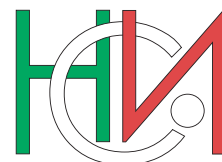


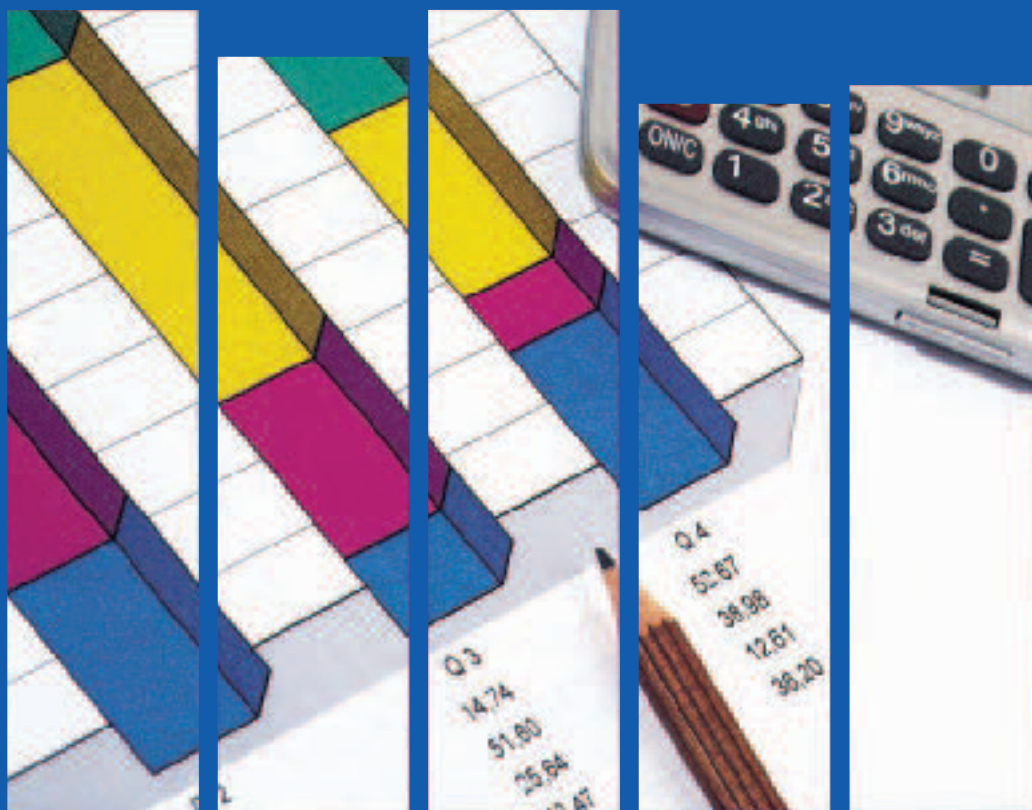
РЕПУБЛИКА БЪЛГАРИЯ  
REPUBLIC OF BULGARIA



НАЦИОНАЛЕН СТАТИСТИЧЕСКИ ИНСТИТУТ  
NATIONAL STATISTICAL INSTITUTE

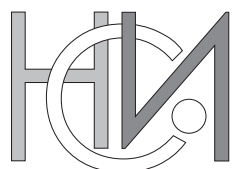
# СТАТИСТИКА STATISTICS

1/2021





**РЕПУБЛИКА БЪЛГАРИЯ**  
**REPUBLIC OF BULGARIA**



**НАЦИОНАЛЕН СТАТИСТИЧЕСКИ ИНСТИТУТ**  
**NATIONAL STATISTICAL INSTITUTE**

# **СТАТИСТИКА**

# **STATISTICS**

**1/2021**

**СОФИЯ, 2021**  
**SOFIA, 2021**

A decorative graphic at the bottom of the page consisting of three curved, overlapping lines that sweep across the width of the page. The lines are black and vary in curvature, creating a sense of movement and flow.

## РЕДАКЦИОННА КОЛЕГИЯ

Главен редактор: д-р Богдан Богданов

Заместник главен редактор: проф. д-р Васил Цанов

Членове:

Проф. д-р Йордан Христосков, проф. д.с.н. Светлана Съикова, проф. д-р Поля Ангелова,  
проф. д-р Димитър Аркадиев, доц. д-р Калоян Харалампиев, доц. д-р Любомир Иванов,  
доц. д-р Екатерина Тошева, проф. Тодор Калоянов, д-р Любен Томев,  
д-р Елка Атанасова, д-р Александър Найденов,  
Деян Славов, д-р Галя Статева

Преводач на руски език: Лариса Петкова

Преводач на английски език: Камен Караджов

Отговорен редактор: Лидия Александрова

Редактори: Мила Трифонова, Соня Златанова

Адрес на редакцията:

София, 1038, ул. „П. Волон” № 2, ет. V

e-mail: [bbogdanov@nsi.bg](mailto:bbogdanov@nsi.bg)

## EDITORIAL TEAM

Chief Editor: Dr. Bogdan Bogdanov

Deputy editor: Prof. Dr. Vasil Tsanov

Members:

Prof. Dr. Jordan Hristoskov, Prof. Dr. Sc. Svetlana Saykova,

Prof. Dr. Polyana Angelova, Prof. Dr. Dimitar Arkadiev,

Assoc. Prof. Kaloyan Haralampiev, Assoc. Prof. Lyubomir Ivanov, Assoc. Prof. Catherine Tosheva,

Prof. Todor Kaloyanov, Dr. Lyuben Tomov, Dr. Elka Atanasova, Dr. Alexander Naidenov,

Deyan Slavov, Dr. Galya Stateva

Russian Translation: Larisa Petkova

English Translation: Kamen Karadjov

Editor: Lidia Aleksandrova

Editors: Mila Trifonova, Sonya Zlatanova

Editorial address:

2, P. Volov St., Sofia 1038, Bulgaria

Vth floor

e-mail: [bbogdanov@nsi.bg](mailto:bbogdanov@nsi.bg)



## СЪДЪРЖАНИЕ

Стр.

### РАЗВИТИЕ НА СТАТИСТИЧЕСКАТА ТЕОРИЯ И ПРАКТИКА

Костадин Георгиев	Производство на експериментална статистика за характеристики	
Галя Статева	на предприятията с данни от интернет .....	9
Валерия Ангелова	Методи и концепции за функционално географско райониране	
	за статистически цели .....	55

### СТАТИСТИЧЕСКИ ИЗСЛЕДВАНИЯ И АНАЛИЗИ

Цветозария Гатева	Организационната структура на Националния статистически институт -	
	синтез между традиции и съвременност .....	91

### ИНФОРМАЦИИ, РЕЦЕНЗИИ, КОНСУЛТАЦИИ

Богдан Богданов	Математиката с поглед към статистиката .....	115
Соня Златанова	Първо преброяване на индустриалните и търговските заведения в	
	Царство България през 1926 година .....	131
Богдан Богданов	В памет на проф. Поля Ангелова .....	147

## СОДЕРЖАНИЕ

		Стр.
<b>РАЗВИТИЕ СТАТИСТИЧЕСКОЙ ТЕОРИИ И ПРАКТИКИ</b>		
Костадин Георгиев Галя Статева	Производство экспериментальной статистики для характеристики предприятий на основе интернет-данных .....	9
Валерия Ангелова	Методы и концепции функционального географического районирования в статистических целях .....	55
<b>СТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ И АНАЛИЗЫ</b>		
Цветозария Гатева	Организационная структура Национального статистического института - синтез традиций и современности .....	91
<b>ИНФОРМАЦИЯ, РЕЦЕНЗИИ, КОНСУЛЬТАЦИИ</b>		
Богдан Богданов	Математический взгляд на статистику .....	115
Соня Златанова	Первая перепись индустриальных и торговых предприятий в Царстве Болгария в 1926 годуу .....	131
Богдан Богданов	В память о профессоре Поле Ангеловой .....	147



## CONTENTS

	Page
<b>DEVELOPMENT OF THE STATISTICAL THEORY AND PRACTICE</b>	
Kostadin Georgiev, Galya Stateva	7
Valeria Angelova	7
<b>STATISTICAL SURVEYS AND ANALYSIS</b>	
Tsvetozaria Gateva	15
<b>INFORMATION, REVIEWS, CONSULTATIONS</b>	
Bogdan Bogdanov	115
Sonya Zlatanova	131
Bogdan Bogdanov	147



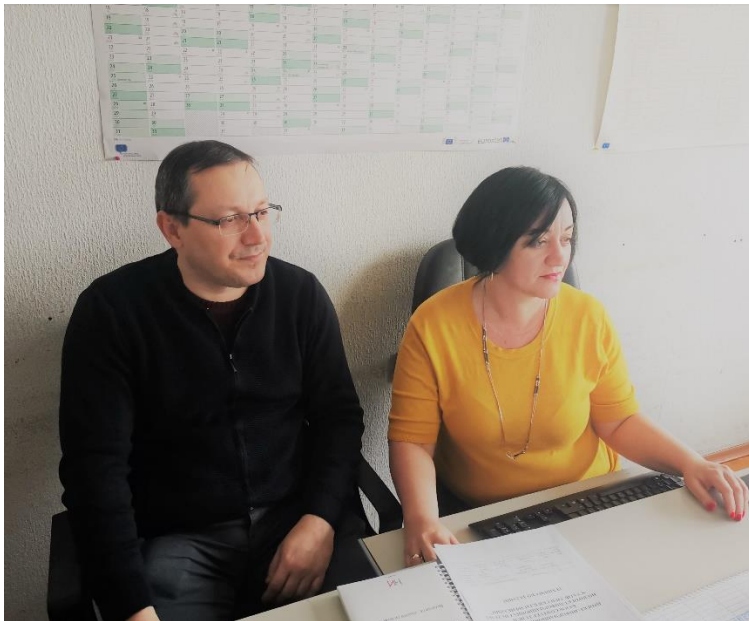


**РАЗВИТИЕ НА СТАТИСТИЧЕСКАТА  
ТЕОРИЯ И ПРАКТИКА**



## ПРОИЗВОДСТВО НА ЕКСПЕРИМЕНТАЛНА СТАТИСТИКА ЗА ХАРАКТЕРИСТИКИ НА ПРЕДПРИЯТИЯТА С ДАННИ ОТ ИНТЕРНЕТ

*Костадин Георгиев\*, Галя Статева\*\**



### I. Въведение

С всяка изминала година онлайн бизнесът става все по-важен, а с наличието на глобалната пандемия COVID е по-важен от всякога. Работният пакет С (WPC) за онлайн базирани характеристики на предприятието (OBEC<sup>1</sup>) в рамките на европейския проект ESSnet on Big data II<sup>2</sup> е свързан с разбирането на онлайн икономическата и бизнес активността на предприятията от гледна точка на националната статистика.

Производството на официална статистика за бизнес характеристиките обикновено е резултат от провеждането на класическо статистическо изследване и/или административни данни. Статистическото изследване има някои недостатъци като увеличаване на тежестта за респондентите или реализиране на високи разходи за националните статистически организации или други национални органи на статистиката. От своя страна, административните данни може да не включват всички променливи, необходими за производството на всеки статистически продукт, а наличните променливи понякога имат значително забавяне във времето. Независимо от това Статистическият

\* Главен експерт в отдел „Информационни системи и приложен софтуер“, НСИ; e-mail: [kgeorgiev@nsi.bg](mailto:kgeorgiev@nsi.bg)

\*\* Държавен експерт в дирекция „Обща методология, координация и анализ на статистическите изследвания“, НСИ; e-mail: [gstateva@nsi.bg](mailto:gstateva@nsi.bg).

<sup>1</sup> Online Based Enterprise Characteristics (OBEC)

<sup>2</sup> [https://ec.europa.eu/eurostat/cros/content/essnet-big-data-1\\_en](https://ec.europa.eu/eurostat/cros/content/essnet-big-data-1_en)

бизнес регистър (СБР) обикновено се използва като рамка за формиране на извадки, анализ на неотговорилите единици и калибриране, както и за статистически оценки. Извлечените от интернет мрежата данни от уебсайтовете на предприятията имат потенциала да „смекчат“ недостатъците в двата вече добре познати източника на данни: статистическото изследване и административните регистри. Процесът по извличане на данни от корпоративните уебсайтове включва незначителна тежест за предприятието (респондента) и използването на актуални „извлечени“ онлайн данни може да доведе до получаване на навременни характеристики на предприятието. Обаче включването на „извлечени“ данни от интернет за конкретен и понякога вече съществуващ статистически продукт определено не е лесна задача. Например може да не е възможно да се свърже еднозначно дадена уебстраница с предприятие, тъй като едно предприятие може да използва много уебстраници или една уебстраница може да се използва от много предприятия.

Ключов резултат от работата по WPC е подобряването на качеството на статистическия бизнес регистър по отношение на характеристики за онлайн присъствието на национално регистрираните фирми, като наличието на уебсайтове, електронна търговия или акаунти в социални медии. Тази проста иновация е от значение за всички членове на Европейската статистическа система (ЕСС), тъй като всяка национална статистическа институция поддържа СБР и едновременно с това е мощно средство, тъй като СБР вече са свързани с набори от данни, които са в основата на икономиката и бизнеса на всяка страна. Това означава, че интеграцията на данните от интернет мрежата в СБР е автоматична, незабавна и безпроблемна.

Основната цел на WPC да използва техники за web-scraping, извличане на знания от текст (text mining) и статистически изводи за събиране и обработка на корпоративна информация с цел подобряване или актуализиране на съществуващата информация, като присъствие в интернет мрежата, вид на икономическата дейност, информация за адресна информация, структура на собствеността и др. в националните СБР, беше постигната успешно.

В рамките на WPC методологията от предишния проект ESSnet on Big data I беше обобщена и разширена с цел използване във всяка държава от ЕСС, като се вземе предвид разнообразието, необходимо за поддържане на различните случаи на използване (use-cases) в статистическата практика. Тъй като уебскрапингът е сравнително нов метод за извличане на данни за статистическите организации, който изисква необходимото внимание по отношение на защита на данните, беше разработен и публикуван ЕСС макет на политиката за уебскрапване<sup>3</sup>, който съдържа основни правни и етични съображения, както и стабилен набор от принципи и практики, които официалните статистически организации могат да следват.

<sup>3</sup> ESS Web scraping policy template, [https://ec.europa.eu/eurostat/cros/sites/crosportal/files/WPC\\_Deliverable\\_C1\\_ESS\\_web scraping\\_policy\\_template\\_2019\\_07\\_15.pdf](https://ec.europa.eu/eurostat/cros/sites/crosportal/files/WPC_Deliverable_C1_ESS_web scraping_policy_template_2019_07_15.pdf)

Националният статистически институт имаше честта и привилегията да бъде водеща институция и активен партньор по изпълнението на дейностите в рамките на работен пакет С.

Настоящата статия има за цел да запознае читателя с постигнатите резултати относно възможностите и предизвикателствата за производство на експериментална статистика за онлайн характеристики на предприятията на европейско и национално ниво.

## II. Статистически контекст на данните за онлайн базирани характеристики на предприятията

Данните, „извлечени“ от интернет (web scraped) могат да бъдат допълнителен източник за производство на статистически продукти, като за тази цел е необходимо да бъдат предварително дефинирани основните използвани концепции и дефиниции. Акцентът е върху статистическите продукти, които са пряк резултат от жизнения цикъл на обработката на големи данни за получаване на онлайн характеристики на предприятието (дефинирани чрез т.нар. „случаи на използване“).

Основните елементи, които формират статистическите продукти, са: статистическа единица, целева съвкупност и наблюдавани променливи (на входно ниво), периодичността и статистическите показатели (на изходно ниво). Стандартизираните основни концепции са допълнителните елементи, гарантиращи, че разпространяваната експериментална статистика е хармонизирана и сравнима между страните от ЕСС.

Първоначално бяха дефинирани пет „случаи на използване“ (use case):

- 1) Списък на URLs адреси (Uniform Resource Locator);
- 2) Променливи в изследването „Използване на ИКТ в предприятията“;
- 3) Данни за откриване на възникващи корпоративни класификации;
- 4) Експериментална статистика чрез невролингвистично програмиране;
- 5) Информация за всички предприятия, продаващи чрез платформата amazon.de.

Първият и вторият „случаи на използване“ са обусловени от контекста на изследването „Използване на ИКТ в предприятията“ и могат да бъдат внедрени в регулярното статистическо производство, докато третият, четвъртият и петият „случаи на използване“ по-скоро имат проучвателен характер, което не води непременно до производство на нова статистика. За апробиране на всеки use-case бяха разработени функционални производствени прототипи, където са описани подробно прилаганите методи, използваните софтуерни решения и работната процедура за изпълнение.

За производството на експериментална статистика за онлайн характеристиките на предприятията са изпълнени само първите два „случая на използване“, а именно: *Списък на URLs адреси* и *Променливи в изследването „Използване на ИКТ в предприятията“*.

