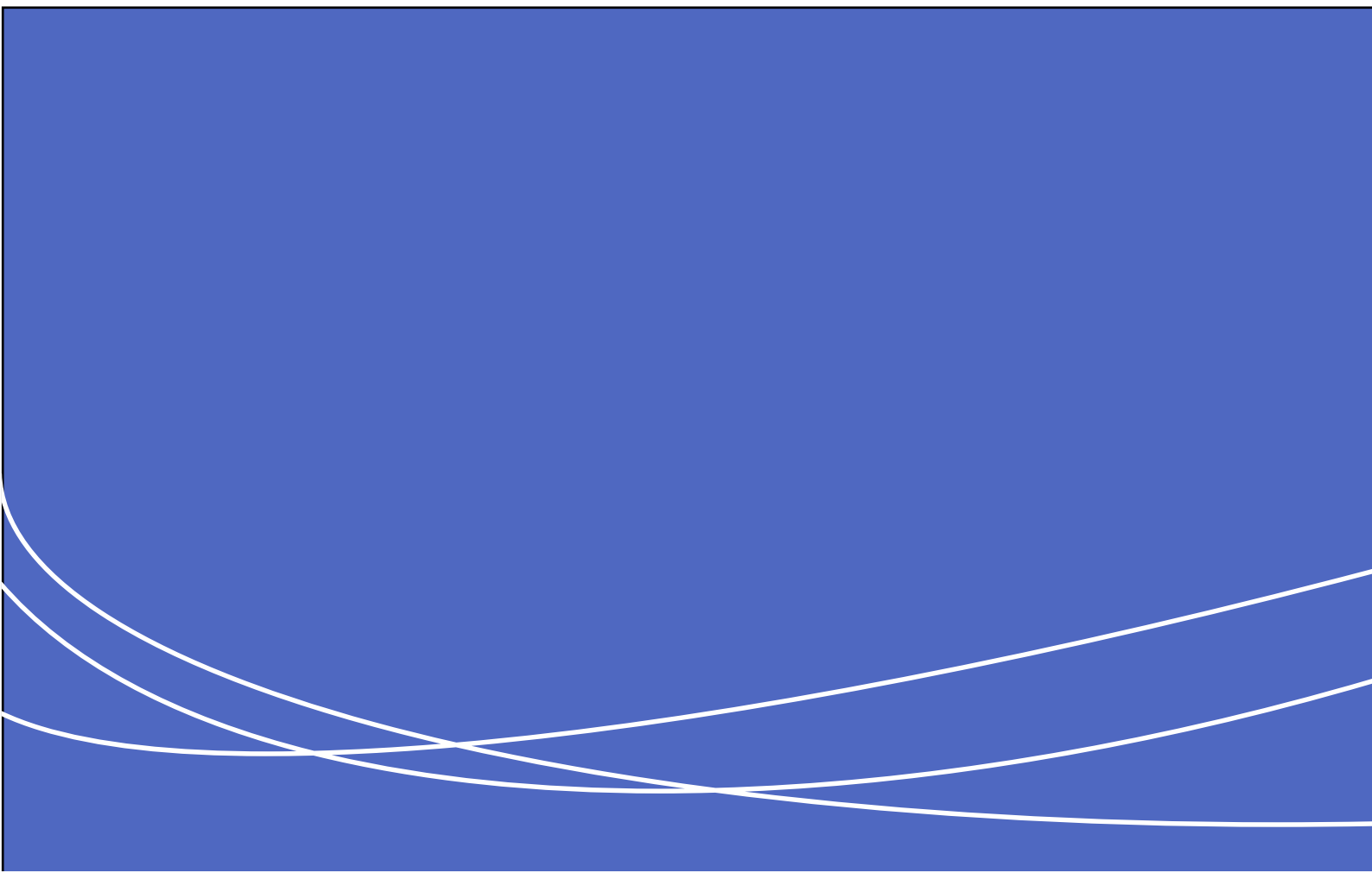
НСИ изразява признателността си към ЕК за финансовата подкрепа, чрез която е постигнато не само подобрене на статистиката за околната среда в България, но и взаимна подкрепа в съвместната работа за устойчиво развитие и добро сътрудничество на национално и европейско ниво.

Референции:

1. Гатев, К. (1995). Въведение в статистиката, София, ЛИА;
2. Калоянов, Т. (2004). Статистика, София, Тракия-М;
3. Цонев, В. (1970). Основи на репрезентативното изучаване, София, ВИИ „К. Маркс”;
4. Европейска нормативна уредба на уебсайта на НСИ (достъпно на: <https://nsi.bg/documents/218>);
5. Българска нормативна уредба на уебсайта на НСИ (достъпно на: <https://nsi.bg/documents/224>);
6. Регламент (ЕС) № 691/2011 относно европейските икономически сметки за околната среда (достъпно на: <https://eur-lex.europa.eu/legal-content/BG/ALL/?uri=CELEX:32011R0691>);
7. Система за екологично икономическо счетоводство (достъпно на: <https://seea.un.org/>);
8. Вътрешноинституционална рамка и политики на НСИ (достъпно на: <https://nsi.bg/documents/229>);
9. Технически доклад за работен пакет 1 от проект по Споразумение за субсидия № 101112756, РД 10-6/18.04.2023 година.

**БИЗНЕС ИНТЕЛИГЕНТНИ СИСТЕМИ И
ИЗКУСТВЕН ИНТЕЛЕКТ**

**BUSINESS INTELLIGENT SYSTEMS AND
ARTIFICIAL INTELLIGENCE**



ИНТЕГРИРАНЕ НА ДАННИ ОТ ИНТЕРНЕТ В ОФИЦИАЛНА СТАТИСТИКА: ПРОЕКТЪТ „ДОВЕРЕНА УМНА СТАТИСТИКА (TSS) - МРЕЖА ЗА ИЗУЧАВАНЕ НА УЕБПРОСТРАНСТВОТО (WIN)“

Мариана Ангелова, Костадин Георгиев***

РЕЗЮМЕ Въведение: Експоненциалният растеж на уеб данните, включително онлайн обяви за работа, фирмени уебсайтове, ценови портали и др. платформи, създава безпрецедентна възможност за модернизиране на официалната европейска статистика. Независимо от това, интегрирането на такива динамични, неструктурирани източници в строгите рамки за качество на националните статистически институти остава методологично и техническо предизвикателство.

Цел и задачи: Финансираната от ЕС мрежа ESSnet Trusted Smart Statistics - Web Intelligence Network (ESSnet WIN, 2021 - 2025) обединява 17 европейски статистически института, които да проектират, изградят и пилотират Web Intelligence Hub (WIH): облачна платформа и методологичен комплекс, способни да събират, обработват и гарантират качеството на уеб данни за рутинно статистическо производство.

Методология: Проектът е структуриран от четири работни пакета: (WP1) управление и обучение на общност от над 70 експерти (WISER); (WP2) надежден софтуер за онлайн обяви за работа (OJA) и онлайн-базирани характеристики на предприятията (OBEC); (WP3) проучвателни примери в недвижимите имоти, строителството, туризма, цените и бизнес регистрите; (WP4) хармонизирани насоки за качество и методология, залежали в модулната референтна архитектура за големи данни BREAL и многопластов работен процес за осигуряване на качеството (ландшафтен дизайн → уеб скрапинг → вземане на проби и обработка на данни → класификация на машинно обучение → валидиране спрямо традиционни източници).

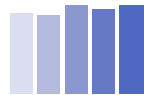
Основни резултати: WIH работи с AWS Athena; OJA и OBEC потоците се изпълняват на тримесечие за над 15 държави; OJA: 25 000 - 60 000 активни реклами/месец, класифицирани в региони ISCO-1 (56 - 62% точност) и NUTS-2; остават нестабилни източници и пропуски в покритието; OBEC: 90% + URL връзки на фирми към уебсайтове; 70 - 80% точност при откриване на електронна търговия, многоезичност или присъствие в социалните медии; Нови пилотни проекти: месечни индекси на наеми/цени за 5 държави; показатели за ново строителство за Германия/Швеция; индекси на хотелските цени от Booking.com; отчитане на трафика от камери като високочестотни показатели за икономическа активност; 9 уебинара, 27 блога, Хакатон и заключителна конференция, ангажирали над 1 200 практикуващи.

Заключения: Уеб данните могат да обогатят, а не да заменят традиционните източници, когато се комбинират с административни регистри и непрекъснат контрол на качеството. Хибридните работни процеси, адаптивните модели за машинно обучение и инструментите с отворен код са от съществено значение за надеждността и мащабируемостта.

Ограничения: Точността на класификацията на машинното обучение, правните/етичните ограничения (GDPR), нестабилните източници и неравномерното географско покритие ограничават незабавната оперативна употреба.