**Списъкът на URLs (use case 1)** представлява база данни, съдържаща URL адреси за всяко предприятие в целевата съвкупност, където може да има един, много или нула

URL адреси за дадено предприятие. Списъкът с URL адреси може вече да е наличен в СБР от друг източник или доставчик на административни данни. В същото време този списък може да бъде изграден и от нулата, като се търсят предприятия чрез уебтърсачки като Bing или Google. Резултатите от търсенето могат да бъдат използвани като успешни опити за отговор на следните два въпроса:

- *Има ли предприятието уебсайт?*
- *Кой URL адрес е най-вероятно да принадлежи на предприятието?*

Обикновено уебтърсенето връща няколко резултата, водещи до различни базови URL адреси за едно предприятие. Основен URL адрес - например www.xyz.com или www.abc.com, може да бъде включен в няколко резултата от търсенето. Като правило уебтърсачката (web scraper) не е необходимо да търси в други страници на даден уебсайт освен на заглавната страница, тъй като информацията може да бъде намерена в основната страница или в раздел „Контакти“.

От даден списък с URL адреси уебсъдържанието може да бъде „извлечено“ с цел да се определи дали дадено URL адрес принадлежи на конкретно предприятие. Ако извлечената информация е в съответствие със съхранената информация в СБР, тогава URL адресът е валиден; в противен случай URL адресът се заменя с URL с най-висока степен на доверителност или се премахва, ако резултатът е под някакъв предварително определен праг.

Тъй като при производството на официална статистика е ясно регламентирано, че могат да се съхраняват само релевантни данни, генерираните в процеса на скрапване данни могат да бъдат изтрети след изграждането на хранилището с URL адреси. Ако общото скрапване не е възможно поради правни проблеми, проблеми с връзката, блокиране на достъпа или др., тогава извличането на URL адреси от друг наличен набор от данни може също да бъде опция. В този случай ще са необходими прилагане на допълнителни методи и модели, както и разработване на специализиран класификатор в контекста на алгоритмите за машинно самообучение (machine learning).

**Променливи в изследването „Използване на ИКТ в предприятията“ (use case 2)** се фокусира върху извличането на информация от URL адрес на предприятие, съответстваща на променливите от традиционното статистическо изследване „Използване на ИКТ в предприятията“. Докато списъкът с URL адреси може да бъде създаден за всички предприятия от генералната съвкупност, то изследването „Използване на ИКТ в предприятията“ съдържа само предприятия с 10 или повече заети в дадени категории на NACE Rev. 2<sup>4</sup>. Извлечените онлайн базирани характеристики на предприятието трябва да отговарят на следните въпроси от изследването „Използване на ИКТ в предприятията“:

- *Уебсайтът на предприятието съдържа ли нещо от изброените функционалности: онлайн поръчки, резервации или електронен магазин (например уебсайтът има ли пазарска количка)?*

<sup>4</sup> Statistical classification of economic activities in the European Community (NACE).

- *Използва ли предприятието социални медии?*
- *Използва ли предприятието Twitter за конкретна цел:*
  - *създаване на имидж или реклама на пазарни продукти на предприятието;*
  - *набиране на персонал;*
  - *други - всеки туит, който не се вписва в другите две категории.*
- *Има ли предприятието специфични характеристики на уебсайта като:*
  - *описание на стоки или услуги, ценоразписи;*
  - *възможност за посетителите да персонализират или проектират онлайн стоки или услуги;*
  - *проследяване или състояние на направени поръчки;*
  - *персонализирано съдържание в уебсайта за редовни/повтарящи се посетители;*
  - *връзки или препратки към профилите на предприятието в социалните медии;*
  - *обяви за свободни работни места или онлайн кандидатстване за работа.*
- *Работи ли предприятието върху предстоящи/нови явления, по-специално свързани с изкуствения интелект (AI) и машинното самообучение (ML).*

Последният въпрос не е част от въпросника на ИКТ изследването, но е включен като допълнителна характеристика, за да се идентифицират иновативните предприятия, които използват AI и ML в текущата си дейност.

## 1. Основни понятия

Основното значение при използването на скрапнати от уебмрежата данни като допълнителен източник за статистическа продукция се крие в процеса на извличане на онлайн базирани характеристики на предприятието (ОВЕС). Както подсказва името, ОВЕС може да се разглежда като всяка характеристика на предприятието, която е извлечена от корпоративните уебстраници с помощта на методите за скрапване. Например URL адресът на предприятието, ако съществува.

Извлечената информация може да бъде директно достъпна чрез друг доставчик на данни или да бъде събрана от самата статистическа организация. Процесът на събиране на данни изисква прилагане на софтуер за скрапване, който директно извлича съдържанието на URL или изобразява уебстраница чрез браузър и тогава извлича изобразеното съдържание. В зависимост от вида на ОВЕС различни части на уебстраницата могат да бъдат подходящи за директно скрапване на интересуващото ни съдържание (ако е идентифицируемо), което би предотвратило натрупване на ненужно големи масиви от данни. Когато скрапнатите данни са вече събрани, извличането на ОВЕС включва или използване на детерминистичен подход, или прилагане на статистическо или машинно самообучение. Както детерминистичният, така и самообучителният подход трябва да бъдат щателно тествани и да са достигнали определено ниво на качество.

Като обобщение, процедурата за извличане на онлайн характеристики на предприятията може да бъде разделена на две основни части:

- Процедура по скрапване: извличане на информация за предприятия чрез скрапване на интернет мрежата - например URL адрес на предприятието (Uniform Resource Locator);
- Процес на извличане: извличане на ОБЕС с помощта на скрапваните данни от стъпка 1.

Извлечените онлайн характеристики на предприятията, ако процесът е бил успешен, след това могат да бъдат свързани със съществуващи единици (предприятия) в СБР и да се използват в статистическия бизнес процес или директно, или като допълнителен източник на данни.

## 2. Вид единица

Статистическият бизнес регистър е единственият източник за генериране на статистически единици за производство на официална бизнес статистика. В резултат на това се постига съгласувана икономическа статистика, съпоставима между секторите, страните, географските области и във времето. Информация от мрежата може да бъде получена за два вида статистически единици от СБР: за предприятия и групи предприятия.

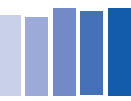
*Предприятието* е най-малката комбинация от юридически единици, която е организационна единица, произвеждаща стоки или услуги, с известна степен на автономност при вземането на решения, особено за разпределение на текущите си ресурси. Всяко предприятие може да се състои от една или повече правни единици или да съдържа само част от правна единица. През последното десетилетие правната организация на предприятията се усложни, тъй като предприятията все повече следват тенденцията да диверсифицират своите дейности на множество правни единици.

Във фазата на идентификация на единиците основно се използва данъчният/ДДС код за потвърждаване на кореспондиращата връзка уебсайтове-предприятия. Трябва да се обърне внимание на случаите, когато в едно и също предприятие има повече от една юридическа единица, тъй като всяка от тях ще използва свой собствен данъчен код.

*Група предприятия* е набор от правни единици, правно и/или финансово обвързани, но с единно управление и контрол, т.е. управление, което определя общата корпоративна политика. Всяко предприятие в групата може да има свой собствен уебсайт, но е по-вероятно за групата предприятия да има единен уебсайт за цялата група, където цялата необходима информация да се намира по-лесно (заедно с връзките към уебсайтове на компаниите в групата).

Въпреки че основното предназначение на ОБЕС е да замести част от въпросите на изследването „Използване на ИКТ в предприятията“, единицата, за която се отнася ОБЕС, невинаги може да бъде самото предприятие. В такива случаи уебсайтът на предприятието може да се разглежда като респондент, за да се изведе статистическа оценка, отнасяща се до URL адресите на предприятията в определена целева съвкупност. Като се има предвид традиционното ИКТ изследване, някои променливи на наблюдение се отнасят не само до уебсайта на самото предприятие, но и до уебсайта на компанията майка или холдинг. По този начин един URL адрес може да доведе до ОБЕС за





множество предприятия. Накратко казано, единиците за извличане на онлайн характеристики са **предприятия и/или уебстраници**. За целите на коректните сравнения предприятието като единица за извличане на ОВЕС е в съответствие с Регламент 7 (ЕИО) № 696/93 на Съвета от 15 март 1993 г. относно статистическите единици за наблюдение и анализ на производствената система в Общността.

*Уебсайтът на предприятието* като единица за извличане на ОВЕС се определя като някакво цифрово решение (в World Wide Web), което едно предприятие има или предлага на своите клиенти.

### 3. Съвкупност и извадка от целевата съвкупност

Целевата съвкупност за официална бизнес статистика обикновено са всички единици в СБР или подсъвкупност от тях, както е случаят с изследването „Използване на ИКТ в предприятията“. По-специално, когато се изчисляват статистически показатели от ОВЕС, може да се извлече оценка въз основа на подгрупата от съвкупността, съставена от предприятия, които имат един или повече URL адреси.

Първата задача е да се отчитат правилно събраните данни от интернет за единиците на съвкупността, т.е. данните, събрани от уебсайтове, трябва да се отнасят към единицата „предприятие“. Наличието на СБР, рамката на съвкупността, съдържаща всички предприятия, включени в целевата съвкупност, предполага използването на СБР като основа за търсене на съответните уебсайтове.

В зависимост от целта на изследването е по-подходящо производството на експериментална ОВЕС статистика да се ограничи до част от съвкупността, която представлява конкретна подгрупа от бизнес статистиката, вместо да се отнася до всички единици в СБР. В зависимост от вида на статистическия показател ОВЕС могат да бъдат извлечени за цялата целева съвкупност или само за извадка от нея.

След като се определи целевата съвкупност, от нея се конструира рамката на извадката. Наблюдава се, че разликата в структурата и съдържанието на уебсайтовете зависи до голяма степен от размера на предприятията, сложността на тяхната организация и икономическите дейности, които те извършват. Следователно, за да се оценят правилно резултатите от дейността по скрапване от интернет мрежата, се препоръчва да се направи стратифицирана извадка от СБР. Извличането на ОВЕС за тази извадка от целевата съвкупност може да бъде от полза по различни причини. Разходите по отношение на изчислителното време може да са значително по-ниски и би било възможно ръчно да се провери дали алгоритъмът за скрапване е дал коректни резултати. В зависимост от статистическия показател статистическата оценка на базата на извадката вече има потенциал да отговори на задължителните изисквания за качество.

Като правило се избират активни предприятия през референтната година, които имат юридическа форма на корпорация или партньорство. За да се подобри точността, стратифицирана извадка от тези единици в СБР може да бъде формирана от референтната съвкупност по няколко възможни начина, например чрез използване на пропорционално разпределение. По този начин броят на единиците във всяка страта е пропорционален на броя на единиците в общата съвкупност. Поради тази причина е

препоръчително да се работи с извадка от предприятия, стратифицирани както по размер по отношение на заетостта и оборота, така и по икономическа активност, за да се вземат предвид разликите в структурата и съдържанието на уебсайтовете, евентуално причинени от тези фактори.

Извадките на ОБЕС включва предприятия с потенциално свързани URL адреси, които имат процент на оценяване над даден праг (оценяване, дадено от онлайн търсачката за всяко предприятие). Това може да зависи още от потенциалните възможности на основния производствен процес - например ако технически е възможно да се скрапват много голям брой уебстраници по време на периода на наблюдение. Статистическият показател може също да бъде оценен с помощта на ОБЕС от извадката на целевата съвкупност и съответно да се калибрират теглата на извадката.

Подходът, основан на регистър към големите данни, категорично означава, че цялата статистическа, административна и уебинформация ще бъде „каталогизирана“ в СБР за многократно използване, като по този начин се гарантира последователност и ще даде конкретна подкрепа на статистическото производство въз основа на СБР.

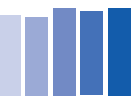
В този смисъл има „двупосочен“ информационен поток към и от регистъра. От една страна, регистърът придобива нова информация по-рано от интернет мрежата и я каталогизира по последователен начин за множество цели, увеличавайки съдържанието и способността си да подпомага статистическото производство. От друга страна, данните от мрежата, попаднали в обхвата на регистъра, се интегрират с всички останали променливи от административни и статистически източници. По този начин неструктурираните данни от мрежата придобиват „структура“ и биха могли да се възползват от цялата налична статистическа класификация: скрапнатите данни получават име, размерност, местоположение и т.н., използвайки всички класификации в статистическия Бизнес регистър. Широкият обхват на СБР се превръща в поддържаща платформа за големи данни.

#### 4. Периодичност

Когато ОБЕС се използват като алтернативен източник на данни за официална статистика, времевата рамка за извличане на ОБЕС трябва да бъде в съответствие с периода на наблюдение, както е дефиниран в методологичното ръководство на изследването „Използване на ИКТ в предприятията“. Периодичността обаче може да бъде увеличена, ако скрапването не натоварва много посетените URL адреси.

#### 5. Променливи на наблюдение

За разлика от класическото изследване за използване на ИКТ в предприятията променливите не се наблюдават чрез събиране на отговори от въпросник, а чрез използване на интернет търсачки, приложно-програмен интерфейс (API) и/или софтуер за скрапване на данни, които потенциално съдържат информация за наблюдаваната променлива. От скрапнатите данни ОБЕС могат да бъдат извлечени или с помощта на алгоритми за машинно самообучение или чрез детерминистичен подход и от една или повече ОБЕС наблюдения. Например променливата „Има ли предприятието уебсайт“ е двумерна променлива със стойности 0/1 или „да“/„не“, която се извлича чрез използване



на списъкът с URL адреси, състоящ се от един, много или нула URL адреси за дадено предприятие.

Други потенциални ОБЕС производни, целеви променливи (да бъдат извлечени от сурови текстови данни, скрапнати от уебстраници) могат да бъдат следните:

- *Какво продават предприятията: основни продадени продукти/услуги;*
- *Как предприятията продават: канали за продажба (например карта за онлайн пазаруване, услуги за резервация, услуги за доставка);*
- *На кого продават предприятията - например бизнес към бизнес, бизнес към потребител;*
- *Къде продават предприятията: национални/мултинационални пазари.*

## 6. Статистически показатели

За производство на експериментална статистика (за use case 1 и use case 2) бяха генерирани следните статистически показатели:

- Процент на предприятията, които имат уебсайтове;
- Процент на предприятията, занимаващи се с електронна търговия чрез своя уебсайт;
- Процент на предприятията, които присъстват в социалните медии;
- Процент на предприятията, използващи Twitter за конкретна цел;
- Процент на предприятията със специфични характеристики на уебсайта;
- Процент на предприятията, работещи по предстоящи/нови явления, по-специално AI и ML.

Всички изброени по-горе статистически показатели могат да бъдат директно оценени от съответните онлайн характеристики на предприятията.

В случая, когато ОБЕС е извлечена от извадка  $e_1, \dots, e_n$  от от цялата целева съвкупност и когато всяко предприятие  $e_i, i=1, \dots, n$  има кореспондиращо тегло в извадката  $w_i$ .

Единственият показател за *use case 1* е процентът на предприятията, които имат уебсайт:

$$R_{web} = \frac{\sum_{i=1}^n w_i I[u_i \neq \emptyset]}{\sum_{i=1}^n w_i} \quad (1)$$

където  $\sum_{i=1}^n w_i = N$  е броят на единиците в целевата съвкупност,  $I[.]$  функция на показателя и  $u_i$  е набор от всички уебсайтове за единица  $i$ , който е празен, ако няма намерен поне един уебсайт. Този показател може лесно да бъде изчислен по NUTS<sup>5</sup> региони, NACE категории или групиран по броя на заетите в предприятието. Нека  $A_1, \dots, A_p$  бъдат части от цялата целева съвкупност  $\mathcal{A}$ , т.е.:

$$\bigcup_{i=1}^p A_i = \mathcal{A} \text{ and } A_i \cap A_j = \emptyset \quad \forall i \neq j \quad (2)$$

<sup>5</sup> Номенклатура на териториалните единици за статистически цели (NUTS).

тогава процентът на предприятията, имащи уебсайт за специфична група,  $A_k$  е дефинирана чрез:

$$R_{web,k} = \frac{\sum_{i=1}^n w_i I[u_i \neq \emptyset \wedge e_i \in A_k]}{\sum_{i=1}^n w_i I[e_i \in A_k]} . \quad (3)$$

По подобен начин могат да бъдат дефинирани показателите за **use case 2**, но целевата съвкупност се променя и вече съдържа предприятия, имащи уебсайт. Повечето от показателите в use case 2 могат да бъдат дефинирани чрез

$$R_{ICT,q} = \frac{\sum_{i=1}^n w_i I[u_i \neq \emptyset \wedge V_i^q = 1]}{\sum_{i=1}^n w_i I[u_i \neq \emptyset]} . \quad (4)$$

с  $V_i^q$  равно на 1, където ОБЕС за дадено предприятие  $e_i$  отговаря на въпрос  $q$ .  $q$  може да бъде всеки от следните въпроси:

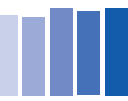
- *Уебсайтът на предприятието съдържа ли нещо от изброените: онлайн поръчки, резервации или електронен магазин (например уебсайтът има ли пазарска количка)?*
- *Има ли уебсайтът на предприятието връзки или препратки към профили на предприятието в социалните медии?*
- *Има ли предприятието специфични функции на уебсайта?*
- *Работи ли предприятието върху предстоящи/нови явления - например: AI и ML?*

За показателя *Използва ли предприятието Twitter за конкретна цел* е необходимо целевата съвкупност да бъде адаптирана, така че да съдържа само предприятия, за които онлайн характеристиката за използване на социални медии е равно на 1.