* Началник на отдел „Управление на метаданни, методология и регистри“, Национален статистически институт, гр. София, България, e-mail: mangelova@nsi.bg

** Главен експерт в отдел „Информационни системи и приложен софтуер“, Национален статистически институт, гр. София, България, e-mail: kgeorgiev@nsi.bg



Практически приложения: Националните статистически институти могат да вграждат WIN модули за мониторинг на пазара на труда, бизнес демографията, нарастването на ИПЦ и на индикаторите за кризи. Споделените уеб обхождащи програми, хармонизираните показатели за качество и откриването с помощта на големи езикови модели (LLM) ще намалят дублирането и разходите.

Оригиналност/стойност: ESSnet WIN предоставя първата паневропейска, готова за производство архитектура и тествани насоки за надеждна интеграция на уеб данни в официалната статистика, предоставяйки план за многократна употреба за националните статистически институти по целия свят.

Ключови думи: интеграция на уеб данни, официална статистика, Web Intelligence Hub, класификация с машинно обучение, хибридни източници на данни

ИНТЕГРАЦИЯ ИНТЕРНЕТ-ДАННЫХ В ОФИЦИАЛЬНУЮ СТАТИСТИКУ: ПРОЕКТ «НАДЁЖНАЯ ИНТЕЛЕКТУАЛЬНАЯ СТАТИСТИКА (TSS) - ВЕБ-ПРОСТРАНСТВЕННЫЙ ИНТЕЛЛЕКТ (WIN)»

Мариана Ангелова, Костадин Георгиев***

РЕЗЮМЕ Введение: Экспоненциальный рост веб-данных, включая онлайн-объявления о вакансиях, корпоративные веб-сайты, ценовые порталы и другие платформы, создаёт беспрецедентные возможности для модернизации официальной европейской статистики. Тем не менее интеграция таких динамичных, неструктурированных источников в строгие рамки качества национальных статистических институтов остаётся методологическим и техническим вызовом.

Цель и задачи: Финансируемая ЕС сеть ESSnet Trusted Smart Statistics - Web Intelligence Network (ESSnet WIN, 2021 - 2025) объединила 17 европейских национальных статистических институтов для проектирования, создания и пилотирования Web Intelligence Hub (WIH) - облачной платформы и методологического комплекса, способных собирать, обрабатывать и обеспечивать качество веб-данных для регулярного статистического производства.

Методология: Проект структурирован на четыре рабочих пакета:

(WP1) управление, обучение сообщества из более чем 70 экспертов (WISER); (WP2) надёжное программное обеспечение для онлайн-объявлений о вакансиях (OJA) и онлайн-базируемых характеристик предприятий (OBEC); (WP3) исследовательские примеры в сфере недвижимости, строительства, туризма, цен и бизнес-регистров; (WP4) гармонизированные руководства по качеству и методологии, заложенные в модульную эталонную архитектуру больших данных BREAL и многоуровневый рабочий процесс обеспечения качества (ландшафтный анализ → веб-скрейпинг → выборка и обработка данных → классификация с помощью машинного обучения → валидация по традиционным источникам).

Основные результаты: WIH функционирует на AWS Athena; конвейеры OJA и OBEC запускаются ежеквартально для более чем 15 стран; OJA: 25 000 - 60 000 активных объявлений в месяц, классифицированных по ISCO-1 (точность 56 - 62%) и регионам NUTS-2; сохраняются нестабильные источники и пробелы в охвате; OBEC: более 90% привязки URL-адресов компаний к веб-сайтам; 70 - 80% точность при выявлении электронной торговли, многоязычного или социально-медийного присутствия; Новые пилотные проекты: ежемесячные индексы арендной платы/цен для 5 стран; ранние строительные индикаторы для Германии/Швеции; индексы цен на отели по данным Booking.com; данные с дорожных камер как высокочастотные прокси-показатели экономической активности; Проведено 9 вебинаров, опубликовано 27 блогов, проведены Хакатон и заключительная конференция с участием более 1 200 специалистов-практиков.

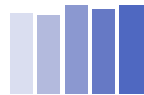
Выводы: Веб-данные могут дополнять, а не заменять традиционные источники, при условии их комбинирования с административными регистрами и постоянным контролем качества. Гибридные рабочие процессы, адаптивные модели машинного обучения и инструменты с открытым исходным кодом являются ключевыми для надёжности и масштабируемости.

Ограничения: Точность классификации с помощью машинного обучения, правовые и этические ограничения (GDPR), нестабильные источники и неравномерный географический охват ограничивают немедленное операционное применение.

Практические результаты: Практические применения: Национальные статистические институты могут внедрять модули WIH для мониторинга рынка труда, бизнес-демографии, роста ИПЦ и

* Начальник отдела «Управление метаданными, методология и регистры», Национальный статистический институт, София, Болгария, e-mail: mangelova@nsi.bg.

** Главный специалист отдела «Информационные системы и прикладное программное обеспечение», Национальный статистический институт, София, Болгария, e-mail: kgeorgiev@nsi.bg.



показателей кризиса. Совместное использование веб-сканеров, гармонизованных показателей качества и поиск данных с использованием больших языковых моделей (LLM) сократят дублирование и затраты.

Оригинальность / ценность: ESSnet WIN предоставляет первую панъевропейскую, готовую к промышленной эксплуатации архитектуру и протестированные руководства по надёжной интеграции веб-данных в официальную статистику, предлагая многоразовый план для национальных статистических институтов по всему миру.

Ключевые слова: интеграция веб-данных, официальная статистика, Web Intelligence Hub, классификация с помощью машинного обучения, гибридные источники данных

INTEGRATING WEB DATA INTO OFFICIAL STATISTICS: THE 'TRUSTED SMART STATISTICS (TSS) -WEB INTELLIGENCE NETWORK (WIN)' PROJECT

Mariana Angelova*, Kostadin Georgiev**

SUMMARY Introduction: The exponential growth of web data, including online job adverts, firm websites, price portals, creates an unprecedented opportunity to modernize official European statistics. Regardless of this, integrating such dynamic, unstructured sources into the rigid quality frameworks of the national statistical institutes (NSIs) remains methodologically and technically challenging.

Aim and objectives: The EU-funded ESSnet Trusted Smart Statistics - Web Intelligence Network (ESSnet WIN, 2021 - 2025) united 17 European national statistical institutes to design, build and pilot the Web Intelligence Hub (WIH): a cloud platform and methodological stack able to harvest, process and quality-assure web data for routine statistical production.

Methodology: The project is structured of four work packages: (WP1) governance, training and community of 70+ experts (WISER); (WP2) production-grade software for Online Job Adverts (OJA) and Online-Based Enterprise Characteristics (OBEC); (WP3) exploratory use-cases in real-estate, construction, tourism, prices and business registers; (WP4) harmonized quality and methodological guidelines anchored in the modular BREAL big-data reference architecture and layered quality-assurance workflow (landscaping → web-scraping → sampling & data-wrangling → ML classification → validation against traditional sources).

Main results: WIH is operational on AWS Athena; OJA and OBEC pipelines run quarterly for 15+ countries; OJA: 25 k - 60 k live adverts/month classified to ISCO-1 (56 - 62% accuracy) and NUTS-2 regions; unstable sources and coverage gaps remain; OBEC: 90% + URL linkage of firms to websites; 70 - 80% accuracy detecting e-commerce, multilingual or social-media presence; New pilots: monthly rent/price indices for 5 countries; early construction indicators for Germany/Sweden; hotel-price indices from Booking.com; traffic-camera counts as high-frequency economic activity proxies; 9 webinars, 27 blogs, hackathon and final conference engaged > 1 200 participants.

Conclusions: Web data can enrich, not replace, traditional sources when combined with administrative registers and continuous quality control. Hybrid workflows, adaptive ML models and open-source tooling are essential for credibility and scalability.

Limitations: Accuracy of ML classification, legal/ethical constraints (GDPR), volatile sources, and uneven geographic coverage constrain immediate operational use.

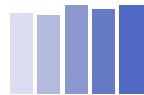
Practical implications: NSIs can embed WIH modules for labour market monitoring, business demography, CPI augmentation and crisis indicators. Shared crawlers, harmonized quality metrics and LLM-assisted discovery will reduce duplication and costs.

Originality/value: ESSnet WIN delivers the first pan-European, production-ready architecture and tested guidelines for trustworthy web-data integration into official statistics, providing a reusable blueprint for NSIs worldwide.

Keywords: web data integration, official statistics, Web Intelligence Hub, machine learning classification, hybrid data sources

* Head of Metadata, Methodology and Registers Management Department, National Statistical Institute, Sofia, Bulgaria, e-mail: mangelova@nsi.bg

** Chief expert in Information Systems and Application Software Department, National Statistical Institute, Sofia, Bulgaria, e-mail: kgeorgiev@nsi.bg



ИНТЕГРИРАНЕ НА ДАННИ ОТ ИНТЕРНЕТ В ОФИЦИАЛНА СТАТИСТИКА: ПРОЕКТЪТ „ДОВЕРЕНА УМНА СТАТИСТИКА (TSS) - МРЕЖА ЗА ИЗУЧАВАНЕ НА УЕБПРОСТРАНСТВОТО (WIN)“

Мариана Ангелова*, Костадин Георгиев**

Въведение

Експоненциалният растеж на данни от интернет създава безпрецедентни възможности за подобряване на традиционното статистическо производство. Независимо от това, интегрирането на неструктурирани, динамични данни от интернет в строгите рамки на официалната статистика поставя значителни методологични, технически и свързани с качеството предизвикателства.

Проектът „Доверена умна статистика (TSS) - мрежа за изучаване на уебпространството (WIN)“ (ESSnet Trusted Smart Statistics - Web Intelligence Network (ESSnet WIN) (април 2021 - март 2025 г.), финансиран от Европейския съюз, цели да отговори на предизвикателствата в Европейската статистическа система (ECC). ESSnet WIN е развитие на постиженията на проектите ESSnet Big Data I и ESSnet on Big Data II проведени през периода 2016 - 2020 година. За четири години проектът обединява 17 национални статистически института и партньорски организации от цяла Европа. Фокусът на проекта е да разработи и тества Европейска система за събиране и анализ на уебданни (WIN) (облачна среда, управлявана от AWS Athena), която като модерно средство да служи като платформа за събиране, анализиране и обработка на данни от мрежата и превръщането ѝ във висококачествен източник на уебданни, методологии и алгоритми, подготвени да бъдат използвани за производство на официална иновативна статистика на европейско и национално ниво. Проектът включва следните направления на работа:

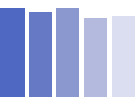
- разработване на WIN за мащабно, централизирано извличане на данни от интернет;
- напредък на WIN и внедряване на случаите на използване: онлайн обяви за работа (OJA); и онлайн базирани характеристики на предприятията (OBEC) в етапа на статистическо производство;
- проучване на потенциала за разширяване на WIN чрез използване на нови източници на данни от интернет за недвижими имоти, туризъм, строителство, статистика на цените и икономически индикатори;
- разработване на методологични документи и документи за качество за гарантиране надеждност и съответствие със стандартите на официалната статистика.

Проектът е структуриран в четири взаимосвързани работни пакета (WPs), всеки от които се фокусира върху различни аспекти на интеграцията на данни от интернет:

- WP1: Изграждане на система за събиране и анализ на уебданни (WIN) в рамките на ECC и извън нея, с фокус върху изграждане на компетенции, споделяне на знания, подкрепа на потребителите и ангажираност;

* началник на отдел „Управление на метаданни, методология и регистри“, Национален статистически институт, гр. София, България, e-mail: mangelova@nsi.bg

** главен експерт в отдел „Информационни системи и приложен софтуер“, Национален статистически институт, гр. София, България, e-mail: kgeorgiev@nsi.bg



- WP2: Напредък на WIH и прехвърляне на случаите OJA и OBEC в статистическо производство;
- WP3: Изследване на нови източници на данни и случаи на приложение, за да се разширят възможностите на WIH;
- WP4: Разработване на методологични и качествени основи за генериране на статистика в рамките на WIH.

Консорциумът на проекта включва следните статистически институти и партньори:

- Główny Urząd Statystyczny (GUS) - Полша (PL)
- Bundesanstalt Statistik Oesterreich (STATA) - Австрия (AT)
- Национален статистически институт (NSI) - България (BG)
- Tilastokeskus (TILASTOKESKUS) - Финландия (FI)
- Institut National De La Statistique Et Des Etudes Economiques (INSEE) - Франция (FR)
- Direction De L'animation De La Recherche Et Des Etudes Statistiques (DARES) - Франция (FR)
- Amt Fur Statistik Berlin-Brandenburg (SSI - BBB) - Германия (DE)
- Statistisches Bundesamt (DESTATIS) - Германия (DE)
- Hessisches Statistisches Landesamt (HSL) - Германия (DE)
- Istituto Nazionale di Statistica (ISTAT) - Италия (IT)
- Lietuvos Statistikos Departamentas (SL) - Литва (LT)
- Centraal Bureau Voor De Statistiek (CBS) - Холандия (NL)
- Instituto Nacional De Estatistica Portugal Statistics (POR) - Португалия (PT)
- Statistiska Centralbyran (SCB) - Швеция (SE)
- Statisticni Urad Republike Slovenije (SURS) - Словения (SI)
- UK Office for National Statistics (ONS) - Обединеното кралство (UK)
- Swiss Federal Statistical Office (FSO) - Швейцария (CH)

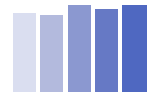
Географското и институционално разнообразие позволява кръстосано между участващите страни тестване на методологии (например OJA в Нидерландия/Австрия/Полша/Германия) и осигурява широка приложимост на резултатите.

Основно изложение

Работен пакет: WP1: Координация, подкрепа и разпространение

Основна цел на този работен пакет е гарантиране на ефективното управление на проекта, повишаване на осведомеността относно WIH и нейния потенциал за производство на европейска и национална статистика, споделяне на знания и развитие на умения и изграждане на капацитет в рамките на ЕСС.

За изпълнение на задачата за ефективно управление на проекта, координация, работа в мрежа и изграждане на общност се провеждаха редовни двумесечни срещи, водени от координатора на проекта (Полската статистика), с участието на лидерите на работните пакети, представители на Евростат и ключови заинтересовани страни, за да се синхронизират задачите с разработката на WIH. Целта е събиране на актуална информация за напредъка в работата и последващите резултати да се



разпространят в рамките на проекта и извън него. Вътрешната колаборация се осъществява чрез използване на Wiki на ESSnet WIN за споделяне на документи и проследяване на проблеми.

Създаден е и рецензентски борд (група WISER) включващ независими експерти, които оценяват ключовите резултати, за да гарантират качество. Рецензентският борд е общност от повече от 70 експерти (статистически институти, академични среди, международни организации), с основна цел ръководство на разработката на WIN, тестване на инструменти и предоставяне на обратна връзка. Бордът провежда успешно обучение за WIN, OJA и OBEC инструменти и поредица от сесии за обратна връзка за усъвършенстване на изискванията, в резултат на които са реализирани повече от 20 предложения.

В изпълнение на комуникационни дейности е изградена и изпълнявана стратегия за позициониране на WIN като инструмент за модернизация на статистическото производство. Основните дейности включват:

- Социални мрежи: Платформи: X (Twitter) и LinkedIn - по време на изпълнение на проекта са направени 207 публикации, 732 772 прегледа на X и 126 341 прегледа на LinkedIn. Последователите са 219 за X и 763 за LinkedIn, като през 2024 г. има 278 нови последователи в LinkedIn. През декември 2022 г. стартира YouTube канал, който има 10 видеа, 2 191 прегледа, 79 абонати, 400 часа гледане;
- Блогове: 27 публикации на портала CROS, обхващащи всички WP и случаи на приложение (например „Измерване на строителна активност чрез онлайн данни“, „Използване на големи данни за статистика на туризма“).

Успешно е организиран Хакатон - шестседмично онлайн предизвикателство (есен 2024 г.) с участието на десет екипа. Участниците разработват инструменти с отворен код за откриване на електронна търговия/социални мрежи на уебсайтове на предприятия, използвайки набор от данни от 4000 URL (Нидерландия, Австрия, Полша, Германия).

Комуникационните дейности на международни събития обхващат координиране на присъствието на членовете на проекта в рамките на международни събития и конференции. В рамките на тези дейности са представени повече от 14 презентации и е проведена заключителна конференция на проекта в гр. Гданск, Полша през месец февруари 2025 г. с повече от 80 участници. Заключителната конференция отчита, че дейностите по проекта са голяма стъпка към използването на убеданни в официалната статистика, въпреки че осигуряването на качеството на данните не е лесна задача. Повишаването на знанията и компетентността в ЕСС изгражда потенциал на общността за използване на убеданни при производството на статическа информация.

Работен пакет: WP2: Разработка на софтуер за OJA и OBEC

Целта на този работен пакет е да доразвие случаите на използване за OJA и OBEC в производствена среда. Работната програма включва четири основни задачи: разработване на софтуер за събиране, обработка и анализ на данни на OBEC; разработване на софтуер за събиране, обработка и анализ на OJA данни; подготовка на предложените показатели за преминаване към производствена среда и прилагане на изискванията, дефинирани от групата WISER.