Може да се случи така, че за определени предприятия  $e_{(1)}$ , ...,  $e_{(k)}$  не се скрапват достатъчно данни, така че съответните ОБЕС да не отговарят на нито един от горните въпроси. За да се елиминират техническите причини, се препоръчва използването на софтуер за скрапване, който може да се справи с вградения JavaScript, и уебстраниците да се „идвличата“ (скрапват) по няколко пъти и по различно време. Дори и да се елиминират техническите проблеми необходимата информация все още може да не е скрапната, тъй като структурата на уебстраниците не е стандартизирана и по този начин процедурата по извличане на ОБЕС се проваля.

### **III. Жизнен цикъл за обработка на големи данни: онлайн характеристики на предприятията**

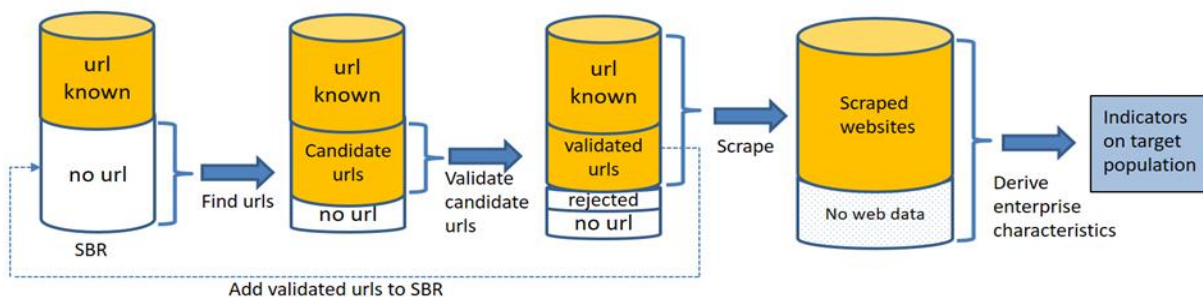
Процесът на уебскрапване за откриване на характеристики на предприятието се вписва естествено в по-общия процес на жизнения цикъл на обработка на големи данни. Този обобщен процес осигурява отправна точка за всяка държава - членка на ЕСС, която би искала да извлече онлайн базирани характеристики на предприятията на национално



ниво. За да улесни приложимостта, в която и да е държава от ЕСС, екипът на WPC схематично представи процеса за определяне на онлайн характеристики на предприятията по възможно най-стандартизиран начин (вж. фиг. 1), но при следните две допускания:

- Всяка национална статистическа служба има статистически бизнес регистър (СБР), който съдържа поне имената и идентификационните номера на фирмите. Прието е, че процентът на предприятията с известен URL адрес може да варира между 0 и 100%.
- Всяка държава има някаква национална или международна търсачка (и), която може да се използва във фазата на извличане на URL адреси.

**Фиг. 1. Обща схема за процеса на скрапване на онлайн характеристики на предприятията**



Основните **фази на процеса** на определяне на онлайн характеристики на предприятието са, както следва:

1. Извличане на URL адреси за фирми без налични URL адреси. Резултатите от търсенията са **кандидат-URL адреси** за включване в списъка с URL. На схемата ясно се вижда, че дори след няколко търсения, че все още има предприятия, за които не може да бъде намерен URL адрес. Възможните причини за това могат да бъдат: че предприятието наистина няма уебсайт или че прилаганата стратегия за търсене не е достатъчно ефективна и не може да го намери.

2. **Валидиране или отхвърляне** на кандидат-URL адреси чрез сравняване на познатите за предприятието данни в СБР с данни от резултатите от търсенето. Извличането и валидирането/отхвърлянето на URL адреси може да бъде итеративно.

3. Във фазата на **скрапване** предприятията, за които URL адресът вече е известен в СБР и потвърдените, намерени URL адреси, се използват за скрапване на уеб-съдържание и съхраняване на изтритите уебсайтове.

4. **Извличане на онлайн характеристики на предприятието** за целевата съвкупност: това може да бъде целият СБР или част от него, например съвкупността на изследването „Използване на ИКТ предприятията“. На този етап е важно да се отбележи, че предприятията, за които няма уебданни, също трябва да бъдат отчетени и оценени чрез калибриране, използвайки общия брой единици в СБР.

5. **Актуализиране на СБР** с валидираните URL адреси, получени от последователни итерации и други статистически процеси в допълнение към основния цикъл. Това е показано на фиг. 1 като стрелка от валидираните URL адреси към СБР. Тук би могъл да се включи и допълнителен индикатор дали даден URL адрес в СБР произхожда от административен източник, или от процеса на извличане от уебмрежата.

Най-важните решения за различните фази на процеса могат да бъдат изразени и чрез фазите на GSBPM<sup>6</sup> модела по отношение на жизнения цикъл на големите данни: Събиране, Обработване, Анализ и Разпространение.

Процесът на **събиране** на ОБЕС данни (*фаза 4. Събиране, GSBPM*) е съставен от четири подпроцеса. Първо се изисква идентифициране на списък на компаниите, за които ще се събират данни (целева съвкупност), с основни атрибути като например името на предприятието. След това се конструира списък с потенциални адреси на уебсайтове, като се използват отговорите от търсачката на уебсайтове за всяко предприятие в СБР или други административни регистри. В подпроцес 3 се извършва частично „обхождане“ на потенциални адреси на уебсайтове (начални страници) чрез прилагане на механизъм за класиране на вероятности кой уебсайт е възможно най-добрият избор (като по този начин се търсят идентификационни данни на уебсайта). На последната стъпка се избира „първият най-добър“ уебсайт за всяко предприятие, след което може да се прави разширено събиране/скрапване на данни от намерените уебсайтове за получаване на специфични онлайн характеристики на предприятията.

В процеса на **обработка** (*извличане, почистване, интегриране, агрегиране и представяне, фаза 5. Обработване, GSBPM*) събраните уебданни първо се почистват технически (премахване на html тагове) и съдържателно (премахване на „стоп“ думи и т.н.). Вземат се решения за това кои части от уебстраниците се запазват като входни данни и кои части повече не са ни необходими. Текстовите части, които могат да се използват като бизнес идентификатори, представляват особен интерес. Необходими са и експертни решения какви методи да са прилагат с цел трансформиране на суровите текстове в структурирано съдържание. Тези методи могат да варират от базовия подход „кошница от думи“ (bag-of-words) до методи, запазващи контекстуална информация, като doc2vec, word2vec, sentence2vec или предварително обучени мрежи за текстови анализ. Процесът на преобразуване на текст в променливи се извършва чрез прилагане на детерминистичен подход и методи за машинно самообучение за извличане на характеристики.

**Анализирането** на уебданни (*фаза 6. Анализ, GSBPM*) в комбинация с вече съществуващите данни за предприятията в СБР е от решаващо значение в много от етапите на процеса: валидирането/отхвърлянето на кандидат-URL адреси се нуждае от внимателен анализ. Уебсайт на предприятие и предприятие са две различни концепции, и поради тази причина се изисква предварително двете понятия да се свържат на концептуално ниво. Един уебсайт може да съдържа информация за множество

<sup>6</sup> Generic Statistical Business Process Model (GSBPM), <https://statswiki.unecce.org/display/GSBPM/Generic+Statistical+Business+Process+Model>



предприятия или обратно, едно предприятие може да притежава или публикува на множество уебсайтове.

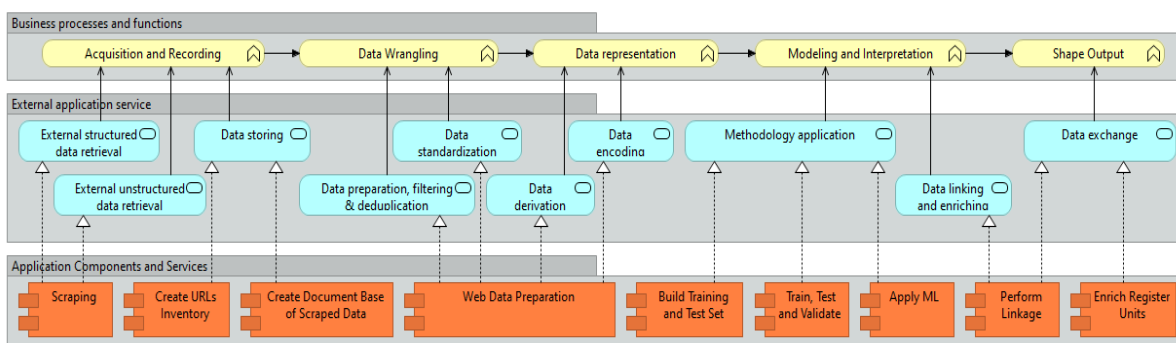
Изчисляването на ОБЕС за определена съвкупност въз основа на скрапнати уебданни се нуждае от моделиране и интерпретация. Изискват се решения относно дефинициите на характеристиките на предприятията - например на микро ниво - дефиниране на критерии кога дадено предприятие се счита, че извършва електронна търговия чрез своя уебсайт или е представено в някоя от социалните медии. Освен това в тази фаза, трябва да се има предвид, че почти винаги има подмножество на СБР, за което не може да бъде намерен URL адрес. Коригирането на това на макрониво може да се извърши чрез свързване на измерваната съвкупност към целевата съвкупност.

Разпространението на ОБЕС не се различава съществено от традиционните процеси на разпространение на статистика с изключение на това, че техниките за разпространение на уебданни са сравнително нови и би било добре да се добави разширено обяснение на използваните методи, което да улесни потребителите на експерименталната статистика за ОБЕС.

#### IV. Референтна ИТ архитектура за ОБЕС данни

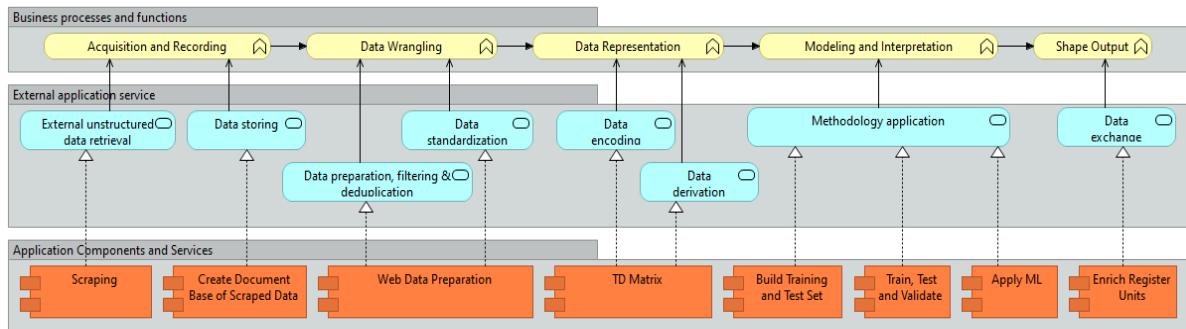
Референтната архитектура за ОБЕС данни се състои от описание на приложната и информационната архитектура на процеса на производство на експериментална статистика за случаите на използване 1 и 2. Описанието се основава на BREAL<sup>7</sup> архитектурата, резултат от работата по работен пакет F, проект ESSnet on BD II (вж. фиг. 2 и 3).

**Фиг. 2. Общо графично представяне на приложната архитектура на Списък на URLs (use case 1)**



<sup>7</sup> Big Data REference Architecture and Layers (BREAL), [https://ec.europa.eu/eurostat/cros/sites/crosportal/files/WPF\\_Deliverable\\_F1\\_BREAL\\_Big\\_Data\\_REference\\_Architecture\\_and\\_Layers\\_v.03012020.pdf](https://ec.europa.eu/eurostat/cros/sites/crosportal/files/WPF_Deliverable_F1_BREAL_Big_Data_REference_Architecture_and_Layers_v.03012020.pdf)

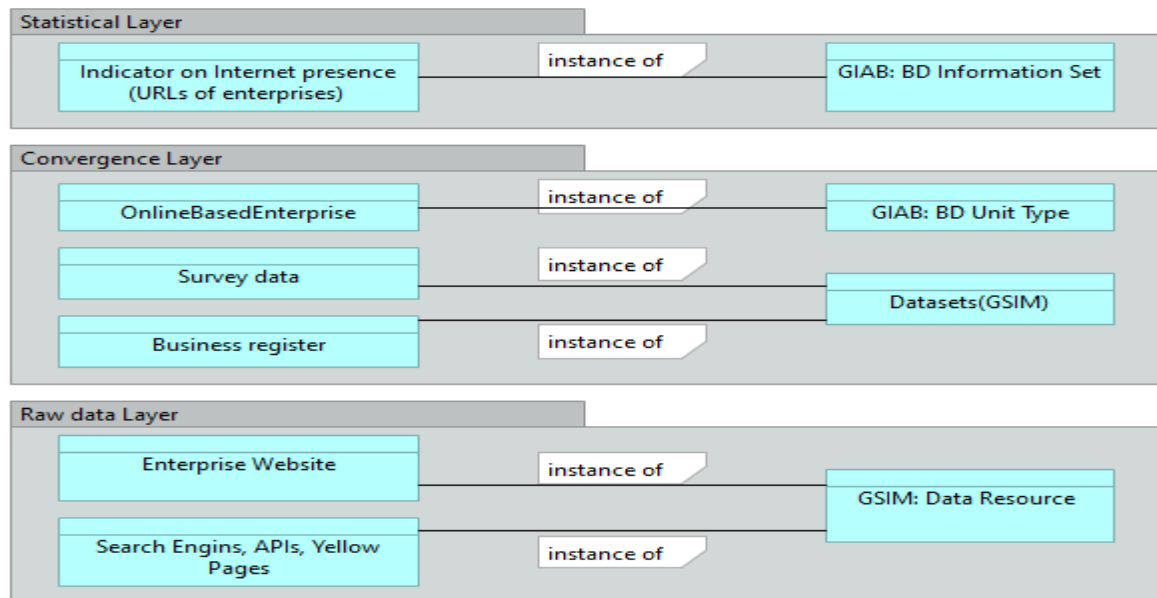
**Фиг. 3. Общо графично представяне на приложната архитектура на Променливи в изследването „Използване на ИКТ в предприятията“ (use case 2)**



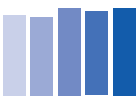
Бизнес процесите и функциите описват ОБЕС бизнес функциите, производни от модела BREAL. Всяка бизнес функция задейства следващата, като всяка бизнес функция се обслужва от една или повече услуги. Всяка услуга за приложения се реализира от един или повече компоненти на приложението. Компонентите на приложението могат да бъдат програми, модули, скриптове, класове или функции. Те могат да бъдат самостоятелни или част от система и да бъдат написани на различни програмни езици.

Описанието на общата информационна архитектура за ОБЕС за случаите на използване 1 и 2 (фиг. 4 и 5, респективно) също се основава на BREAL.

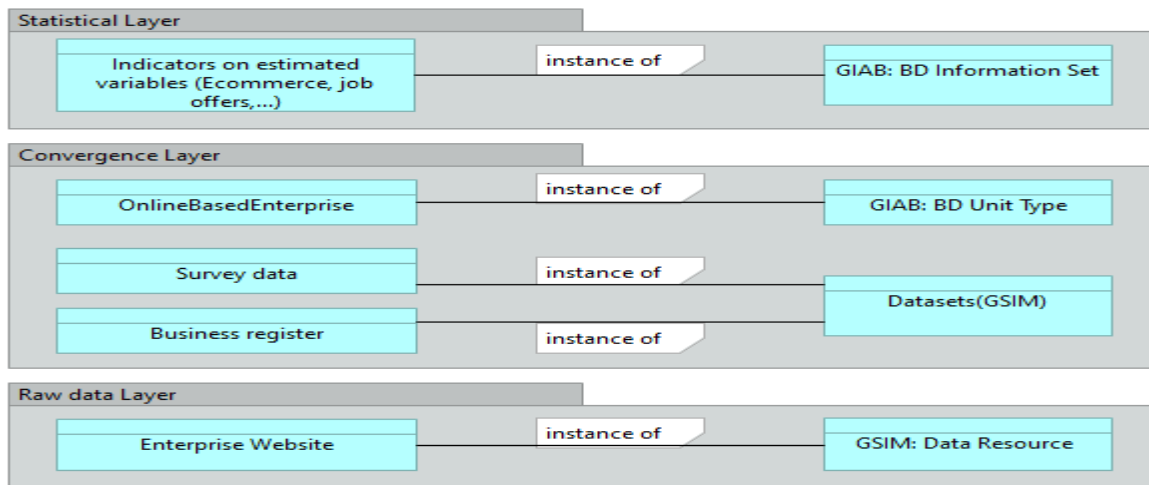
**Фиг. 4. Информационна архитектура на Списък на URLs (use case 1)**







**Фиг. 5. Информационна архитектура на Променливи в изследването „Използване на ИКТ в предприятията“ (use case 2)**



Нивото за „суровите“ данни описва първоначалните източници на данни на ОБЕС по отношение на BREAL. То обхваща процеса на придобиване и запис на ОБЕС. Междинното архитектурно ниво описва информационните обекти, получени от изходни данни чрез процесите за пренасочване и представяне на данни на ОБЕС. Статистическото ниво описва информационни обекти, получени от информационните източници на междинното ниво чрез процесите за моделиране, интерпретиране и извеждане на изходи за ОБЕС.