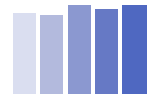
Случай на приложение OJA - онлайн обяви за свободни работни места

Фокусът при приложението на онлайн обявите за свободни работни места е експериментална

статистика за търсенето на работна сила. Изпълнените основни дейности включват усъвършенствани структурни метаданни за онлайн обявите за свободни работни места, дефинирани правила за качество на данните (например дубликати, отворени/затворени обяви), и валидирани URL за извличане. Разработени са експериментални променливи „професия“ от онлайн обявите за работа на първо ниво от Международната класификация на професиите (ISCO) и „регионални тенденции“ (ниво 2 на Класификация на териториалните единици за статистически цели - NUTS). С помощта на WIH е проведено анализиране и оценяване на качеството на данните.

Инициативата „Онлайн обяви за свободни работни места“ (OJA), разработена в рамките на Европейска система за събиране и анализ на уедбанни (WIN), използва уебизвличане и анализ на данни за получаване на информация за пазара на труда от онлайн портали за работа в европейските държави. Данните са стандартизирани, хармонизирани и направени достъпни чрез WIN. По-долу е направен синтез на ключови аспекти на OJA:

- **Източник на данни и достъп** - данните на OJA са онлайн обяви за работа в реално време извлечени от портали за работа и платформи в цяла Европа. Достъпът на потребителите се осъществява чрез скриптове на Python/R и SQL заявки в WIN. Основният набор от данни се намира в схемата `wih_oja_latest`, като таблицата `wih_oja_blended` е основната таблица.
- **Ключовите атрибути са:**
 - метаданни за обявата - първата дата на регистрация и последната дата на регистрация, език и стабилност на източника;
 - геопространствени атрибути: държава, региони NUTS (ниво 2 и 3 на NUTS), идентификатори на града;
 - професия - (ниво 1 и 2 от ISCO), умения (ниво 0 и 1) и образователна степен (Международна стандартна класификация на образованието);
 - икономическа активност - икономическа дейност по кодове на Европейската класификация на икономическите дейности (NACE), вид на договора за наемане на работа, размер на работната заплата.
- **Методологична рамка:**
 - потокът от данни с необработените обяви за работа се обработват чрез алгоритми за машинно обучение (ML) за класифициране на променливите професия, умения и икономически дейности по съответните класификации ISCO и NACE;
 - оценяване на качеството - данните на OJA все още не са класифицирани като официална статистика, а служат като експериментални показатели. Предизвикателство при проведените аотирания е ръчна валидация при кръговете на аотиране и разкрива точност между 40 - 70% при класифицирането по ISCO (първо ниво). Получените резултати изискват предпазливост при интерпретацията. Относно представителността на данните се наблюдават разлики с традиционните статистически изследвания (например Изследване на работната сила) поради отклонения в обхвата на източника и наличието на динамика при порталите за онлайн обяви за свободни работни места.
- **Ключови случаи на приложение** - при анализа на пазара на труда могат да се проследяват тенденциите в обявите за работа по професия, регион и сектор на икономическа дейност. Предимство на използването на данните от OJA е възможността за идентифициране на нови



умения и сезонните колебания. Регионалното разпределение на обявите за работа е полезно за регионалната политика, например възможността да се направи справка за броя на обявите за работа по държава и по NUTS ниво 3 за конкретни професии (например „Химици“ в Белгия). Генерирането на тримесечни/годишни тенденции (например обяви за работа през първото тримесечие на 2023 г.) с корекции за припокриващи се обяви през годините е предимство при разглеждането на динамичните редове.

- **Предизвикателства и ограничения** - основно качество на данните и различията в обхвата. Качеството на данните е ограничено от класифицирането, базирано на машинно обучение, което е несигурно, особено за подробни кодове за професии/умения. За публикуваните променливи се препоръчват доверителни интервали (относителни дялове). Друг съществен фактор за недоброто качество е стабилността на източниците, която варира, тъй като е възможно някои портали да изчезнат или да променят форматите си, което да повлияе на сравнителния анализ. Пропуските в обхвата са породени от нееднаквото представяне на всички държави/региони, а различната популярност на порталите за онлайн обяви за работа и съвместимостта при извличане на данни влияят върху обема на данните. Дезагрегирането на публичен и частен сектор не е налично поради недостатъчни и непълни метаданни.
- **Обратната връзка от потребителите** показва необходимост от консолидирана, актуализирана документация за ОЈА скриптове (Python/R/SQL) и метаданни (напр. методи за машинно обучение). За подобряването на метаданните е необходимо да се изпълни изискването за прозрачност относно точността (прецизност) на показателите за качество на машинното обучение. Поради проявения засилен интерес към създадените таблици за дезагрегиране по професии, умения и регионални тенденции се създава необходимост от „табло за управление“.
- **Примерни приложения** - Мониторингът на търсенето на умения в региони (например на трето ниво на NUTS) е полезен при съгласуване на програмите за обучение, което е съществено за подкрепата на политиките. При икономически изследвания сравнението на тенденциите на ОЈА с традиционни статистически изследвания е важно условие за идентифициране на несъответствия на пазара на труда. При експериментални резултати се предлагат таблици на променливата професия на 1 ниво на ISCO с доверителни интервали, за да се отчетат грешките при класифицирането.

В заключение може да се обобщи, че ОЈА предоставя динамичен, подробен поглед върху пазара на труда, но изисква внимателна интерпретация поради методологични ограничения. Допълнителният му принос и добавена стойност е в допълването на традиционната статистика с актуални, уеб-извлечени данни, особено за нововъзникващи професии и регионален анализ. Бъдещата работа цели подобряване на документацията, прозрачността на метаданните и индикаторите за качество.

Случай на приложение ОБЕС - онлайн базирани характеристики на предприятията

Фокусът при приложението извличане на онлайн характеристики на предприятията (ОБЕС) е откриване на предприятия, които имат уебсайтове, присъстват в социалните медии, имат многоезични сайтове, извършват електронна търговия от техен сайт и подобни.

Инициативата „Онлайн базирани характеристики на предприятия“ (ОБЕС), разработена в рамките на Европейската система за събиране и анализ на уебданни (WIN), се фокусира върху цифровото присъствие на предприятия с десет и повече заети чрез уебизвличане и анализ на данни за техните

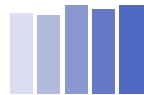
уебсайтове, присъстват ли в социалните медии, имат ли многоезични сайтове, извършват ли електронна търговия от техни сайт и подобни. Данните се събират и обработват чрез платформата Web Intelligence Hub (WIH), като резултатите са насочени към допълване на традиционните изследвания за използване на информационни и комуникационни технологии (ИКТ) от предприятията. По-долу е направен синтез на ключови аспекти на ОБЕС:

- **Дефиниция и обхват** - целевата група са предприятията с десет и повече заети лица със собствен уебсайт. Тази дефиниция е съобразена с въпроса от традиционното статистическо изследване „Използване на ИКТ в предприятия“: „Има ли вашето предприятие уебсайт?“. Обхватът включва уебсайтове, хоствани от предприятието или трети страни (например родителски компании), но изключва профили в социални мрежи или представяне на електронни пазари.
- **Процес на събиране и обработка на данни:**
 - URL адресите на предприятията се взимат от Статистическите бизнес регистри;
 - Трансформират се JSON формат и се зареждат в услугата за придобиване на данни (DAS) на WIH;
 - В WIH-DAS се конфигурира обхождаща машина Crawler с различни параметри (например дълбочина, обработка на грешки) и се стартира извличане;
 - Извлечените данни се обработват с инструменти налични в WIH, като Python, R и др.
- **Ключови индикатори и методологии** - ОБЕС се фокусира върху три основни индикатора, извлечени чрез обработка на HTML - присъствие в социалните мрежи, електронна търговия и наличие на многоезични сайтове. Дефиницията за електронна търговия е дали предприятието поддържа онлайн магазин на своя уебсайт. Дефиницията за присъствие в социалните мрежи е дали предприятието има връзки към профили в платформи за социални мрежи. Дефиницията за многоезични сайтове е дали уебсайтът предлага съдържание на няколко езика;

Предизвикателствата и ограниченията са правни и технически проблеми. Уебизвличането трябва да съответства на правни рамки. Някои предприятия блокират Crawlers и по този начин създават техническите проблеми за осъществяването на извличане на данни от уебсайтовете. Също така динамичната промяна на съдържанието на сайтовете (например страници, генерирани чрез JavaScript) усложнява извличането и обработването на HTML страниците. Възможността връзките към социални мрежи да сочат към съдържание на трети страни, а не към корпоративни профили на предприятия е сериозен риск за влошаване качеството на данните. Откриването на електронна търговия разчита на специфични тагове, които може да пропуснат персонализирани имплементации. Относно представителността - списъците с URL адреси може да не покриват напълно целевата група (например предприятия без леснодостъпни URL адреси).

Обратната връзка от потребителите изяснява дефиницията за електронна търговия и методологията за многоезична поддръжка. Нерешен остава проблемът с възможното блокиране до достъпа до някои сайтове. Изявен е интерес към табла за управление на резултатите от извличанията на данни за предприятията с показатели за техническо качество (напр. неработещи връзки, проблеми с обработването на страници) и табла за управление на резултатите от обработката на данните за електронна търговия, присъствие в социални мрежи и многоезични индикатори за уебсайтовете.

Случаи на приложение са анализа и сравнението на резултатите на ОБЕС с традиционни изследвания за ИКТ, за да се валидират тенденциите в цифровото присъствие на предприятията.



Възможни са също и регионални изследвания чрез анализ на характеристики на уебсайтове (например електронна търговия по адресна регистрация) на нива NUTS3 (на ниво област)/NUTS2 (на ниво статистически район) и идентифициране на празнини в цифровата приемственост (например липса на поддръжка на много езици в конкретни сектори).

В заключение може да се обобщи, че ОБЕС предоставя практическа информация за дигиталния отпечатък на предприятията, използвайки уебизвличане и анализи, базирани на правила. Въпреки че предизвикателства като динамичното съдържание и спазването на законовите разпоредби продължават да съществуват, резултатите от приложението ОБЕС могат да допълват традиционните статистически изследвания и да подкрепят регионалната/икономическата политика. Бъдещата работа има за цел да усъвършенства методологиите и да отговори на нуждите на потребителите за прозрачност и показатели за техническо качество.

Работен пакет: WP3: Нови случаи на използване (use-cases)

Целта на WP3 е да изследва потенциала на нови видове уебизточници на данни, които все още не са интегрирани в WIN, с акцент върху бъдещата им интеграция в WIN. За да се постигне основната цел, пакетът разглежда набор от шест нови случая на използване (use-cases), а именно: Характеристики на пазара за недвижими имоти, Проследяване на строителна активност, Онлайн цени на домакинско оборудване и аудио-визуална, фотографска и информационна техника, Експериментални индекси в статистика на туризма (цени на хотели), Подобряване на бизнес регистрите и Бързи икономически индикатори (камери за наблюдение на трафика).

Характеристики на пазара на недвижими имоти

Основните източници на данни за изпълнението на тази подзадача на проекта са порталите за недвижими имоти в различни държави. Целта е да се произвеждат детайлни, месечни индикатори за пазарите на продажби и наеми. В резултат е генерирана експериментална статистика за пазара на недвижими имоти в пет европейски държави (България, Полша, Германия, Финландия и Франция) в рамките на проекта WIN.

Партньорите оценяват източници на данни (уебсайтове) с помощта на чеклист с 30 критерия (например използване на CAPTCHA, страниране, структурирани данни и други). България и Полша разчитат на уебизвличане (Python/Scrapy, BeautifulSoup), докато Финландия и Франция получават данни чрез двустранни споразумения с притежатели на данни (Oikotie.fi, SeLogger). Германия използва смесица от извличане и API.

Относно качеството на данните могат да се изведат често срещаните проблеми, които включват динамични промени в уебсайта (например промени в HTML структурата), липсващи променливи (например номера на етажи, цени) и дубликати (например припокриващи се обяви в портали). Полша свързва адресите с териториални регистри, а Германия и Финландия сравняват уебданните с официалната статистика за валидиране. Като показатели се използват задължителните променливи (например цена, площ, стаи), които са класифицирани в категории (например ценови диапазони на м², брой стаи). Партньорите изготвят динамични редове за тенденции и териториални разпределения.

Констатации и резултати:

- Стабилност - уеб данните показват колебания, но като цяло последователни тенденции с промените на пазара (напр. офертите за наеми в Полша се увеличават след 2022 г. поради

притока на бежанци). Данните от Oikotie на Финландия са в съответствие с официалната статистика за продажбите/наемите;

- Ограничения - дублирането (напр. споделени обяви в портали), непълните метаданни и регионалното недостатъчно представителство (напр. селските райони в Германия, някои области в България) ограничават надеждността;
- Потенциал - уедданните са разглеждани като допълнение към традиционната статистика, полезни за нови индикатори или напречен анализ, въпреки предизвикателствата в стандартизацията.

В заключение проектът демонстрира жизнеспособността на уедданните за анализ на недвижимите имоти, но подчертава необходимостта от надеждна валидация, стратегии за премахване на дублирането и адаптивност към промените в източниците. Съвместните методологии и партньорства (напр. със собствениците на портали) са от решаващо значение за преодоляване на техническите и качествените пречки.

Проследяване на строителна активност

Основните източници на данни за изпълнението на тази подзадача на проекта са порталите за недвижими имоти в различни държави. Проучването има за цел да допълни официалната статистика чрез използване на уедбизвличане и API за събиране на подробни, навременни данни за новопостроени жилища. Използвани са уедданни от портали за недвижими имоти за разработване на ранни индикатори за строителни дейности в Германия (Хесен и Берлин-Бранденбург) и Швеция.

За изпълнение на целите на проекта като източници на данни са използвани следните портали: ImmobilienScout24 (Германия) и Hemnet/Booli (Швеция), които са извличани с помощта на инструменти като Python (Scrapy), R и API. Германският партньор HSL осигурява данни чрез директни споразумения с доставчици на портали, докато SSI-BBB използва хибридни методи за извличане/API. Шведският партньор SCB се фокусира върху продадените имоти, за да направи изводи за тенденциите в строителството.

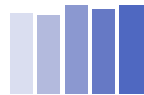
Показателите включват месечни обеми на обявите, разпределение на цените/размерите и географско агрегиране (региони по NUTS3). Германия набляга на дедубликацията и филтрирането за имоти „за първи път“; Швеция анализира сроковете за продажби, за да направи извод за завършване на строителството.

Като предизвикателства може да се посочат качеството на данните, обусловено от липсващите променливи (напр. номера на етажите, адреси) и динамичните промени в уебсайта (напр. промени в HTML структурата), които усложняват извличането на данни. Селските райони показват недостатъчно покритие, докато градските райони имат свръх представяне.

Друг важен проблем са дубликати, например обявите за един и същ имот в различни портали, повтарящите се публикации изискват дедубликация чрез съпоставяне на адреси, усложнено от непълни/непоследователни метаданни.

Временни пропуски, например при краткосрочните обяви (напр. наеми в Берлин с високо търсене), са допуснати при седмичното извличане на данни, рискувайки недостатъчен обхват.

В съответствие с официалните данни, уедданните до голяма степен отразяват тенденциите в официалната статистика за строителството, но не отчитат активността в селските райони. Извлечените от Швеция данни съвпадат с официалните обеми на продажбите на апартаменти, но подценяват



къщите, вероятно поради факта, че за самостоятелно построените къщи не се публикуват обяви в изследваните портали. Разделянето между градски и селски райони изкривява представителността. Градските райони (напр. Франкфурт, Берлин) показват по-високо покритие, докато селските региони (напр. Фогелсбергкрайс) имат оскъдни обяви.

Ранни индикатори: Данните от портала откриват пикове в строителството по-рано от официалните статистически изследвания, но нестабилността и несъответствията в дефинициите (напр. етикети „ново строителство“) ограничават надеждността.

Извличането на данни от мрежата предлага рентабилно допълнение към традиционната статистика в реално време, но изисква строга валидация. Ключовите ограничения включват:

- Техническа поддръжка - извличащите данни се нуждаят от постоянни актуализации, за да се справят с промените в уебсайта;
- Концептуални пропуски - специфичните за портала дефиниции за „ново строителство“ често не съответстват на официалните критерии;
- Пристрастие в покритието - селските/отдалечените райони и самостоятелно построените къщи са недостатъчно представени;
- Сложност на дедубликацията - справянето с несъответствията и дублиращите се обяви изисква усъвършенствани методи (напр. геокодиране, машинно обучение).

Макар и обещаващи за прогнозиране на текущите събития и градските пазари, убеданните сами по себе си не могат да заменят официалната статистика. За надежден мониторинг на строителството се препоръчват хибридни подходи - комбиниране на данни от портала с традиционни източници.