Почти всички приложни услуги и за двата случая на използване на ОБЕС са проектирани да бъдат достатъчно гъвкави, за да могат да бъдат оперативно съвместими, репликирани или споделяни между националните статистически организации в рамките на ЕСС.

По време на проекта всички ИТ услуги и данни се управляваха на местно ниво, но в бъдеще всички услуги (с изключение на услугите за „Свързване и обогатяване на данни“ и „Обмен на данни“, използващи локални данни) могат да бъдат споделени сред националните статистически организации в ЕСС.

В рамките на проектните дейности са внедрени някои обобщени софтуерни решения, които са публично достъпни чрез специализирания WPC GitHub на следния адрес: <https://github.com/EnterpriseCharacteristicsESSnetBigData>.

## **V. Производство на експериментална статистика за ОБЕС в Националния статистически институт**

Експертният екип от НСИ, работещ по дейностите на работен пакет С, произведеха и публикуваха експериментална статистика за ОБЕС за 2019 и 2020 година. Всички получени резултати в табличен вид са достъпни на CROS портала на Евростат (рубрика „Експериментална статистика“): [https://ec.europa.eu/eurostat/cros/content/WPC\\_Experimental\\_statistics\\_en](https://ec.europa.eu/eurostat/cros/content/WPC_Experimental_statistics_en), заедно с кратки методологични бележки и справочни метаданни (отчети за качество във формат ESQRS

и метаданни за потребителите във формат ESMS). Методологичните бележки описват процесите, използваните методи и софтуерни решения за производството на експериментални данни за ОБЕС.

Изчислените експериментални показатели са получени на базата на описаните методи, понятия и допускания в първата част на настоящата статия.

В изложението, което следва, са илюстрирани основните процесни стъпки, необходими за изпълнение на use-case 1 и use-case 2 с цел достигане до експериментални резултати за онлайн характеристики на предприятията.

#### 1. Актуализиране на URLs адреси на предприятия (use-case 1)

За извличането на онлайн характеристики на предприятия е необходимо първоначално да се намерят техните URL адреси и да се конструира списък. Целта на това упражнение се състои от следните основни стъпки:

#### **Подготовка на начални данни и софтуер**

Процесът на актуализиране на интернет адресите на предприятия започва с дефиниране на съвкупността им. Съвкупността се състои от предприятия с 10 и повече заети, които към пролетта на 2020 г. са 28 251 на брой. За тях е взета информация от статистическия Бизнес регистър<sup>8</sup> в НСИ със следните полета: ЕИК<sup>9</sup>, наименование, интернет адрес, електронна поща, пощенски адрес, стационарен телефон, мобилен телефон, населено място, код по NUTS 3<sup>10</sup> и код по икономическа дейност<sup>11</sup>.

След определяне на съвкупността се подготвят операционната среда и софтуерът за обработка. Използван е обикновен компютър с Windows операционна система и език за програмиране Python<sup>12</sup>. Избраният софтуер за обработка е URLs Finder<sup>13</sup>, част от Starter Kit<sup>14</sup>, разработен на Python. Използвана е версия 1.0 на Starter Kit (специализирано средство за статистически експерти и програмисти), като софтуерът е доработен, а доработките по-късно са включени във версия Starter Kit 2.0.

За да работи URLs Finder се нуждае от следните модули и компоненти на Python:

- pandas<sup>15</sup> - бърз, мощен, гъвкав и лесен за използване инструмент с отворен код за анализ и манипулация на данни;

<sup>8</sup> <https://www.nsi.bg/node/13207/>

<sup>9</sup> Единен идентификационен код от Регистър Булстат (<https://www.registryagency.bg/bg/registri/registar-bulstat/>)

<sup>10</sup> Номенклатура на териториалните единици за статистика - малки региони (<https://ec.europa.eu/eurostat/web/nuts/background>)

<sup>11</sup> NACE Rev. 2 (<https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF>, [https://www.nsi.bg/sites/default/files/files/pages/uplf/Methodology\\_KID.pdf](https://www.nsi.bg/sites/default/files/files/pages/uplf/Methodology_KID.pdf))

<sup>12</sup> <https://www.python.org/>

<sup>13</sup> <https://github.com/EnterpriseCharacteristicsESSnetBigData/StarterKit/tree/master/URLsFinder>

<sup>14</sup> <https://github.com/EnterpriseCharacteristicsESSnetBigData/StarterKit>

<sup>15</sup> <https://www.python.org/>

- glob<sup>16</sup> - разширение за файлова пътека в стил Unix<sup>17</sup>;
- BeautifulSoup<sup>18</sup> - библиотека, която улеснява изрязването на информация от уебстраници;
- requests<sup>19</sup> - позволява да се изпращат HTTP<sup>20</sup>/1.1 заявки изключително лесно;
- re<sup>21</sup> - предоставя операции за регулярни изрази върху текст;
- numpy<sup>22</sup> - основният пакет за научни изчисления с Python;
- time<sup>23</sup> - предоставя различни функции, свързани с времето;
- unquote<sup>24</sup> - заменя %xx кодирания с еквивалента им от един символ;
- urlparse<sup>25</sup> - разделя URL<sup>26</sup> на шест компонента в съответствие на общата структура на URL със следните имена scheme://netloc/path;parameters?query#fragment;
- tqdm<sup>27</sup> - предоставя лента за напредъка;
- datetime<sup>28</sup> - предоставя класове за манипулиране на дати и часове;
- logging<sup>29</sup> - дефинира функции и класове, които прилагат гъвкава система за регистриране на събития за приложения и библиотеки;
- sklearn<sup>30</sup> - инструмент за Машинно самообучение<sup>31</sup> и анализ на Python.

Променливите с входна информация и данни за работата на URLs Finder са следните:

- version - идентификация на файловете с извлечена информация по дата, идентификационен номер или друго;
- title - име на проекта, използвано и за имена на файлове;
- startpath - директория, в която се намира csv<sup>32</sup> файл с информация за предприятия от съвкупността;
- scraperspath - директория, в която се запазват csv файлове с извлечена информация за предприятия;
- logpath - директория, в която се записва информация за регистриране на събития от работата на софтуера;

<sup>16</sup> <https://docs.python.org/3/library/glob.html>

<sup>17</sup> <https://bg.wikipedia.org/wiki/Unix>

<sup>18</sup> <https://pypi.org/project/beautifulsoup4/>

<sup>19</sup> <https://pypi.org/project/requests/>

<sup>20</sup> [https://en.wikipedia.org/wiki/Hypertext\\_Transfer\\_Protocol](https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol)

<sup>21</sup> <https://docs.python.org/3/library/re.html>

<sup>22</sup> <https://numpy.org/>

<sup>23</sup> <https://docs.python.org/3/library/time.html>

<sup>24</sup> <https://docs.python.org/3/library/urllib.parse.html>

<sup>25</sup> <https://docs.python.org/3/library/urllib.parse.html>

<sup>26</sup> <https://en.wikipedia.org/wiki/URL>

<sup>27</sup> <https://pypi.org/project/tqdm/>

<sup>28</sup> <https://docs.python.org/3/library/datetime.html>

<sup>29</sup> <https://docs.python.org/3/library/logging.html>

<sup>30</sup> <https://scikit-learn.org/stable/>

<sup>31</sup> [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)

<sup>32</sup> [https://en.wikipedia.org/wiki/Comma-separated\\_values](https://en.wikipedia.org/wiki/Comma-separated_values)

- `blacklistpath` - директория, в която се намира csv файл с черен списък с URL адреси;
- `startfile` - име на csv файл с информация за предприятия от съвкупността;
- `scrapefile` - име на csv файл с извлечена информация от уебсайтове;
- `sapifile` - име на csv файл с извлечена информация от търсачката Duck Duck Go<sup>33</sup>;
- `toscrapefile` - име на csv файл с URL адреси от уебсайтове, от които ще бъде извлечена информация;
- `logfile` - име на регистрационен файл за събития;
- `blacklistfile` - име на csv файл с черни списъци с URL адреси;
- `csv_delimiter` - разделител на csv файла, например: „;“;
- `csv_encoding` - кодиране на csv файла, например: „utf-8“;
- `headers` - информация за HTTP заявка.

За изпълнение на софтуер URLs Finder е използван Jupyter Notebook<sup>34</sup>. В Jupyter Notebook последователно се извикват команди за зареждане на модулите на софтуер и за изпълнение на техните методи. Първо се посочва пътят до директорията, в която се намира софтуерът. Дават се стойности на променливите. Зарежда се модулът за регистриране на събития за приложения и библиотеки и се изпълнява неговият метод за започване на работа за регистриране.

### **Намиране на кандидат-интернет адреси на предприятия**

Намирането на кандидат-интернет адреси на предприятия започна със зареждане на модула на софтуера за извличане на данни от интернет и с инициализиране на променливите. Продължи със зареждането на информацията за предприятията от съвкупността взета, от статистическия Бизнес регистър и с информацията за нежеланите интернет адреси от черния списък като жълти страници, новинарски сайтове, интернет директории и други.

### **Изпращане на заявки с информация за предприятията към търсеща машина**

Методът `querySearchEngine` на URLs Finder е използван за изграждане на списък с кандидат-интернет адреси на предприятията. Методът изпраща заявки с имената на предприятията към търсещата машина. Търсещата машина, използвана от софтуера, е Duck Duck Go. Търсачката връща списък с до 10 предполагаеми интернет адреса за всяко предприятие (фиг. 6).

<sup>33</sup> <https://duckduckgo.com>

<sup>34</sup> <https://jupyter.org/> - в URLs Finder е включен Jupyter Notebook файл с тестови данни, команди и резултати. В тази статия няма да бъдат посочвани командите. Ще бъдат демонстрирани само резултати.

**Фиг. 6. Предложения за интернет адреси на НСИ от уебтърсачката**

Out [12]:	Has equal domain	ID	Link position	Name	Suggested URL	URL	Has Simple Suggested URL
0	1.0	000695146	0.0	National Statistical Institute	https://www.nsi.bg/en	https://www.nsi.bg	1
1	0.0	000695146	1.0	National Statistical Institute	http://www.insse.ro/cms/en	https://www.nsi.bg	0
3	0.0	000695146	3.0	National Statistical Institute	https://www.niss.org/	https://www.nsi.bg	1
4	0.0	000695146	4.0	National Statistical Institute	https://ec.europa.eu/eurostat/web/links	https://www.nsi.bg	0
5	0.0	000695146	5.0	National Statistical Institute	https://www.ons.gov.uk/	https://www.nsi.bg	1
6	0.0	000695146	6.0	National Statistical Institute	https://www.statistics.gov.nw/statistical-publ...	https://www.nsi.bg	0
7	0.0	000695146	7.0	National Statistical	https://nis.gov.kh/	https://www.nsi.bg	1

За да се избегне злоупотреба с ресурсите на търсещата машина и евентуална забрана за използване, заявките към търсачката по име за всяко предприятие се изпълняват през 6 секунди. Това увеличава времето за работа на две денонощия, повишава риска от прекъсване поради грешка в софтуера, загуба на храняване, загуба на интернет свързаност и други. Използваният софтуер не поддържа функция за продължаване от мястото на прекъсване. Поради това съвкупността от предприятия беше разделена на 100 множества без повторение, всяко с 282 или 283 предприятия. Така при прекъсване може да се продължи от множеството, където е възникнало прекъсване, без да се налага повторно търсене за всички вече намерени предложения. Търсенето върху всяко множество продължава средно по 33 минути (фиг. 7). В резултат на търсенето за съвкупността от 28 251 предприятия бяха предложени около 250 хил. потенциални интернет адреса.

**Фиг. 7. Успешно приключване на търсене на кандидат-URLs адреси на предприятия с търсеща машина за 4 от 100-те подмножества от изследваната съвкупност от предприятия**

```

.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 837067392 СТРОЙРЕСУРС ООД : 100% | ██████████ | 283/283 [32:46<00:00, 6.98s/it]

.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 837067442 ХРАНИТЕЛНИ СТОКИ - ШУМЕН АД : 100% | ██████████ | 283/283 [32:46<00:00, 7.01s/it]

.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 837068124 ШУМЕН - ПЪТНИЧЕСКИ АВТОТРАНСПОРТ ООД : 100% | ██████████ | 283/283 [32:50<00:00, 6.95s/it]

.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 837068284 ВОДОСНАБДЯВАНЕ И КАНАЛИЗАЦИЯ-ШУМЕН ООД : 100% | ██████████ | 283/283 [33:01<00:00, 7.00s/it]

.\sbr_data\SBR_Data_2020.csv

```

Всички неуспешни търсения се натрупват в обект с табличен формат. Използваната версия на софтуера не записва таблицата във файл и тази информация е загубена. В новата версия на софтуера този недостатък е отстранен. Грешките се

класифицират като проблеми с връзката, HTTP<sup>35</sup> грешка, изчерпване на времето, твърде много пренасочвания, грешка в заявката и обща грешка.

### **Намиране на страници за извличане на информация за предприятията**

За всеки предложен потенциален интернет адрес на предприятие са извлечени първите 10 и последните 10 интернет връзки от страницата на интернет адреса. Извлечените връзки се филтрират и се запазват само тези с еднакъв домейн, като например потенциалният интернет адрес, а дублираните адреси се премахват. Отново процесът се извършва върху 100-те множества на съвкупността от предприятията с цел възобновяване след прекъсване без загуба на информация и повторно извличане. Всеки предложен адрес се проверява дали присъства в черния списък с адреси. Ако присъства, се изключва от предложените възможни адреси за предприятието. Така количеството на потенциалните интернет адреси на предприятията е сведено до около 90 хиляди. Търсенето върху всяко множество продължава средно по 30 минути (фиг. 8). В резултат на търсенето за съвкупността от 28 251 предприятия бяха предложени около 700 хил. страници за извличане на информация за потенциални интернет адреси.

**Фиг. 8. Успешно приключване на търсене на страници на кандидат-интернет адреси на предприятия на 3 от 100-те подмножества от изследваната съвкупност от предприятия**

```

29
.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 838166048 https://www.lifebites.bg/yanko-yanev/ : 100%|██████████| 896/896 [30:34<00:00, 2.30s/it]

30
.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 838166596 https://lemma-ecoinvest.com/proekti/ : 100%|██████████| 899/899 [29:34<00:00, 1.30s/it]

31
.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 838168266 http://www.infostock.bg/infostock/control/quotes/HES : 100%|██████████| 894/894 [28:21<00:00, 3.26s/it]

32
.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)

```

Отново информацията за възникналите грешки не е достъпна, като проблемът е отстранен в новата версия на софтуера.

### **Извличане на данни за предприятията от страници на кандидат-интернет адреси на предприятия**

Извличане на информация за предприятията от предложените интернет адреси и техните страници се проведе върху всяко от 100-те множества без повторение на цялата съвкупност от предприятия. Продължи 2 седмици, като всяко множество се обработва за около 3 часа средно (фиг. 9).

<sup>35</sup> <https://bg.wikipedia.org/wiki/HTTP>

**Фиг. 9. Успешно приключване на извличане на информация за предприятията от предложените интернет адреси и техните страници**

```

74
.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 836227335 https://genik.eu/index.php/coffee/home : 100% | ██████████ | 7665/7665 [2:09:29<00:00, 2.13it/s]
75
.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 836227520 https://www.perota.bg/terms-travel.asp : 100% | ██████████ | 6730/6730 [1:58:23<00:00, 1.29s/it]
76
.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)
processed: 834025349 https://www.nasamnatam.com/reklama.html : 100% | ██████████ | 6611/6611 [1:55:57<00:00, 1.63it/s]
77
.\sbr_data\SBR_Data_2020.csv
All records (28251, 8)

```

След приключване на работата на този етап, софтуерът връща данни за открита информация за предприятията на всяка страница от списъците с кандидат-URLs адреси (фиг. 10). Структурата на таблицата е следната:

- ID - ЕИК на предприятието;
- Name - име на предприятието;
- URL - интернет адрес на предприятието, който знаем от статистическия Бизнес регистър;
- Suggested URL - кандидат- (потенциалният, предложеният от търсачката) интернет адрес на предприятието;
- Link position - пореден номер на предложеният от търсачката интернет адрес за предприятието;
- Has equal domain - 1, ако домейнът на известния и предложеният адрес са равни, иначе 0;
- Has Simple Suggested URL – 1, ако предложеният интернет адрес съдържа само протокол, домейн и език „en“, иначе 0;
- URL to scrape - интернет страницата, от която са получени данните;
- Status code - код за статус от HTTP заявката към URL to scrape;
- Has ID - 1, ако ЕИК на предприятието е открит в текста на интернет страницата, от която са получени данните, иначе 0;
- Has Name - 1, ако името на предприятието е открито в текста на страницата, иначе 0;
- Has Phone - 1, ако стационарният телефон на предприятието е открито в текста на страницата, иначе 0;
- Has GSM - 1, ако мобилният телефон на предприятието е открито в текста на страницата, иначе 0;
- Has Address - 1, ако адресът на предприятието е открито в текста на страницата, иначе 0;
- Has Populated place - 1, ако населеното място на предприятието е открито в текста на страницата, иначе 0;

- Has Email - 1, ако електронната поща на предприятието е открито в текста на страницата, иначе 0;
- Has equal Email and URL Domains - 1, ако домейните на електронната поща на предприятието и предложения интернет адрес са равни, иначе 0.