Онлайн цени на домакински уреди и аудио-визуална, фотографска техника и оборудване за обработка на информация

Източници на данни за този компонент са сайтове за електронна търговия в Швеция и България. Резултат представлява докладът за методология и резултати за онлайн цени, който подробно описва работата на SCB и NSI в рамките на проекта WIN за изследване на убизвличане за онлайн цени на домашни уреди и аудио-визуално оборудване, с фокус върху интегрирането на убеданни с административни източници, за да се подобри измерването на индекса на потребителските цени (ИПЦ).

Този компонент на проекта има за цел да оцени възможността на използването на онлайн цени за домакински уреди (напр. телевизори, лаптопи), извлечени от интернет, за допълване на традиционните данни за ИПЦ. Ранните усилия в Швеция и България включват проучване на данни, разработване на инструменти за извличане на данни и първоначално събиране на цени.

Опитът на България е извличане на цени за уреди като перални машини и телевизори от три портала, изчислявайки месечни средни стойности и експериментални индекси. Основните препятствия включват липсата на данни от скенери (ограничаващо изчисляване на претегления индекс) и несъответствия в класификацията между онлайн/физическите магазини.

Шведският партньор участва в комбинирането на източници на данни като тества интегриране на уб-извлечени данни с данни от касови регистри, за да оцени тегловните коефициенти, подобрявайки точността на средната цена. Целта е да се премине от еднакви тегла (текуща практика при уб-извлечени данни) към използване на обеми на продажби от касови регистри.

Методологията включва следните елементи:

- Източници на данни - извличане на данни от мрежата (Python/iMacros) за онлайн цени, данни от касови апарати (количества продажби) от две компании;
- Средни стойности - сравнени са три показателя:
 - Истинска средна стойност - данни от касови апарати (претеглени спрямо действителните продажби);
 - Средна стойност, извлечена от мрежата - равни тегла (текущ метод);
 - Оценена средна стойност - линейен регресионен модел, използващ данните от касовите апарати на една компания за прогнозиране на теглата за друга, приложен към данни, извлечени от мрежата.
- Ключов извод - оценената средна стойност (използвайки комбинирани данни) е по-близка до истинската средна стойност в 91.5% от случаите, намалявайки отклонението в сравнение с равното тегло.

Предизвикателствата и ограниченията са в качество на данните. Промени в структурата на уебсайта, липсващи променливи (напр. атрибути на продукта) и кратки времеви серии (седмични срещу месечни) усложняват анализа. Резултатите са ограничени до ограничен брой компании и продуктови групи. По-широката приложимост (напр. облекло) се нуждае от тестване. Седмичният анализ (поради технически ограничения) може да не отразява месечните тенденции в ИПЦ.

При оценката на теглото линейната регресия работи, но алтернативни методи (напр. експоненциални модели) или критерии за сортиране (напр. „нови на склад“) биха могли да подобрят точността. При извличане на данни от мрежата може да се пропуснат намалени цени или продукти, които не са сортирани по популярност, изкривявайки средните стойности.

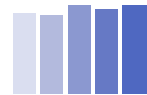
Бъдещата работа включва разширяване до повече продуктови групи/търговци на дребно, тестване на месечни данни и проучване на подобрения в изчисляването на индекса.

Извличането на данни от мрежата е обещаващо за подобряване на данните за ИПЦ, особено когато е интегрирано с административни източници. Пилотният проект в Швеция демонстрира, че комбинирането на уебданни с обеми на продажбите намалява грешките в оценката, въпреки че техническите и концептуалните предизвикателства (напр. динамични уебсайтове, съпоставяне на продукти) изискват постоянно внимание.

Експериментални индекси в статистиката на туризма (цени на хотели)

Фокусът е експериментални индекси за туризма, използвайки данни за цени на хотели, за да се допълни традиционната статистика. Източникът на данни е Booking.com, извлечени двуседмично (основни данни) и месечно (подробни данни) от 2022 - 2024 г., обхващащи имена на имоти, местоположения, цени и удобства. Участниците, GUS и NSI, имат за цел да изследват осъществимостта на интегриране на такива данни в статистиката на туризма, справяйки се с предизвикателства като динамични структури на уебсайтове и мерки срещу извличане.

Предизвикателства, като политиките против извличане на данни (блокиране на IP адреси, CAPTCHA), динамичните промени в HTML и ограниченото показване на резултати, усложняват събирането на данни. Обработката и свързването на данни включва почистване/нормализиране - суровите данни са структурирани, липсващите стойности (напр. цени на ~1 - 2%) са маркирани и променливите са



стандартизирани. Извлечените данни са съпоставени със статистически регистри, за да се коригира неправилното класифициране (напр. хотели спрямо краткосрочни места за настаняване).

По време на проекта се наблюдава сезонност в данните - най-високи средни цени на наемите през лятото (юли-август) както за Полша, така и за България, съобразени с туристическите пикове. Възстановяването след пандемията е очевидно, с покачване на цените от 2022 до 2024 година. Обхванати са над 25 000 уникални имота годишно, с незначителни липсващи данни (0.9 - 2.3% за цените) при извличане. Алгоритъмът SIFT постига 75% точност при идентифициране на дублирани имоти в различни платформи, използвайки съпоставяне на характеристики на изображения.

Получените резултати могат да се приложат при импутация: медианните цени от извлечените данни биха могли да запълнят празнини в проучванията за пътувания (напр. валидиране на отчетените от респондентите разходи за настаняване) и при пазарни тенденции: показатели като общ брой обекти по вид, брой легла/стаи и регионални ценови тенденции подпомогнаха анализа на туристическата инфраструктура.

Необходими са постоянни актуализации на скриптовете за извличане на данни поради промени в уебсайта и пропуски в данните: селските райони и по-малките имоти са недостатъчно представени в сравнение с градските центрове. Бъдещият потенциал за сътрудничество е в партньорствата с платформи (напр. достъп до API), което би подобрило надеждността на данните и сложността на извличането им.

Извличането на данни от мрежата предлага жизнеспособен и рентабилен метод за проследяване на тенденциите в туризма, особено на градските пазари. Полезността му обаче зависи от преодоляване на техническите бариери и съгласуване на извлечените данни с официалните класификации. Проектът подчертава потенциала за интегриране на нетрадиционни източници на данни в статистиката за туризма, като същевременно подчертава необходимостта от адаптивни методологии.

GUS и NSI тестват методи за импутиране на липсващи данни за цените в статистиката за туризма, използвайки извлечени от мрежата данни от Expatistan.com и Trip.com. Проучването има за цел да подобри точността на наблюденията за разходите за пътуване чрез интегриране на информация за цените в реално време.

Събирането на данни има за цел да адресира липсващи или остарели данни за цените в извадкови проучвания за туристически пътувания и разходи (напр. проучването „Участие на поляци в пътуванията“ в Полша). Използвани са месечни извлечени данни от Expatistan.com (2022 - 2024 г.) за девет продукта (мляко, хляб, бензин и др.) и пет услуги (такси, храна, настаняване и др.) в 26 полски/16 български града. Липсващите данни варират от 20 до 35% за променливи като капучино или билети за кино. Извлечени са данни от Trip.com за полети (15 997 полета) от Варшава/София до седем дестинации (Атина, Лондон и др.), като се отбелязват типовете самолети (най-често Airbus A319) и времето за резервация (цените често са по-високи, когато се резервират седмица в сравнение с месец преди заминаване).

Методологията включва „Golden Records“ - висококачествени данни от анкетата се използват като референтни стойности за валидиране и импутиране на липсващи данни. Цените, извлечени от мрежата (напр. разходи за полети, цени на храна), актуализират остарелите данни от анкетата, отразявайки пазарните условия в реално време. Наблюдавани са следните ценови тенденции:

- средните цени в България са стабилни (52 - 63 евро), докато тези в Полша се повишават рязко (79 - 145 евро) през периода;

- сезонни пикове през лятото (юли - август) за настаняване и полети;
- цените на бензина са по-нестабилни в България; тези в Полша са стабилни;
- данните, извлечени от интернет спрямо тези от анкетата, показват увеличени средни разходи с 6.35% за пътувания до България и 7.32% за вътрешни пътувания в Полша.

Честите промени в структурите на порталите изискват постоянни актуализации на скриптовете за извличане на данни. Липсващите променливи (напр. цени на жилищата, данни за полети) ограничават анализа, но са смекчени чрез валидиране „Golden Records“.

Извличането на данни от интернет предлага жизнеспособен метод за импутиране на ценови данни, подобрявайки точността на туристическата статистика, като отразява пазарната динамика в реално време. Поддържането на софтуер за извличане и справянето с пропуските в данните обаче остават ключови предизвикателства. Подходът е обещаващ за подобряване на надеждността на проучванията, особено на бързо променящите се пазари.

Подобряване на бизнес регистрите

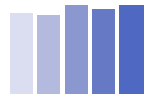
Тази подзадача е насочена към подобряване на статистическите бизнес регистри (SBR) с помощта на уебданни, като се фокусира върху две ключови задачи: намиране на URL адреси (свързване на бизнеса с техните уебсайтове) и обогатяване на SBR (подобряване на променливи като кодове на NACE и контактната информация).

Намирането на URL адреси включва свързване на бизнеси в SBR с техните уебсайтове с помощта на онлайн данни. Страните партньори използват различни подходи:

- Търсачки - Австрия, Хесен (Германия) и Швеция използват API на Google Search, извличайки резултати и прилагайки машинно обучение (напр. Random Forest), за да валидират съвпаденията. Предизвикателствата включват динамични промени в уебсайтовете, правни/етични проблеми (изтичане на информация от търсачки) и обработка на фалшиви положителни резултати;
- Данни от трети страни - Нидерландия (CBS) използва DataProvider, търговски набор от URL адреси, и подобрява свързването чрез логистична регресия, за да съпостави URL адресите с единици на SBR;
- Регистри на домейни - Финландия използва национални регистри на домейни (напр. домейни .fi), за да намери URL адреси, използвайки ping тестове и размито съвпадение, за да свърже домейни с бизнеси.

Обогатяване на SBR включва подобряване на NACE кода. NACE кодовете са от решаващо значение за SBR, но често са непълни/неправилни. Уебданни (текст от уебсайт) са използвани за:

- Откриване на грешно класифициране - ML модели (напр. Naive Bayes, SVM) идентифицират вероятно неправилни NACE кодове чрез сравняване на регистрирани кодове с функции на текста на уебсайта (напр. ключови думи, TF-IDF). Тестването показва, че 58 - 80% от грешките могат да бъдат открити с нисък брой фалшиви положителни резултати;
- Предсказване на NACE кодове - Модели (невронни мрежи, XGBoost) предсказват NACE кодове от текста на уебсайта. Правилното класифициране варира в зависимост от икономическата дейност (напр. дейността на музеите е по-лесна за предсказване от творческата дейност на художниците), но все още ръчното редактиране е по-точно от пълната автоматизация;
- Обогатяване на SBR с контактната информация - Хесен (Германия) извлича имейл адреси



и имена на мениджъри от страници, използвайки регулярни изрази за намиране на имейли (Regex отчитайки обфускация като „info[at]example.com“) и Named Entity Recognition (NER) речници за извличане на имена на мениджъри. Точността е ~78% за имейли и ~78% за имена.

Сложността на NACE представлява предизвикателство. Йерархичните NACE кодове (2 - 5 цифри) са трудни за предвиждане от текст, особено за малки предприятия или сектори с неясни описания. Качество на данните е проблем. Липсващ и непоследователен текст на уебсайта (напр. селски предприятия, сайтове, които не са на английски) ограничава точността. Анализа на инструментариума показва, че няма универсално решение - методите варират в различните страни (търсачки срещу данни от трети страни).

Бъдещи насоки са използването на хибридни подходи, включващи комбиниране на уедбанни с административни източници (напр. данъчни данни), инструменти с отворен код - разработване на инструменти за многократна употреба на URL адреси и NACE модели за прогнозиране, които да се мащабират в НСИ и извън NACE - проучване на други променливи (напр. устойчивост, патенти), използвайки уедбанни.

Уедбанните предлагат значителен потенциал за подобряване на SBR, особено за намиране на URL адреси и подобряване на NACE кода. Въпреки че напълно автоматизираното класифициране остава предизвикателство, машинното обучение (ML) и обработката на естествения език (NLP) могат да се комбинират с ръчно редактиране, подобрявайки качеството на структурирания отчет (SBR) и надеждността на официалната статистика.

Бързи икономически индикатори (камери за наблюдение на трафика)

Компонентът е насочен към разработване на по-бързи икономически индикатори, използващи данни от камери за наблюдение на трафика, с фокус върху измерването на „натовареността“ като показател за градската икономическа активност. Националната статистическа служба на Обединеното кралство (ONS) е пионер в този подход, а Шведската статистика тества методологията за приложимостта в друга страна. Основната цел е създаване на навременни, високочестотни индикатори за градска активност (напр. брой пешеходци/превозни средства), които да допълнят традиционната икономическа статистика.

Методи и технически подход:

- Източник на данни - публично достъпни камери за трафик (Великобритания: 1000+; Швеция: 25 камери в Стокхолм);
- Облачна архитектура - Google Cloud Platform (GCP) за мащабируемост, използваща Cloud Functions, Pub/Sub и BigQuery;
- Модел за дълбоко обучение - Faster-RCNN за откриване на обекти (автомобили, пешеходци, велосипедисти и др.), със статично филтриране на маска за намаляване на фалшивите положителни резултати;
- Почистване на данни - адресиране на дефектни изображения (напр. монохроматични области, повтарящи се редове);
- Статистика/Индикатори - анализ на времеви серии с импутация (пакет imputeTS) и сезонни корекции (TRAMO/SEATS) за извеждане на тенденции в натовареността.

Доказателство за концепцията в Швеция включва техническа настройка чрез локално внедряване на

предварително обучения модел на Обединеното кралство върху лаптоп (без облачна инфраструктура) и обработка на данни от седмично агрегиране на данни от камери.

Обединеното кралство успешно генерира седмични индикатори за натовареност, публикувани като експериментална статистика. Швеция - репликирана система, като наблюдавани подобни модели (напр. пикови часове през делничните дни, пикове на пешеходците през уикенда). Предизвикателствата пред Швеция включват ограниченото покритие от камерите, въздействието на времето върху осветлението (напр. сняг, тъмнина) влияещо на точността, регулаторни пречки пред използването на облачна среда, ограничения брой камери (фокус върху градските райони, малко в селските райони) и дефинирането на „натовареност“ (изисква изясняване на рамките за ограничена популация спрямо произволни процеси, без ясен консенсус).

Силните страни на подхода са ефективност - използвана е съществуващата обществена инфраструктура (камери за трафик), навременност - високочестотни (ежедневни/седмични) актуализации (по-бързи от традиционните изследвания), поверителност - избегнати са лични идентификатори.

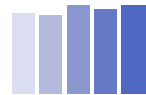
Бъдещите възможности са допълнителни данни от сензори, разработване на мащабируеми Python среди и облачни решения за производство и тестване в различни региони (напр. селски райони, различни климатични условия) за подобряване на надеждността.

Подходът с камери за движение предлага жизнеспособен, рентабилен метод за по-бързи икономически показатели. Макар че е технически осъществимо, практическите предизвикателства (инфраструктура, покритие на данните) изискват решение. Бъдещата работа трябва да се съсредоточи върху интегрирането на множество източници на данни, за да се подобри надеждността и приложимостта в цяла Европа.

Работен пакет: WP4: Методология и качество

Работният пакет 4 на WIN е посветен на методологията и качеството при интегрирането на данни от интернет в официалното статистическо производство. Целта е създаване на цялостна рамка, която да гарантира надеждност, проследимост и съпоставимост на уеб-източниците с традиционните статистически данни. В основата на тази рамка стои архитектурата BREAL (Big data REference Architecture and Layers), която предоставя модулна, мащабируема и отворена структура за събиране, съхранение и анализ на големи обеми онлайн информация. BREAL не е просто технологичен справочник, а жив организъм, който постоянно се адаптира към променящите се условия на уебпространството и към регулаторните изисквания на ЕС. Тя обхваща модули (слоеве) за стратегия и управление на процесите, които позволяват прехода от експериментални изследвания към рутинно производство без загуба на качество и без увеличаване на оперативния риск. Методологията описва минималните насоки, качество и архитектура, метриците за оценка на средата и фазите на вход и преработка, подходите за оформяне на извадки, коригиране на дубликати, подобренията в самата архитектура BREAL и не на последно място - оценките на качеството за двата водещи случая на приложение: онлайн обяви за работа (OJA) и онлайн базирани характеристики на предприятията (OBEC).

Първата фаза на методологията е позната като оценка на средата (landscaping) и представлява систематично сканиране, каталогизиране и оценяване на уебизточниците преди те да бъдат включени в статистическия процес. Този процес се разделя на три основни подетапа - каталогизиране, измерване



и избор. В рамките на каталогизирането се извършва идентификацията на URL адресите, които потенциално могат да предоставят релевантна информация, и се използват комбинации от търсачки, специализирани API и семантични техники за извличане на метаданни. Следва измерването, при което всеки идентифициран източник се оценява по редица критерии - релевантност, стабилност, историческа наличност, правна съвместимост и техническа достъпност. За целта се прилагат модели за множествени критерии, включително Analytic Hierarchy Process, особено когато става дума за трудни за оценяване портали за работа, които често променят структурата и обема на публикуваната информация. На последно място идва изборът - прилагане на прагови стойности и тегловни схеми, които гарантират, че само най-надеждните и представителни източници остават в рамките на проучването. Целият процес се повтаря периодично, за да се отразяват промените в уебпространството и да се гарантира актуалността на данните.