**Фиг. 10. Данни за информация за предприятията за всяка страница за кандидат-интернет адресите на предприятията**

as mail	Has GSM	Has ID	Has Name	Has Phone	Has Populated place	Has Simple Suggested URL	Has equal Email and URL Domains	Has equal domain	ID	Link position	Name	Status code	Suggested URL	URL	URL to scrape
3.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	127041015	7.0	ИНТЕРНЕТ СЪРВИС ООД	200.0	https://optinet.bg/	NaN	https://optinet.bg/
3.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	127041015	7.0	ИНТЕРНЕТ СЪРВИС ООД	200.0	https://optinet.bg/	NaN	https://optinet.bg /интернет/
3.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	127041015	7.0	ИНТЕРНЕТ СЪРВИС ООД	200.0	https://optinet.bg/	NaN	https://optinet.bg /видеонаблюдение/
3.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	127041015	7.0	ИНТЕРНЕТ СЪРВИС ООД	200.0	https://optinet.bg/	NaN	https://optinet.bg /контакти/
3.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	127041015	7.0	ИНТЕРНЕТ СЪРВИС ООД	200.0	https://optinet.bg/	NaN	https://optinet.bg /общо-условия/
3.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	127041015	7.0	ИНТЕРНЕТ СЪРВИС ООД	200.0	https://optinet.bg/	NaN	https://optinet.bg /политика-за- поверителност
3.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	127041015	7.0	ИНТЕРНЕТ СЪРВИС ООД	200.0	https://optinet.bg/	NaN	https://optinet.bg /политика-за- поверителност/
3.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	127041015	9.0	ИНТЕРНЕТ СЪРВИС ООД	200.0	https://inetdec.nra.bg/	NaN	https://inetdec.nra.bg/

Отново информацията за възникналите грешки не е достъпна, като проблемът е отстранен в новата версия на софтуера.

Извличането на информацията се извършва чрез сравняване на низове с малки букви, като при съвпадение на низовете се присвоява стойност 1, а при разлика се присвоява стойност 0. В бъдеще може да се приложи алгоритъм<sup>36</sup> за Fuzzy logic<sup>37</sup> за откриване на алтернативно изписване на низове за телефони, GSM, адреси и други. Получените резултати от този алгоритъм могат да се използват за по-прецизно оценяване на наличието на търсената информация. Също така може да се пренапише софтуерът, за да открива повече от един низ за дадено поле - например телефон, GSM и други.

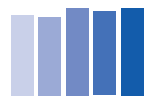
Следващата стъпка обединява множествата без повторение от предприятия в един масив. Софтуерът добавя две нови полета към таблицата:

- Has URL - 1, ако знаем интернет адреса на предприятието от статистическия Бизнес регистър, иначе 0;
- sum - сумата по редове на колоните Has equal Email and URL Domains, Has Email, Has Name, Has Phone, Has Address, Has ID and Has Populated place, групирани по полета ID, Name, Suggested URL, Link position and Status code и агрегирани в колона sum.

<sup>36</sup> <https://pypi.org/project/fuzzywuzzy/>

<sup>37</sup> [https://en.wikipedia.org/wiki/Fuzzy\\_logic](https://en.wikipedia.org/wiki/Fuzzy_logic)





След това софтуерът извършва премахване на дублираните записи по всички полета, като запазва само първия намерен. Резултатът е таблица с 62 160 записа в 18 колони (фиг. 11).

**Фиг. 11. Таблица с данни за открита информация за предприятията по страници на кандидат-интернет адреси**

Suggested URL	Link position	Status code	Has Address	Has Email	Has GSM	Has ID	Has Name	Has Phone	Has Populated place	Has Simple Suggested URL	Has equal Email and URL Domains	Has equal domain	sum	URL	Has URL
https://free-images.com/display/atanas_teshovs...	8.0	200.0	0	0	0.0	0	0	1	0	0	0	0	20.0	http://osnatpk.com/gd_teshovski.php	1
https://www.wikiwand.com/bg/Потребителска_кооп...	1.0	200.0	0	0	0.0	0	0	1	0	0	0	0	9.0	NaN	0
https://www.wikizero.com/bg/Потребителска_кооп...	4.0	200.0	0	0	0.0	0	0	1	0	0	0	0	2.0	NaN	0
https://www.multitran.com/m.exe?a=3&11=15&l2=3...	6.0	200.0	0	0	0.0	0	0	1	0	0	0	0	20.0	NaN	0

### Подготовка на модел за Логистична регресия за намиране на интернет адреси на предприятията

Получените данни от предишния етап се използват за намиране на интернет адреси на предприятията с използване на Логистична регресия. За целта е приложен методът  $\text{prerepareLR}^{38}$  на софтуера URLs Finder, като данните са разделени на 70% обучително множество и 30% тестово множество. Методът изчислява променлива Score, като:

$$\text{Score колона} = \text{sum колона} - \text{sum колона} * \text{Link position колона} / 100$$

Методът избира само записите с най-висок Score от всички уникални ЕИК на предприятията и ги зарежда в модела на логистичната регресия. Методът използва полетата Has Simple Suggested URL, Has Address, Has Email, Has ID, Has Name, Has Phone, Has Populated place и Has equal Email and URL Domains за независимата променлива и Has equal domain за зависимата променлива на модела на логистичната регресия. Методът връща следните обекти в резултат на прилагането му, които се използват за предсказване на интернет адресите на предприятията от наличните данни и за оценяване на качеството на полученото предсказване:

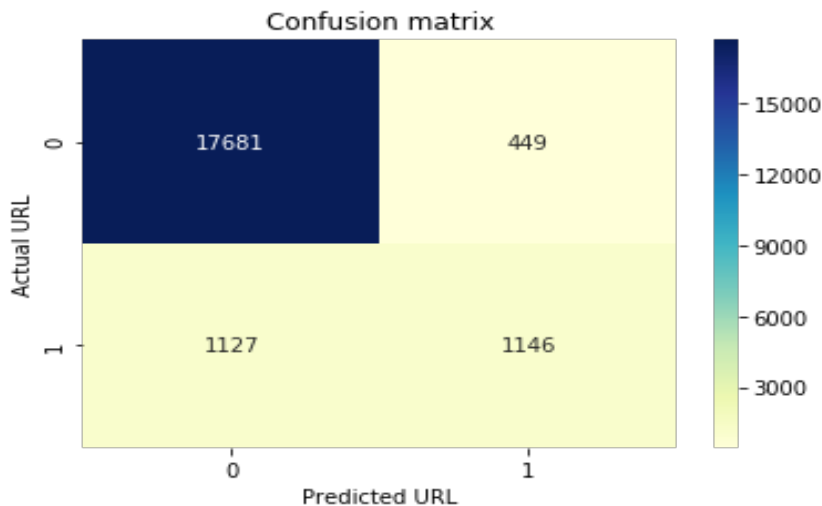
- обучително множество на независимата променлива;
- тестово множество на независимата променлива;
- обучително множество на зависимата променлива;
- тестово множество на зависимата променлива;
- предсказани резултати от тестово множество на зависимата променлива;
- табличен обект с независимата променлива;

<sup>38</sup> Методът е пренаписан в новата версия на софтуера, като са променени начинът на формиране на Score променливата, независимата и зависимата променлива.

- обект с моделът на логистичната регресия.

От тестовото множество на зависимата променлива и предсказаните резултати от тестово множество на зависимата променлива е получена следната Матрица на неточностите<sup>39</sup> (фиг. 12):

**Фиг. 12. Матрица на неточностите за use-case 2**

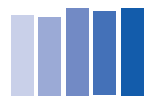


От матрицата може да се заключи, че:

- точност - моделът е коректен в 92% от случаите;
- прецизност - когато моделът предсказва интернет адреси на предприятия, то моделът е коректен в 72% от случаите;
- чувствителност - когато знаем интернет адреса на предприятието, то моделът коректно предсказва интернет адреса на предприятието само в 50% от случаите;
- специфичност - когато не знаем интернет адреса на предприятието, то моделът коректно не предсказва интернет адреса на предприятието в 98% от случаите и предсказва некоректно само в 2% от случаите;
- F1 оценка - със стойност от 0.59 (скала от 0 до 1) показва значителен дисбаланс между прецизността и чувствителността на модел и има какво да се желае за подобряването му въпреки високата точност, която отчита;
- MCC<sup>40</sup> - със стойност от 0.56 (скала от -1 до 1) показва, че моделът е по-близо до перфектното предсказване (1) отколкото до произволното (0), но все още е далече от идеалните резултати.

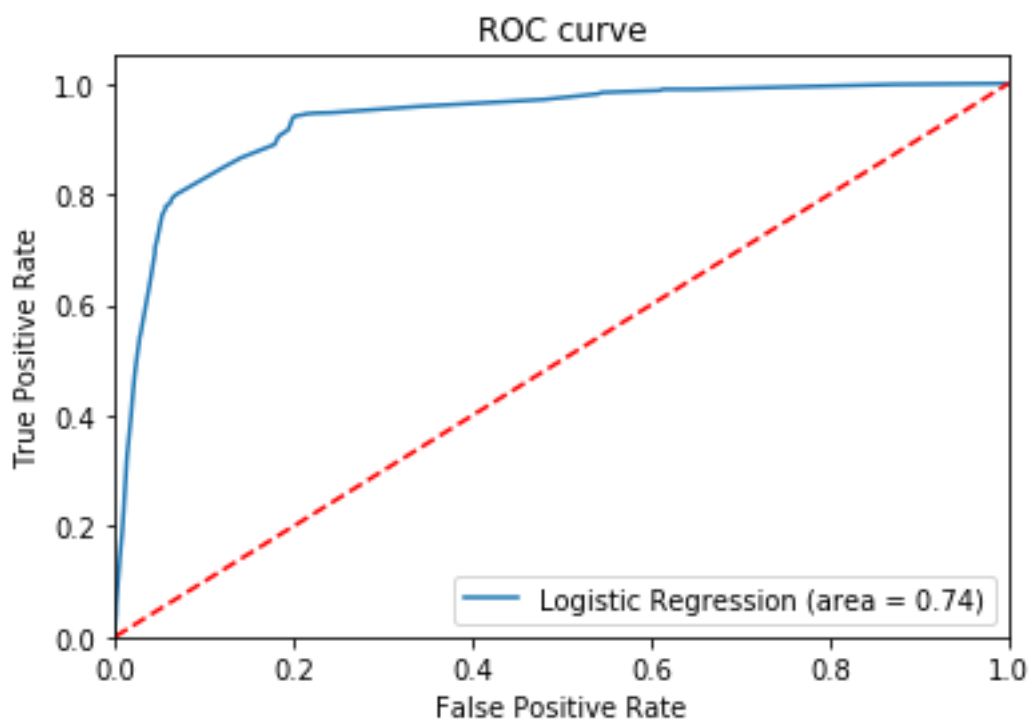
<sup>39</sup> Confusion matrix - [https://en.wikipedia.org/wiki/Confusion\\_matrix](https://en.wikipedia.org/wiki/Confusion_matrix)

<sup>40</sup> Matthews Correlation Coefficient (MCC) - [https://en.wikipedia.org/wiki/Matthews\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Matthews_correlation_coefficient)



От ROC-кривата<sup>41</sup> (фиг. 13) се вижда, че площта под синята линия е значителна и тя се приближава към стойност 1. Това показва, че моделът има потенциал за предсказване на наличието на URLs адреси на предприятията.

**Фиг. 13. ROC-крива**



### **Прилагане на избрания модел върху данните**

С метода prepareP на URLsFinder моделът на логистичната регресия се прилага върху таблицата с данни за открита информация за предприятията по страници на кандидат URLs адресите. Резултатът е таблица със следната структура (фиг. 14):

- ID - ЕИК на предприятието;
- Name - име на предприятието;
- URL - интернет на адрес на предприятието, който знаем от статистическия Бизнес регистър;
- Suggested URL - кандидат- (потенциалният, предложеният от търсачката) интернет адрес на предприятието;
- Link position - пореден номер на предложението от търсачката интернет адрес за предприятието;
- Score - променливата score от модела на логистичната регресия;
- predict - 1, ако логистичната регресия предсказва, че кандидат-интернет адрес на предприятието е търсеният интернет адрес на предприятието, иначе 0;
- 0 - вероятността предсказанието на логистичната регресия да е грешно;

<sup>41</sup> Receiver Operating Characteristic curve - [https://en.wikipedia.org/wiki/Receiver\\_operating\\_characteristic](https://en.wikipedia.org/wiki/Receiver_operating_characteristic)

- 1 - вероятността предсказанието на логистичната регресия да е вярно.

**Фиг. 14. Таблица с предсказания за интернет адреси на предприятия**

Out [21]:

ID	Name	Link position	Score	Suggested URL	URL	predict	0	1	
0	000000858	ТПКИ АТАНАС ТЕШОВСКИ	8.0	18.40	<a href="https://free-images.com/display/atanas_teshovs...">https://free-images.com/display/atanas_teshovs...</a>	<a href="http://osnatpk.com/jgd_teshovski.php">http://osnatpk.com/jgd_teshovski.php</a>	0	0.994551	0.005449
1	000003402	ПОТРЕБИТЕЛСКА КООПЕРАЦИЯ МАКЕДОНИЈА	1.0	8.91	<a href="https://www.wikiwand.com/bg/Потребителска_кооп...">https://www.wikiwand.com/bg/Потребителска_кооп...</a>	NaN	0	0.994551	0.005449
2	000003402	ПОТРЕБИТЕЛСКА КООПЕРАЦИЯ МАКЕДОНИЈА	4.0	1.92	<a href="https://www.wikizero.com/bg/Потребителска_кооп...">https://www.wikizero.com/bg/Потребителска_кооп...</a>	NaN	0	0.994551	0.005449
3	000003402	ПОТРЕБИТЕЛСКА КООПЕРАЦИЯ МАКЕДОНИЈА	6.0	18.80	<a href="https://www.multitran.com/m.exe?a=3&amp;11=15&amp;12=3...">https://www.multitran.com/m.exe?a=3&amp;11=15&amp;12=3...</a>	NaN	0	0.994551	0.005449
4	000025060	ОБЛАСТЕН КООПЕРАТИВЕН СЪЮЗ - БЛАГОВЕВГРАД	3.0	3.88	<a href="https://www.alo.bg/obiavi/zapoznanstva-eskort/">https://www.alo.bg/obiavi/zapoznanstva-eskort/...</a>	NaN	0	0.994551	0.005449
62152	838178157	ИНФРАСТРОЙ - ИНЖЕНЕРИНГ ООД	9.0	0.00	<a href="http://www.eurobuildingengineering.com/bg/proj...">http://www.eurobuildingengineering.com/bg/proj...</a>	<a href="http://www.ise-yambol.com/">http://www.ise-yambol.com/</a>	0	0.993792	0.006208
62153	838180083	ПРОМИШЛЕНА ЕНЕРГЕТИКА АД	0.0	18.00	<a href="https://www.prom-energo.bg/">https://www.prom-energo.bg/</a>	<a href="https://www.prom-energo.bg/bg/">https://www.prom-energo.bg/bg/</a>	1	0.085908	0.914092
62154	838180083	ПРОМИШЛЕНА ЕНЕРГЕТИКА АД	8.0	3.68	<a href="https://www.euba.bg/index.php?option=com_conte...">https://www.euba.bg/index.php?option=com_conte...</a>	<a href="https://www.prom-energo.bg/bg/">https://www.prom-energo.bg/bg/</a>	0	0.948508	0.051492
62155	838185899	СТАНЕВ КОНСЕРВ ЕООД	3.0	0.00	<a href="https://dawhois.com/site/stanevkonserv.com.html">https://dawhois.com/site/stanevkonserv.com.html</a>	<a href="http://stanevkonserv.com/">http://stanevkonserv.com/</a>	0	0.993792	0.006208
62156	838188870	НЕВА - МЕТАЛ ООД	5.0	4.75	<a href="https://enametel.com/">https://enametel.com/</a>	NaN	0	0.975192	0.024808
62157	838190394	МУЛТИФАРМ - 95 ООД	0.0	18.00	<a href="http://multipharm.eu/page/3/za-nas.html">http://multipharm.eu/page/3/za-nas.html</a>	<a href="http://multipharm.eu/">http://multipharm.eu/</a>	1	0.407451	0.592549
62158	838190394	МУЛТИФАРМ - 95 ООД	6.0	0.00	<a href="https://www.puls.bg/reference/pharmacy/">https://www.puls.bg/reference/pharmacy/</a>	<a href="http://multipharm.eu/">http://multipharm.eu/</a>	0	0.993792	0.006208
62159	838190394	МУЛТИФАРМ - 95 ООД	7.0	0.93	<a href="http://www.zdrave.bg/?c=h&amp;a=i&amp;d=◆◆◆◆◆◆">http://www.zdrave.bg/?c=h&amp;a=i&amp;d=◆◆◆◆◆◆</a>	<a href="http://multipharm.eu/">http://multipharm.eu/</a>	0	0.989774	0.010226

62160 rows × 9 columns

От таблицата са вземат само тези редове, за които полето predict има стойност 1 и полетата URL и Suggested URL се различават, като например ред 163 и 184 (фиг. 15). За тези филтрирани редове предложените интернет адреси се сравняват с известните от СБР регистър и се подготвя таблица с актуализирани адреси на предприятията. Ръчно са прегледани 280 нови предложения за URLs адреси на предприятията, от които 230 са потвърдени; 267 адреси са ревизирани, от които 75 адреси са по-добри от вече известните адреси. Така от известни 12 058 URLs адреси на предприятия бяха актуализирани 75 и бяха добавени нови 230, с което общият брой на адресите станаха 12 288.