След като източниците са подбрани, започва процесът на уебизвличане. На този етап предизвикателствата са многобройни и често непредсказуеми. Уебсайтовете използват динамични технологии, които генерират съдържание чрез JavaScript, правейки класическото извличане чрез статични HTTP заявки неефективно. Освен това се наблюдават чести пренасочвания, промени в URL структурата, въвеждане на CAPTCHA или rate limiting механизми, които ограничават достъпа. Спазването на етичните и правните стандарти на ЕС, включително GDPR, изисква внимателно филтриране на личните данни и изключване на източници, които не предоставят прозрачни условия за ползване. Техниките на извличане се движат между два полюса - специфично извличане, което е насочено към конкретни елементи от страницата (например заглавие на обява, заплата, локация) и общо извличане, при което се сваля цялата страница и след това се прилага допълнително филтриране. Независимо от избрания подход, често се сблъскваме с проблеми като непълни данни, липсващи ключови променливи или технически грешки, които намаляват завършеността на крайния набор от данни. За справяне с тези проблеми се прилагат техники за повторно извличане, използване на прокси мрежи, ротация на потребителски агенти и автоматично откриване на структурни промени чрез алгоритми за сравнение.

След като данните са извлечени, те преминават през фазата на Sampling & Data Processing. Тук се прилагат както вероятностни, така и невероятностни методи за оформяне на извадки. Вероятностните включват стратифицирани извадки, които позволяват изчисляване на доверителни интервали и оценяване на точността. Невероятностните методи, като калибровка спрямо административни регистри или тегловни схеми, се използват за коригиране на покритието и за намаляване на предубежденията, породени от невъзможността за пълно обхващане на уебпространството. Data Wrangling включва почистване на текстове, елиминиране на дубликати, нормализиране на дати и валути, както и преобразуване на свободния текст в стандартни класификации като NACE и ISCO. Моделирането се извършва с помощта на машинно обучение - класификация на текстове за откриване на иновации, автоматично кодиране на професии и икономически дейности, коригиране на предубеждения чрез техники като re-weighting или post-stratification. Особено внимание се обръща на т.нар. „handling dynamic populations“ - откриване на concept drift и периодично преобучаване на моделите. Например когато продуктите баркодове се променят или когато терминологията в уебсайтовете се обновява, моделите трябва да се адаптират бързо, за да не се загуби точността на индикаторите.

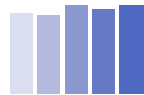
Оценката на качеството е вградена във всеки етап от процеса. За онлайн обявите за работа (OJA) ключовите проблеми са свързани със стабилността на източниците и точността на класифициране.

Платформите за работа често променят структурата си, добавят нови категории или изтриват стари обяви, което води до прекъсвания във времевите редове. Точността на автоматичното кодиране по ISCO на първо ниво варира между 56% и 62%, което е под прага за оперативно използване. Класифицирането по икономическа дейност и образование са още по-ниски - около 30% и 25%. За да се компенсира това, се използва система от показатели за качество, която включва проследяване на релевантността на източника, наличието му във времето и стабилността на версиите. За ОБЕС резултатите са по-обещаващи - свързването на URL адреси към бизнес регистри достига 82 - 90% точност, а откриването на електронна търговия и присъствие в социални мрежи варира между 60% и 90%, като има значителна вариация между държавите. Предизвикателствата при валидацията включват недостатъчен или прекомерен обхват, причинен от непълни бизнес регистри или непредставителни уебизточници, както и предубеждения, породени от обучителните множества - например свръхпредставяне на големи предприятия с уебсайтове и подпредставяне на микро- и малки фирми.

Независимо от постигнатия напредък, проектът идентифицира няколко ключови предизвикателства. Нестабилността на източниците остава водещ проблем - платформите за работа се реструктурират, променят URL схеми или изчезват напълно, което нарушава надлъжните анализи. Точността на моделите за класифициране на професии и икономически дейности остава недостатъчна за нуждите на оперативната статистика, особено когато се изисква високо ниво на агрегираност. Правните и технически бариери продължават да ограничават мащаба на извличане - от GDPR изискванията за анонимизация до динамичното съдържание, генерирано от JavaScript. Concept drift е друг фундаментален проблем - моделите губят точност с течение на времето, тъй като уебсъдържанието се променя - например терминологията за иновации се обновява, което води до грешно класифициране на текстове. За справяне с тези предизвикателства, екипът препоръчва редица мерки - добавяне на допълнителен слой „Стратегия и управление на процеса“ в BREAL, който да улесни прехода от експеримент към производство; редовни анотационни кампании за подобряване на обучителните множества; валидиране на хибридни данни, които комбинират уеб и традиционни източници; централизирани платформи като Web Intelligence Hub, които да приоритизират инструменти с отворен код и прозрачност на метаданните; и стандартизиране на критериите за landscaping и споделяне на crawlers между институциите, за да се намали дублирането на усилията.

Заклучение

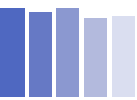
Резултатите от проекта показват, че Web Intelligence Hub вече е оперативен за тестване на OJA и ОБЕС и е интегриран с DataLab за обработка. Приложението OJA включва експериментална статистика за свободни работни места по ISCO първо ниво и NUTS2 региони, но също така разкрива ограниченията, породени от волатилността на източниците - например закриване на платформи за обяви води до прекъсване на времевите редове. При ОБЕС са постигнати висока точност при свързване на URL (90%+) и умерена точност при откриване на електронна търговия (70 - 80%). Показателите включват още наличие в социални мрежи - Facebook 85%, LinkedIn 75%, както и откриване на многоезични уебсайтове чрез анализ на домейни. Резултатите се използват за обогатяване на бизнес регистри с контактна информация и подобряване на класифицирането по NACE класификацията. Новите области на приложение включват ранни индекси за наемни пазари в Полша и Германия, табла за цени на хотели в Швеция и модели за импутация на липсващи данни от класически туристически проучвания, както и регионални метрики за новопостроени имоти в Австрия.



Натрупаните уроци показват, че данните от интернет трябва да се разглеждат като допълнение, а не като заместител на традиционните източници. Хибридните подходи са съществени, тъй като убеданните рядко отговарят самостоятелно на строгите изисквания за официална статистика. Ограниченията на ОЈА и ОБЕС подчертават необходимостта от комбиниране с административни регистри и класически проучвания. Динамичната природа на уебпространството изисква адаптивни рамки и непрекъснато преобучение на моделите. Проактивното управление на качеството, включително редовни анотации и валидация на хибридни данни, е критично за поддържане на доверието в резултатите. Регулаторните изисквания, като GDPR, налагат балансиран подход, който да гарантира както поверителността, така и актуалността на статистиката.

Проектът ESSnet WIN полага солидна основа за интегриране на убеданни в официалната европейска статистическа система. Ключовите постижения включват функционален център за изучаване на уебпространството и за мащабируемо събиране на данни, експериментална статистика за ОЈА (пазар на труда) и ОБЕС (характеристики на предприятията), методологични рамки (BREAL, насоки за качество), които да насочват бъдещата работа.

Въпреки това, предизвикателства като качеството на данните, съответствието с регулаторните изисквания и сложността на интеграцията остават. Чрез използване на ангажираността на общността (WISER), приемане на нововъзникващи технологии (LLM) и приоритизиране на хибридни методологии, ЕСС може да отключи пълния потенциал на убеданните за предоставяне на навременна, подробна и надеждна статистика.



РЕФЕРЕНЦИИ

1. **Web Intelligence Network.** Достъпно на: <https://cros.ec.europa.eu/dashboard/web-intelligence-network> (достъп: 10.07.2025).
2. **Web Intelligence Network Project Overview.** Достъпно на: <https://cros.ec.europa.eu/book-page/web-intelligence-network-project-overview> (достъп: 10.07.2025).
3. **Online Job Advertisements (OJA).** Достъпно на: <https://cros.ec.europa.eu/wih/oja> (достъп: 10.07.2025).
4. **Online based enterprise characteristics (OBEC).** Достъпно на: <https://cros.ec.europa.eu/wih/obec> (достъп: 10.07.2025).
5. **Web Intelligence Network Reports.** Достъпно на: <https://cros.ec.europa.eu/book-page/web-intelligence-network-reports> (достъп: 10.07.2025).