**Фиг. 15. Таблица с предсказания за интернет адреси на предприятия, за които полето predict има стойност 1**

Out [23]:

ID	Name	Link position	Score	Suggested URL	URL	predict	0	1	
21	000065763	СМА МИНЕРАЛ БУРГАС ВАР ЕООД	0.0	29.00	<a href="http://smameneralbg.com/">http://smameneralbg.com/</a>	<a href="http://smameneralbg.com/">http://smameneralbg.com/</a>	1	0.150551	0.849449
44	000110952	ПОДЕМ ЕООД	2.0	4.90	<a href="http://podem-eood.com/">http://podem-eood.com/</a>	<a href="http://podem-eood.com/">http://podem-eood.com/</a>	1	0.272833	0.727167
46	000111036	ПРОИЗВОДИТЕЛНА КООПЕРАЦИЯ НА ИНВАЛИДИТЕ ЦАРЕВЕЦ	0.0	4.00	<a href="http://pki-tzarevetz.com/">http://pki-tzarevetz.com/</a>	<a href="http://pki-tzarevetz.com/">http://pki-tzarevetz.com/</a>	1	0.247597	0.752403
82	000275929	АВТОМОТОР КОРПОРАЦИЯ АД	0.0	4.00	<a href="http://leasing.citroen.bg/">http://leasing.citroen.bg/</a>	<a href="http://www.citroen.bg/">http://www.citroen.bg/</a>	1	0.449579	0.550421
163	000620115	ИХБ ЕЛЕКТРИК АД	0.0	39.00	<a href="http://www.ihbelectric.com/">http://www.ihbelectric.com/</a>	<a href="http://www.ihbelectric.com/bg/">http://www.ihbelectric.com/bg/</a>	1	0.054059	0.945941
184	000627259	НАЦИОНАЛНА ПОТРЕБИТЕЛНА КООПЕРАЦИЯ НА СЛЕПИТЕ ...	0.0	42.00	<a href="http://npksb.com/">http://npksb.com/</a>	NaN	1	0.352453	0.647547



## **Резултати**

На базата на събраните, обработени и анализирани данни за URLs адреси на предприятията се установи, че за 2020 г. 43.5% от предприятията от съвкупността на изследването „Използване на ИКТ в предприятията“ имат интернет адрес. За 37.8% от предприятията, имащи между 10 и 49 заети, са намерени интернет адреси, за предприятията, имащи между 50 и 249, този процент е 67.8, а големите предприятия с повече от 250 заети имат и най-голямо отношение на интернет адресите с 85.5%. Над половината предприятия в област София (столица) имат интернет адреси, докато в областите Благоевград и Видин предприятията, имащи интернет адреси, е под една четвърт (фиг. 16). Близо две трети от предприятията в сектори на икономическа дейност „Създаване и разпространение на информация и творчески продукти; Далекосъобщения“ и „Производство и разпространение на електрическа и топлинна енергия и на газообразни горива“ имат интернет адрес. Едва една четвърт от предприятията в сектор „Хотелиерство и ресторантьорство“ имат интернет адрес, което е възможно да се дължи на използване на споделени платформи за резервация или използване на интернет адреси с имената на обектите, а не с имената на предприятията (фиг. 16).

За пълнота на анализа е направено сравнение между официалните данни от изследването „Използване на ИКТ в предприятията“ и експерименталните данни от този процес

(вж.

[https://ec.europa.eu/eurostat/cros/sites/crosportal/files/WPC\\_Experimental\\_statistics\\_BG\\_2020\\_Results.pdf](https://ec.europa.eu/eurostat/cros/sites/crosportal/files/WPC_Experimental_statistics_BG_2020_Results.pdf)).

Фиг. 16. Интернет адреси на предприятията с 10 и повече заети по области

Интернет адреси на предприятията с 10 и повече заети по области

Код по NUTS	Област	Предприятия	Интернет адреси	Отношение на Интернет адресите към Предприятията (%)
		брой	брой	
BG311	Видин	177	43	24.3
BG312	Монтана	336	90	26.8
BG313	Враца	379	145	38.3
BG314	Плевен	666	204	30.6
BG315	Ловеч	392	163	41.6
BG321	Велико Търново	744	283	38
BG322	Габрово	471	231	49
BG323	Русе	848	400	47.2
BG324	Разград	270	77	28.5
BG325	Силистра	243	64	26.3
BG331	Варна	2208	1001	45.3
BG332	Добрич	468	151	32.3
BG333	Шумен	476	179	37.6
BG334	Търговище	298	84	28.2
BG341	Бургас	1678	617	36.8
BG342	Сливен	454	166	36.6
BG343	Ямбол	330	125	37.9
BG344	Стара Загора	1057	470	44.5
BG411	София (столица)	8760	4932	56.3
BG412	София	682	254	37.2
BG413	Благоевград	1400	337	24.1
BG414	Перник	361	99	27.4
BG415	Кюстендил	373	106	28.4
BG421	Пловдив	2908	1347	46.3
BG422	Хасково	699	231	33
BG423	Пазарджик	792	287	36.2
BG424	Смолян	409	104	25.4
BG425	Кърджали	372	98	26.3

Фиг. 17. Интернет адреси на предприятията с 10 и повече заети по сектори на икономическа дейност

Интернет адреси на предприятията с 10 и повече заети по сектори на икономическа дейност

Сектор на икономическа дейност	Предприятия	Интернет адреси	Отношение на Интернет адресите към Предприятията (%)
	брой	брой	
Административни и спомагателни дейности (N)	1256	528	42
Доставяне на води; Канализационни услуги, управление на отпадъци и възстановяване (E)	253	138	54.5
Други дейности (S)	15	9	60
Операции с недвижими имоти (L)	542	239	44.1
Преработваща промишленост (C)	7466	3842	51.5
Производство и разпространение на електрическа и топлинна енергия и на газообразни горива (D)	133	86	64.7
Професионални дейности и научни изследвания (M)	1278	739	57.8
Строителство (F)	3180	1260	39.6
Създаване и разпространение на информация и творчески продукти; Далекосъобщения (J)	1235	813	65.8
Транспорт, складиране и пощи (H)	2107	687	32.6
Търговия; Ремонт на автомобили и мотоциклети (G)	7761	3164	40.8
Хотелиерство и ресторантьорство (I)	3025	783	25.9

## 2. Извличане на онлайн характеристики на предприятията от интернет (use-case 2)

### Подготовка на начални данни и софтуер

Процесът на намиране на онлайн характеристиките на предприятията започва с определяне на съвкупността им. Съвкупността от СБР е същата използвана за актуализирането на интернет адресите на предприятията, а именно предприятия с 10 и повече заети, които към пролетта на 2020 г. са 28 251 на брой. Информацията е същата, като са актуализирани URLs адресите на предприятията съгласно резултатите от предишния процес (use-case 1).

Използваните операционна среда, софтуер, модули и компоненти са същите като избраните за предишния процес. Входните променливи за софтуера са същите. Софтуерът се изпълнява като в предишния процес. Пропуска се само стъпката със събиране на предложения за интернет адреси на предприятията от търсачка, тъй като уебсайтовете на предприятията, от които ще извличаме информация са вече известни.

### Намиране на страници за извличане на информация за предприятията

За всяко предприятие от съвкупността, на което знаем интернет адреса (общо 12 288 на брой), са извлечени първите 20 и последните 20 интернет връзки от страницата на URL адреса. Извлечените връзки се филтрират и се запазват само тези с идентичен домейн и интернет адрес, а дублираните URLs адреси се премахват. Отново стъпката се извършва върху разделената на 100 множества без повторение съвкупност на предприятията с цел възобновяване след прекъсване без загуба на информация и повторно извличане. Търсенето върху всяко множество продължава средно по четири минути и половина (фиг. 18). В резултат на търсенето за съвкупността от 12 288 адреса на предприятия са предложени около 220 хил. страници за извличане на информация за онлайн характеристики на предприятията (ОБЕС).

**Фиг. 18. Успешно приключване на търсене на страници за характеристики на предприятия на 4 от 100-те подмножества от изследваната съвкупност от предприятия**

```
In [13]: dfnt=uf.slice_urls_to_scrape(timeout=30,sleep=0.5,urlsatstart=20,urlsatend=20, slice=15, url='URL', what='find_urls_to_scrape')
<
>
processed: 836146456 http://www.tehniks.bg/ : 100% ██████████ 123/123 [03:49<00:00, 3.84s/it]
Slice: 37
Load file with (12288, 2) rows and columns: .\sbr_data\url.csv
DataFrame columns: ['ID', 'URL']
processed: 836146560 http://tivigroup.com/ : 100% ██████████ 123/123 [04:11<00:00, 1.30s/it]
Slice: 38
Load file with (12288, 2) rows and columns: .\sbr_data\url.csv
DataFrame columns: ['ID', 'URL']
processed: 836147896 http://www.akva3.com/ : 100% ██████████ 123/123 [04:19<00:00, 1.36s/it]
Slice: 39
Load file with (12288, 2) rows and columns: .\sbr_data\url.csv
DataFrame columns: ['ID', 'URL']
processed: 836148005 http://monolitbg.com/ : 100% ██████████ 123/123 [05:18<00:00, 2.10s/it]
```

За разлика от предишния процес за use-case 1 в този процес грешките се съхраняват във файл (фиг. 19), като общият им брой е 551, или 4.5% не отговорили. С подобряване на софтуера (направено в новата версия) могат да бъдат намалени грешки от типа Request exception. Грешки от вида Connection problem и Timeout occurred могат да бъдат намалени, като стъпката се изпълни повторно за тези URLs адреси на предприятията.

**Фиг. 19. Грешки при намиране на страници за извличане на информация за характеристики на предприятията от интернет адресите им**

ID	URL	Error
0 200225006	pizzasiciliana.bg	Request exception
1 200608912	https://www.plama.bg/	Connection problems
2 201088101	https://fashionsupreme.co.uk/contact/	Timeout occurred
3 202634919	www.chemcos.eu	Request exception
4 202854154	www.globewilliams.com	Request exception
5 203089329	www.megastroi.eu	Request exception
6 203352806	https://markanpro.bg/	Connection problems

### **Извличане на данни за характеристики на предприятията от уебсайтовете им**

Преди да започне извличане на данни за характеристиките на предприятията от уебсайтовете им, се подготвят различни тематични списъци с ключови думи, които служат за търсене на различни он-лайн характеристики (Фигура 20), такива като:

- извършване на онлайн търговия през сайта на предприятието (88 ключови думи);
- наличие на обяви за работа на сайта на предприятието (37 ключови думи);
- профили на предприятието в социални мрежи (14 ключови думи);
- информация за контакт;
- политика за бисквитки;
- политика за лични данни;
- условия за използване;
- сертификация на предприятието по различни стандарти;
- профили на потребители на сайта.

От изброените характеристики, само за първите три бяха изчислени експериментални статистически данни.



**Фиг. 20. Ключови думи за характеристики на предприятията от уебсайтовете им**

```
uf.get_obecwords()

Load file with (11, 1) rows and columns: .\sbr_data\OBEC_words_contact.txt
Load file with (13, 1) rows and columns: .\sbr_data\OBEC_words_cookie.txt
Load file with (88, 1) rows and columns: .\sbr_data\OBEC_words_ecommerce.txt
Load file with (12, 1) rows and columns: .\sbr_data\OBEC_words_gdpr.txt
Load file with (12, 1) rows and columns: .\sbr_data\OBEC_words_iso.txt
Load file with (37, 1) rows and columns: .\sbr_data\OBEC_words_job.txt
Load file with (14, 1) rows and columns: .\sbr_data\OBEC_words_socialmedia.txt
Load file with (10, 1) rows and columns: .\sbr_data\OBEC_words_tou.txt
Load file with (6, 1) rows and columns: .\sbr_data\OBEC_words_user.txt

OBEC term matrix words:
['address', 'contact', 'mail', 'phone', 'адрес', 'връзка', 'поща', 'мейл', 'контакт', 'свържеш', 'телефон', 'cookie', 'бисквитка', 'браузър', 'изтрива', 'индивидуални', 'настройки', 'поверителност', 'политика', 'сайт', 'сигурност', 'съхранява', 'трепти', 'функционалност', 'american', 'bag', 'card', 'cart', 'commerce', 'dhl', 'easypay', 'ecommerce', 'e-commerce', 'escont', 'epay', 'eshop', 'e-shop', 'estore', 'e-store', 'euro', 'express', 'mastercard', 'online', 'order', 'password', 'paypal', 'postepay', 'rapido', 'shop', 'speedy', 'store', 'value', 'valuta', 'visa', 'акаунт', 'безплатно', 'валута', 'вземи', 'влизна', 'вноски', 'връщане', 'вход', 'ддо', 'добави', 'достав', 'достъп', 'еконт', 'желания', 'записи', 'изплащане', 'идентификация', 'количка', 'комисионна', 'кредит', 'куриер', 'лв', 'лева', 'лист', 'любими', 'магазин', 'марк', 'наличност', 'намалени', 'оферта', 'парола', 'плащане', 'пратка', 'популярни', 'поръчай', 'поръчка', 'потвържд', 'проверка', 'превод', 'предложени', 'прода', 'продукт', 'промоции', 'проследяване', 'път', 'разход', 'рекламаци', 'слагам', 'спииди', 'спийди', 'списък', 'стоки', 'търго', 'търсен', 'цена', 'цени', 'ценова', '679', 'свс', 'данни', 'егн', 'забравен', 'закон', 'защита', 'лични', 'потребител', 'право', 'рекламента', '9001', '13485', '14001', '15050', '18001', '22000', '25424', '27001', '50001', 'iso', 'бдо', 'сертификат', 'candidate', 'career', 'experience', 'job', 'looking', 'position', 'vacanc', 'work', 'възможност', 'възнаграждение', 'екип', 'завършил', 'кадри', 'кандидат', 'кариер', 'конкурс', 'места', 'место', 'местожителство', 'наш', 'обяв', 'опит', 'отворени', 'позиции', 'предложе', 'при нас', 'работа', 'работни', 'свободни', 'стаж', 'стани', 'търси', 'умения', 'часът', 'човек', 'човешки', 'обучени', 'facebook', 'flickr', 'google', 'instagram', 'linkedin', 'muraspace', 'picasa', 'pinterest', 'slideshare', 'tumbler', 'twitter', 'xing', 'yammer', 'youtube', 'гаранции', 'договор', 'използване', 'клиент', 'общи', 'ползване', 'правила', 'продавач', 'условия', 'влез', 'забравена', 'регистр']
```

Извличане на информация за характеристики на предприятията от уебсайтовете и техните страници върху избраните ключови думи е проведено върху всяко от 100-те множества без повторение на съвкупността от предприятия с URLs адреси. Продължителността на този процес е 4 денонощия, като всяко множество се обработва за около 50 минути средно (фиг. 21).

**Фиг. 21. Успешно приключване на извличане на информация за предприятията от предложените интернет адреси и техните страници**

```
dfnes=uf.slice_urls_to_scrape(timeout=10, sleep=0.5, slice=90, what='word_count')

Slice: 90
DataFrame columns: ['ID', 'URL', 'URL to scrape']
processed: 833160422 https://www.swe-flex.com/сервис/клиентска-служба/ : 100%|██████████| 2240/2240 [55:57<00:00, 1.14s/it]

Slice: 91
DataFrame columns: ['ID', 'URL', 'URL to scrape']
processed: 833161111 https://www.dagaplus.com/za-nas/ : 100%|██████████| 2091/2091 [48:41<00:00, 1.36s/it]

Slice: 92
DataFrame columns: ['ID', 'URL', 'URL to scrape']
processed: 123635807 https://www.informator.bg/cgi-bin/index.pl?_state=AjaxMapAddress&q=rp. Стара Загора, ул. Димчо Стаев №30Б : 34%|██████| 709/2108 [18:04<29:57, 1.29s/it]
processed: 123635807 https://www.informator.bg/cgi-bin/index.pl?_state=AjaxMapAddress&q=t]
processed: 131262158 https://plesio.bg/search.html?codes=1352407,2035359,2049546,2159724,3023974,3023990,2342901,2343010,2343029,2411172,2901986,2901994,2977613&sort=priceasc
processed: 131262158 https://plesio.bg/search.html?codes=1352407,2035359,2049546,2159724,3023974,3023990,2342901,2343010,2343029,2411172,2901986,2901994,2977613&sort=priceasc
processed: 833163129 https://www.raya-press.com/inovacii-konkurentnosposobnost/ : 100%|██████████| 2108/2108 [52:58<00:00, 1.42s/it]
```

След приключване на работата на този етап софтуерът връща данни за броя на откритите ключови думи на всяка от избраните страници от уебсайтовете на предприятията (фиг. 22).

**Фиг. 22. Брой на откритите ключови думи на всяка от избраните страници от уебсайтовете на предприятията**

ID	URL	URL to scrape	address	contact	mail	phone	адрес	връзка	поща	...	използване	клиент	общ	ползване	п
0	020149610	https://www.lyulin-metali.com/	https://www.lyulin-metali.com/tel:+359878103092	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	
1	020149610	https://www.lyulin-metali.com/	https://www.lyulin-metali.com/	0.0	0.0	2.0	0.0	1.0	0.0	...	0.0	1.0	0.0	0.0	
2	020149610	https://www.lyulin-metali.com/	https://www.lyulin-metali.com/3a-nac/	0.0	0.0	1.0	0.0	1.0	0.0	...	0.0	2.0	0.0	0.0	
3	020149610	https://www.lyulin-metali.com/	https://www.lyulin-metali.com/lyulin-metali?chat	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	
4	020149610	https://www.lyulin-metali.com/	https://www.lyulin-metali.com/ycnyri/	0.0	0.0	1.0	0.0	1.0	0.0	...	0.0	2.0	0.0	0.0	
5	020149610	https://www.lyulin-metali.com/	https://www.lyulin-metali.com/продуктова-рама/	0.0	0.0	1.0	0.0	1.0	0.0	...	0.0	2.0	0.0	0.0	
6	020149610	https://www.lyulin-metali.com/	https://www.lyulin-metali.com/3a-nac/	0.0	0.0	1.0	0.0	1.0	0.0	...	0.0	0.0	0.0	0.0	

Установените грешки при работата на софтуера са 921, или по-малко от 0.5% от всички проверени страници (фиг. 23).

**Фиг. 23. Грешки при извличане на ключови думи от страниците на уебсайтовете на предприятията**

ID	URL	URL to scrape	Error
0	108692434	http://www.techno-lux.com/	Timeout occurred
1	115545438	https://tvsatcom.bg	https://tvsatcom.bg/category/shows/kinomaraton/ Connection problems
2	131335001	https://efellows.bg/	https://efellows.bg/about-us Timeout occurred
3	201302854	https://www.oplus.bg	https://www.oplus.bg/catalogue/hartieni_izdeli... Timeout occurred
4	201302854	https://www.oplus.bg	https://www.oplus.bg/catalogue/hartieni_izdeli... Timeout occurred
5	201302854	https://www.oplus.bg	https://www.oplus.bg/catalogue/hartieni_izdeli... Timeout occurred
6	201302854	https://www.oplus.bg	https://www.oplus.bg/catalogue/hartieni_izdeli... Timeout occurred
7	201302854	https://www.oplus.bg	https://www.oplus.bg/catalogue/hartieni_izdeli... Timeout occurred
8	201302854	https://www.oplus.bg	https://www.oplus.bg/catalogue/hartieni_izdeli... Timeout occurred
9	201302854	https://www.oplus.bg	https://www.oplus.bg/catalogue/hartieni_izdeli... Timeout occurred
10	201302854	https://www.oplus.bg	https://www.oplus.bg/catalogue/hartieni_izdeli... Timeout occurred
11	201302854	https://www.oplus.bg	https://www.oplus.bg/catalogue/hartieni_izdeli... Timeout occurred

### Подготовка на модел за Логистична регресия за намиране на ОБЕС

Получените данни от предишния етап са използвани за намиране на онлайн характеристики на предприятията с използване на Логистична регресия. За целта с помощта на методи от софтуера е приготвен файл, съдържащ следните полета (фиг. 24):

- ID - ЕИК на предприятието;
- URL - интернет адреса;

- полета с тематични ключови думи за съответната търсена онлайн характеристика, съдържащи единица, ако думата е намерена в наблюдаваните страници от уебсайта на предприятието, иначе 0;
- ОБЕС - известната страница от сайта на предприятието, на която се среща търсената онлайн характеристика;
- Known ОБЕС - поле с 1, ако знаем търсената характеристика за даденото предприятие, иначе 0;
- Link position - необходимо поле за предишния процес за намиране на интернет адрес;
- Sum - сума от полетата с ключови думи за характеристиката;
- Score - необходимо поле за предишния процес за намиране на интернет адрес.

**Фиг. 24. Файл за ML с Логистична регресия за намиране на ОБЕС (в случая за наличие на обяви за работа на сайта на предприятието)**

```
df[0].info(verbose=False)
df[0].sample(5)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10533 entries, 0 to 10532
Columns: 44 entries, ID to Score
dtypes: float64(39), int64(2), object(3)
memory usage: 3.5+ MB
```

ID	URL	candidate	career	experience	job	looking	position	vacanc	work	...	умения	част	човек	човешки	обучени
9657	816089656	https://www.holcim.bg/bg	0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	...	1.0	1.0	1.0	1.0
10122	831430207	https://amperei.net/bg/	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	1.0	1.0	0.0	1.0	0.0
4738	130589420	https://antares-bg.net/	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	1.0	1.0	1.0	1.0	1.0
9748	822106665	https://www.vikpz.com/	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0
6306	175264381	http://www.call-ex.com/	0.0	1.0	1.0	1.0	0.0	1.0	0.0	...	1.0	1.0	0.0	1.0	1.0

5 rows x 44 columns

idate	career	experience	job	looking	position	vacanc	work	...	умения	част	човек	човешки	обучени	ОБЕС	Known ОБЕС	Link position	sum	Score
0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	...	1.0	1.0	1.0	1.0	1.0	https://www.holcim.bg/bg/kariera-v-kholsim/svo...	1	1	26.0	25.74
0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	1.0	1.0	0.0	1.0	0.0	NaN	0	2	14.0	13.86
0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	1.0	1.0	1.0	1.0	1.0	NaN	0	2	22.0	21.78
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	https://www.vikpz.com/page/15/karieri.html	1	1	7.0	6.93
0.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	...	1.0	1.0	0.0	1.0	1.0	http://www.call-ex.com/bg/careers/	1	1	23.0	22.77

Полученият файл се използва за намиране на онлайн характеристика на предприятието. С помощта на метод от софтуера данните са разделени на 70% обучително множество и 30% тестово множество за машинно самообучение с Логистична регресия. За избраното произволно състояние<sup>42</sup> 3333 са получени 244

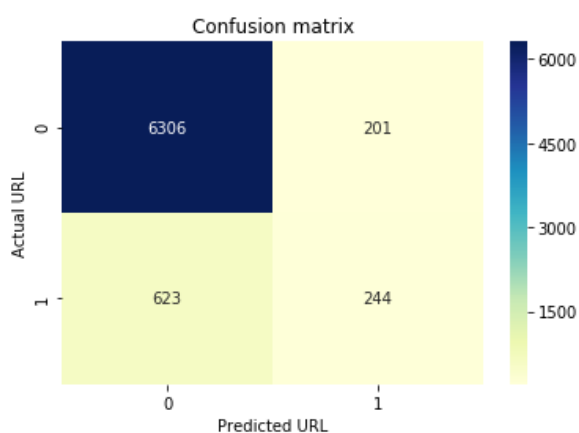
<sup>42</sup> random\_state - [https://scikit-learn.org/stable/glossary.html#term-random\\_state](https://scikit-learn.org/stable/glossary.html#term-random_state)

истински положителни, 6 306 истински отрицателни, 201 фалшиво положителни и 623 фалшиво отрицателни резултата за онлайн характеристиката за наличие на обяви за работа на сайта на предприятието (фиг. 25).

**Фиг. 25. Матрица на неточностите за use-case 2**

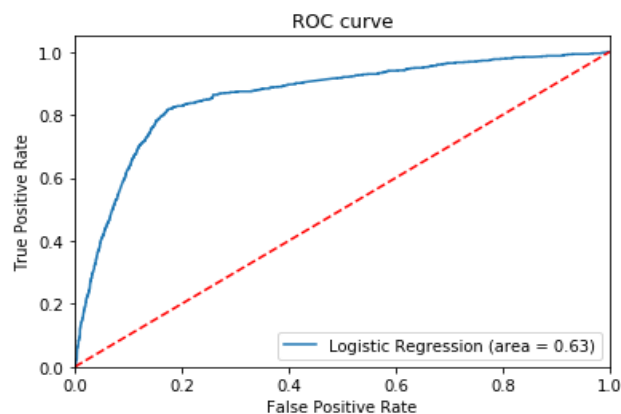
```
lr = ml.logistic_regression_fit(test_size=0.7, random_state=3333,
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10533 entries, 0 to 10532
Columns: 44 entries, ID to Score
dtypes: float64(39), int64(2), object(3)
memory usage: 3.6+ MB
None
```

```
ml.prepareCM(lr[3], lr[4])
```



От матрицата може да се направи заключението, че моделът по-скоро предсказва произволно отколкото точно, което може да се дължи на неподходящи ключови думи за онлайн характеристиката на предприятието или на голям шум в използването на тези думи по интернет страниците. Същият извод се налага и от ROC-кривата (фиг. 26), където се вижда, че площта над синята линия е значителна и синята линия е далече от стойност 1.

**Фиг. 26. ROC-крива за намиране на ОБЕС (в случая за наличие на обяви за работа на сайта на предприятието)**



Въпреки не толкова добрите резултати на модела той е приложен за произволни състояния от 1 до 100 на Логистичната регресия за намиране на онлайн характеристики на предприятията за наличие на електронен магазин и обяви за работа на сайтовете на предприятията. Резултатите от 100-те итерации са интегрирани и изчистени от повторения по ЕИК на предприятието за всяка от двете наблюдавани характеристики (фиг. 27).

**Фиг. 27. Резултат от 100 итерации на Логистичния модел за наличие на обяви за работа на сайтовете на предприятията**

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1956 entries, 368 to 1052772
Columns: 10 entries, ID to 1
dtypes: float64(4), int64(3), object(3)
memory usage: 168.1+ KB
```

Out[19]:

	ID	URL	OБЕС	Known OБЕС	Link position	sum	Score	predict	0	1	
368	000070536	http://www.galateabg.net/	http://galateabg.net/main_bg/10870_karieri.html	1	1	13.0	12.87	1	0.163721	0.836279	
980	000179553	http://gt-vratza.com/	http://gt-vratza.com/tyrgove-konkursi/	1	1	10.0	9.90	0	0.861697	0.138303	
1297	000220021	http://monek-bg.com/		NaN	0	2	13.0	12.87	1	0.477315	0.522685
2743	000443113	https://www.addisan.bg/		NaN	0	2	13.0	12.87	1	0.382640	0.617360
3195	000551929	https://www.silistra.avtogara.eu/bg/	https://www.silistra.avtogara.eu/bg/	1	1	3.0	2.97	0	0.928014	0.071986	
3347	000600649	http://mihalkovo.com/bg/		NaN	0	2	23.0	22.77	1	0.361865	0.638135

След ръчно експертно валидиране е установено, че от предложените 1 917 URLs адреса за електронна търговия на предприятията 1 329 са верни, а от предложените 1 956 интернет адреса за страници с обяви за работа на сайтовете на предприятията само 1 652 са верни.

### 3. Присъствие/профили на предприятията в социалните мрежи

За намиране на профилите на предприятията в социалните мрежи е използван SocialMediaProfiles<sup>43</sup> софтуер на Python, част от средството Starter Kit. Софтуерът проверява сайтове за наличие на връзки към профили в следните социални мрежи:

- Facebook;
- Twitter;
- Youtube;
- LinkedIn;
- Instagram;
- Xing;
- Pinterest.

След подаване на 12 288-те URLs адреса на предприятия, които са известни от use-case 1, софтуерът намира общо 4 398 профила в социални мрежи на предприятия (фиг. 28).

<sup>43</sup> <https://github.com/EnterpriseCharacteristicsESSnetBigData/StarterKit/tree/master/SocialMediaProfiles>

**Фиг. 28. Намиране на профили в социалните мрежи на предприятията със софтуер SocialMediaProfiles**

```

Website currently being scraped: http://www.novsvjat.com/

The length of the scrapped content: 29653 characters
Number of links on website: 15
https://www.facebook.com/%d0%9d%d0%9e%d0%92-%d0%a1%d0%92%d0%af%d0%a2-162209530544830/
https://www.facebook.com/%d0%9d%d0%9e%d0%92-%d0%a1%d0%92%d0%af%d0%a2-162209530544830/
Total number of unique social media links found: 1

Website currently being scraped: http://osnatpk.com/gd_teshovski.php

The length of the scrapped content: 9992 characters
Number of links on website: 40
No social media links have been found.
Preparing to scrape subpages...
Scraping InternalURL_type1: http://osnatpk.com/gd_teshovski.php / mailto:osnatpkblagoevgrad@abv.bg?subject=zapitvane
Exception occurred during processing the following URL:mailto:osnatpkblagoevgrad@abv.bg?subject=zapitvane
ExternalURL_type1: http://osnatpk.com
ExternalURL_type1: http://osnatpk.com/blg_nov_sviat.php

```

## Резултати

На базата на получените данни от уебсайтовете на предприятията с 10 и повече заети е установено, че:

- 4.7% от предприятията имат собствен електронен магазин;
- 10.8% от предприятията с уебсайт извършват електронна търговия през уебсайта си;
- 5.8% от предприятията имат собствена страница с обяви за работа в интернет;
- 13.4% от предприятията с уебсайт имат страница с обяви за работа на него;
- Най-много предприятия имат електронни магазини в София (столица) - 6.9%, а най-малко - в област Видин - 0.6% (фиг. 29);
- От предприятията с уебсайтове най-много електронни магазини имат в област Хасково - 15.2%, а най-малко - отново в област Видин - 2.3% (фиг. 29);
- За предприятията от сектор „Производство и разпространение на електрическа и топлинна енергия и на газообразни горива“ (D) не са открити електронни магазини, докато предприятията от сектор „Други дейности“ (S) е най-вероятно да имат такъв (фиг. 30);
- Най-много предприятия имат страници с обяви за работа в област София (столица) - 11%, а най-малко - в област Видин - 0.3% (фиг. 31);
- От предприятията с уебсайтове най-много страници с обяви за работа имат в област София (столица) - 19.5%, а най-малко - отново в област Кърджали - 1% (фиг. 32);
- Предприятията от сектор „Производство и разпространение на електрическа и топлинна енергия и на газообразни горива“ (D) имат най-много страници с обяви за работа на техните сайтове - 27.9%, докато предприятията от сектор „Хотелиерство и ресторантьорство“ (I) имат най-малко - 7.4% (фиг. 33);

- 19.9% от предприятията с уебсайт имат профили в социални мрежи, като най-популярна сред тях е Facebook - с 18.5% (фиг. 34);
- От предприятията с уебсайтове най-много профили в социални мрежи имат в област Силистра - 29.7%, а най-малко - в област Габрово - 13% (фиг. 34);
- Предприятията от сектор „Хотелиерство и ресторантьорство“ (I) имат най-много профили в социални мрежи на техните сайтове - 29.1%, докато предприятията от сектор „Доставяне на води; Канализационни услуги, управление на отпадъци и възстановяване“ (E) имат най-малко - 12.3% (фиг. 35).

**Фиг. 29. Електронни магазини на предприятията с 10 и повече заети по области**

Електронни магазини на предприятията с 10 и повече заети по области

Код по NUTS	Област	Предприятия	Интернет адреси	Електронни магазини	Отношение на Електронните магазини към Предприятията (%)	Отношение на Електронните магазини към Интернет адресите (%)
		брой	брой	брой		
BG311	Видин	177	43	1	0.6	2.3
BG312	Монтана	336	90	6	1.8	6.7
BG313	Враца	379	145	11	2.9	7.6
BG314	Плевен	666	204	24	3.6	11.8
BG315	Ловеч	392	163	10	2.6	6.1
BG321	Велико Търново	744	283	26	3.5	9.2
BG322	Габрово	471	231	21	4.5	9.1
BG323	Русе	848	400	34	4	8.5
BG324	Разград	270	77	7	2.6	9.1
BG325	Силистра	243	64	3	1.2	4.7
BG331	Варна	2208	1001	113	5.1	11.3
BG332	Добрич	468	151	17	3.6	11.3
BG333	Шумен	476	179	11	2.3	6.1
BG334	Търговище	298	84	9	3	10.7
BG341	Бургас	1678	617	63	3.8	10.2
BG342	Сливен	454	166	13	2.9	7.8
BG343	Ямбол	330	125	14	4.2	11.2
BG344	Стара Загора	1057	470	54	5.1	11.5
BG411	София (столица)	8760	4932	606	6.9	12.3
BG412	София	682	254	12	1.8	4.7
BG413	Благоевград	1400	337	30	2.1	8.9
BG414	Перник	361	99	8	2.2	8.1
BG415	Кюстендил	373	106	9	2.4	8.5
BG421	Пловдив	2908	1347	158	5.4	11.7
BG422	Хасково	699	231	35	5	15.2
BG423	Пазарджик	792	287	22	2.8	7.7
BG424	Смолян	409	104	9	2.2	8.7
BG425	Кърджали	372	98	3	0.8	3.1

**Фиг. 30. Електронни магазини на предприятията с 10 и повече заети по сектори на икономическата дейност**

Електронни магазини на предприятията с 10 и повече заети по сектори на икономическа дейност

Сектор на икономическа дейност	Предприятия	Интернет адреси	Електронни магазини	Отношение на Електронните магазини към Интернет адресите (%)	Отношение на Електронните магазини към Интернет адресите (%)
	брой	брой	брой		
Административни и спомагателни дейности (N)	1256	528	20	1.6	3.8
Доставяне на води; Канализационни услуги, управление на отпадъци и възстановяване (E)	253	138	1	0.4	0.7
Други дейности (S)	15	9	4	26.7	44.4
Операции с недвижими имоти (L)	542	239	7	1.3	2.9
Преработваща промишленост (C)	7466	3842	337	4.5	8.8
Производство и разпространение на електрическа и топлинна енергия и на газообразни горива (D)	133	86	0	0	0
Професионални дейности и научни изследвания (M)	1278	739	6	0.5	0.8
Строителство (F)	3180	1260	19	0.6	1.5
Създаване и разпространение на информация и творчески продукти; Далекосъобщения (J)	1235	813	44	3.6	5.4
Транспорт, складиране и пощи (H)	2107	687	15	0.7	2.2
Търговия; Ремонт на автомобили и мотоциклети (G)	7761	3164	818	10.5	25.9
Хотелиерство и ресторантьорство (I)	3025	783	58	1.9	7.4

**Фиг. 31. Страници с обяви за работа на предприятията с 10 и повече заети по области**

Страници с обяви за работа на предприятията с 10 и повече заети по области

Код по NUTS	Област	Предприятия	Интернет адреси	Страници с обяви за работа	Отношение на Страници с обяви за работа към Интернет адресите (%)	Отношение на Страници с обяви за работа към Интернет адресите (%)
		брой	брой	брой		
BG311	Видин	177	43	1	0.6	2.3
BG312	Монтана	336	90	4	1.2	4.4
BG313	Враца	379	145	18	4.7	12.4
BG314	Плевен	666	204	15	2.3	7.4
BG315	Ловеч	392	163	9	2.3	5.5
BG321	Велико Търново	744	283	27	3.6	9.5
BG322	Габрово	471	231	15	3.2	6.5
BG323	Русе	848	400	53	6.2	13.2
BG324	Разград	270	77	7	2.6	9.1
BG325	Силистра	243	64	6	2.5	9.4
BG331	Варна	2208	1001	123	5.6	12.3
BG332	Добрич	468	151	13	2.8	8.6
BG333	Шумен	476	179	16	3.4	8.9
BG334	Търговище	298	84	5	1.7	6
BG341	Бургас	1678	617	49	2.9	7.9
BG342	Сливен	454	166	10	2.2	6
BG343	Ямбол	330	125	13	3.9	10.4
BG344	Стара Загора	1057	470	41	3.9	8.7
BG411	София (столица)	8760	4932	961	11	19.5
BG412	София	682	254	27	4	10.6
BG413	Благоевград	1400	337	23	1.6	6.8
BG414	Перник	361	99	8	2.2	8.1
BG415	Кюстендил	373	106	4	1.1	3.8
BG421	Пловдив	2908	1347	154	5.3	11.4
BG422	Хасково	699	231	14	2	6.1
BG423	Пазарджик	792	287	29	3.7	10.1
BG424	Смолян	409	104	6	1.5	5.8
BG425	Кърджали	372	98	1	0.3	1



**Фиг. 32. Страници с обяви за работа на предприятията с 10 и повече заети по сектори на икономическата дейност**

Страници с обяви за работа на предприятията с 10 и повече заети по сектори на икономическа дейност

Сектор на икономическа дейност	Предприятия	Интернет адреси	Страници с обяви за работа	Отношение на Страници с обяви за работа към Предприятията (%)	Отношение на Страници с обяви за работа към Интернет адресите (%)
	брой	брой	брой		
Административни и спомагателни дейности (N)	1256	528	105	8.4	19.9
Доставяне на води; Канализационни услуги, управление на отпадъци и възстановяване (E)	253	138	23	9.1	16.7
Други дейности (S)	15	9	1	6.7	11.1
Операции с недвижими имоти (L)	542	239	26	4.8	10.9
Преработваща промишленост (C)	7466	3842	327	4.4	8.5
Производство и разпространение на електрическа и топлинна енергия и на газообразни горива (D)	133	86	24	18	27.9
Професионални дейности и научни изследвания (M)	1278	739	191	14.9	25.8
Строителство (F)	3180	1260	113	3.6	9
Създаване и разпространение на информация и творчески продукти; Далекосъобщения (J)	1235	813	197	16	24.2
Транспорт, складиране и пощи (H)	2107	687	127	6	18.5
Търговия; Ремонт на автомобили и мотоциклети (G)	7761	3164	460	5.9	14.5
Хотелиерство и ресторантьорство (I)	3025	783	58	1.9	7.4

**Фиг. 33. Присъствие в социалните мрежи на предприятия с 10 и повече заети**

Присъствие в социални мрежи на предприятия с 10 и повече заети

Социални мрежи	Присъствие брой	Отношение към всички предприятия (%)	Отношение към предприятията с интернет адрес (%)
Предприятия	2450	8.7	19.9
Facebook	2269	8	18.5
Twitter	558	2	4.5
Youtube	592	2.1	4.8
LinkedIn	551	2	4.5
Instagram	330	1.2	2.7
Xing	9	0	0.1
Pinterest	89	0.3	0.7

**Фиг. 34. Присъствие в социалните мрежи на предприятията с 10 и повече заети по области**

Присъствие в социални мрежи на предприятията с 10 и повече заети по области

Код по NUTS	Област	Предприятия брой	Интернет адреси брой	Присъствие в социални мрежи брой	Отношение на Присъствие в социални мрежи към Предприятията (%)	Отношение на Присъствие в социални мрежи към Интернет адресите (%)
BG311	Видин	177	43	6	3.4	14
BG312	Монтана	336	90	17	5.1	18.9
BG313	Враца	379	145	27	7.1	18.6
BG314	Плевен	666	204	33	5	16.2
BG315	Ловеч	392	163	30	7.7	18.4
BG321	Велико Търново	744	283	52	7	18.4
BG322	Габрово	471	231	30	6.4	13
BG323	Русе	848	400	74	8.7	18.5
BG324	Разград	270	77	15	5.6	19.5
BG325	Силистра	243	64	19	7.8	29.7
BG331	Варна	2208	1001	216	9.8	21.6
BG332	Добрич	468	151	35	7.5	23.2
BG333	Шумен	476	179	34	7.1	19
BG334	Търговище	298	84	17	5.7	20.2
BG341	Бургас	1678	617	120	7.2	19.4
BG342	Сливен	454	166	29	6.4	17.5
BG343	Ямбол	330	125	26	7.9	20.8
BG344	Стара Загора	1057	470	73	6.9	15.5
BG411	София (столица)	8760	4932	1007	11.5	20.4
BG412	София	682	254	46	6.7	18.1
BG413	Благоевград	1400	337	62	4.4	18.4
BG414	Перник	361	99	15	4.2	15.2
BG415	Кюстендил	373	106	25	6.7	23.6
BG421	Пловдив	2908	1347	295	10.1	21.9
BG422	Хасково	699	231	57	8.2	24.7
BG423	Пазарджик	792	287	48	6.1	16.7
BG424	Смолян	409	104	24	5.9	23.1
BG425	Кърджали	372	98	18	4.8	18.4

**Фиг. 35. Присъствие в социалните мрежи на предприятията с 10 и повече заети по сектори на икономическата дейност**

Присъствие в социални мрежи на предприятията с 10 и повече заети по сектори на икономическа дейност

Сектор на икономическа дейност	Предприятия брой	Интернет адреси брой	Присъствие в социални мрежи брой	Отношение на Присъствие в социални мрежи към Предприятията (%)	Отношение на Присъствие в социални мрежи към Интернет адресите (%)
Административни и спомагателни дейности (N)	1256	528	115	9.2	21.8
Доставяне на води; Канализационни услуги, управление на отпадъци и възстановяване (E)	253	138	17	6.7	12.3
Други дейности (S)	15	9	2	13.3	22.2
Операции с недвижими имоти (L)	542	239	40	7.4	16.7
Преработваща промишленост (C)	7466	3842	734	9.8	19.1
Производство и разпространение на електрическа и топлинна енергия и на газообразни горива (D)	133	86	12	9	14
Професионални дейности и научни изследвания (M)	1278	739	147	11.5	19.9
Строителство (F)	3180	1260	208	6.5	16.5
Създаване и разпространение на информация и творчески продукти; Далекосъобщения (J)	1235	813	157	12.7	19.3
Транспорт, складиране и пощи (H)	2107	687	134	6.4	19.5
Търговия; Ремонт на автомобили и мотоциклети (G)	7761	3164	656	8.5	20.7
Хотелиерство и ресторантьорство (I)	3025	783	228	7.5	29.1

## VI. Заключение

През последните десетилетия комбинираното използване на данни от различни източници за статистически цели се превърна в консолидирана практика. Наред с използването на административни данни за създаване на статистически регистри експериментирането по отношение на използването на източници на големи данни продължава, за да се актуализира, разшири и валидира наличната информация в статистическия Бизнес регистър. Най-използваният и най-достъпният източник за намиране на големи данни е интернет мрежата. Голямото количество налична информация предоставя нови възможности, но също така и нови предизвикателства пред статистическите експерти по интеграция на данни, предвид структурните разлики между административните и уебданните.

В административните източници идентификацията на единиците е сигурна. В допълнение, интегрирането на данните за единицата е лесно, тъй като те използват общ идентификационен код (данъчен код, ДДС номер и др.). Чрез консолидиран процес на интеграция се присвоява идентификационен код за правните единици и евентуално идентификационен код за предприятия.

За разлика от административните данни големите данни крият сериозни рискове. Някои от тях са очевидни като трудността да се управляват бързо нарастващи обеми данни, водещи до голямо потребление на изчислителни ресурси и ресурси за съхранение. Освен това има технически ограничения за решаване като например дългосрочното време, необходимо на скрапера да обходи цялото съдържание, и ограниченията, свързани с информационната сигурност на уебсайтовете, които пречат на автоматичния достъп. Не са за подценяване и статистическите проблеми като например: трудността да се удостовери качеството на уебинформацията и надеждността на данните, как да се съчетава уебинформацията за дадено предприятие с тази в СБР и как да се гарантира сигурността на класифицирането в съвкупността, към която извлечените от уебмрежата данни принадлежат.

Въпреки посочените рискове използването на уебданни има и очевидни ползи като например обогатяване на статистическото производство с нова информация, подобряване на навременността на статистическите продукти и увеличаване на приложимостта на бизнес статистиката за сметка на по-ниски разходи в сравнение с увеличаването на съществуващите данни. Силен аргумент в тази посока е също, че интернет мрежата е независим източник на данни, докато всички останали източници - административни и статистически, могат да се считат за свързани по някакъв начин и да си влияят взаимно. Цялостната картина, предоставена от уебданните, е със сигурност по-реална, а именно: как едно предприятие вижда себе си и как иска да се представи пред потребителите си.

Сравнително новата идея да се използват големи данни като допълнителен източник за СБР чрез използване на методите за уебскрапване и технологии за извличане на текст с цел интегриране на „структурираните“ бизнес данни с „неструктурираните“ уебданни е добре приета от статистическата общност в ЕСС. Техниките за големи данни променят начина, по който се събират, обработват, анализират и интегрират данни.

Добавената стойност в това отношение се крие точно в информацията, която е скрита в данните и в тяхното проактивно използване, т.е. четенето и използването на данните като отправна точка за създаване на стратегия. Следователно интегрирането на инструментите за анализ на големи данни в контекста на традиционния статистически производствен процес е трудна задача. Това всъщност означава да се комбинират управлявани от данни процеси - въз основа на входни данни, които не идват от статистически източници, нееднородни са, неструктурирани и нестабилни във времето - с процеси, базирани на ориентиран към изхода подход, тъй като в контекста на официалната статистика производственият процес е изграден с оглед получаване на статистически изходи.

Интернет като източник на данни представлява нови възможности и предизвикателства за официалната статистика, която трябва да включи всички иновационни, потенциални източници на данни, колкото е възможно повече в концептуалния дизайн на своите изследвания. Все повече национални статистически организации експериментират с използването на алтернативни източници на данни, за да произведат една и съща или нова статистическа информация по-ефективно и с високо ниво на качество в една наситена среда с много източници на данни.

Има няколко важни задачи, които предстои да бъдат решени от националните статистически служби по отношение на ефективното използване на източници на големи данни като цяло. Основната задача обаче е как на практика да се премине от експериментиране към реално производство на статистика от големи данни. Тази стъпка включва различни аспекти, вариращи от спазване на поверителността на личните данни до необходимостта от изграждане на цялостно нова инфраструктура (методологична, технологична, организационна), както и придобиване на нови умения от експертния състав.

На базата на получените резултати може да се твърди, че извлечените от интернет мрежата данни за онлайн характеристиките на предприятията от техните уебсайтове притежават необходимия потенциал да бъдат допълнителен източник за производство на официална статистика. Това е и основната цел, която се постига чрез изпълнението на първите два случая на използване (описани в настоящата статия), а именно: да се вземе решение за използване на новата информация за производство на още по-подробна статистика за използването на ИКТ в предприятията.

## ПРОИЗВОДСТВО НА ЕКСПЕРИМЕНТАЛНА СТАТИСТИКА ЗА ХАРАКТЕРИСТИКИ НА ПРЕДПРИЯТИЯТА С ДАННИ ОТ ИНТЕРНЕТ

*Костадин Георгиев\*, Галя Статева\*\**

**РЕЗЮМЕ** С всяка изминала година онлайн бизнесът става все по-важен за икономиката, а с наличието на глобалната пандемия COVID е по-важен от всякога. Работният пакет С (WPC) за онлайн базирани характеристики на предприятията (OBEC) в рамките на европейския проект ESSnet on Big data II е свързан с разбирането на онлайн икономическата и бизнес активност на предприятията от гледна точка на националната статистика.

Ключов резултат от работата е подобряването на качеството на статистическия бизнес регистър по отношение на характеристики за онлайн присъствието на национално регистрираните фирми като наличието на уебсайтове, електронна търговия или акаунти в социални медии. В рамките на WPC методологията от предишния проект ESSnet on Big data I беше обобщена и разширена с цел използване във всяка държава от ЕСС, като се вземе предвид разнообразието, необходимо за поддържане на различните случаи на използване (use-cases) в статистическата практика.

Националният статистически институт имаше честта и привилегиата да бъде водеща институция и активен партньор по изпълнението на дейностите в рамките на работен пакет С.

Настоящата статия има за цел да запознае читателя с постигнатите резултати относно възможностите и предизвикателствата за производство на експериментална статистика от източници на „големи данни“ за онлайн характеристики на предприятията на европейско и национално ниво.

**Ключови думи:** онлайн базирани характеристики на предприятията, експериментална статистика, големи данни, статистически Бизнес регистър, Национален статистически институт

\* Главен експерт в отдел „Информационни системи и приложен софтуер“, НСИ; e-mail: kgeorgiev@nsi.bg.

\*\* Държавен експерт в дирекция „Обща методология, координация и анализ на статистическите изследвания“, НСИ; e-mail: gstateva@nsi.bg.

## ПРОИЗВОДСТВО ЭКСПЕРИМЕНТАЛЬНОЙ СТАТИСТИКИ ДЛЯ ХАРАКТЕРИСТИКИ ПРЕДПРИЯТИЙ НА ОСНОВЕ ИНТЕРНЕТ-ДАННЫХ

*Костадин Георгиев\*, Галя Статева\*\**

**РЕЗЮМЕ** С каждым годом онлайн-бизнес становится все более важным для экономики, а в условиях глобальной пандемией COVID важнее, чем когда-либо. Рабочий пакет С (WPC) для онлайн-характеристик предприятий (OBEC) в рамках европейского проекта ESSnet on Big data II связан с пониманием онлайн экономической и бизнес активности предприятий с точки зрения национальной статистики.

Ключевым результатом работы является повышение качества статистического бизнес регистра в отношении характеристик онлайн-присутствия зарегистрированных национальных компаний, включающих наличие веб-сайтов, электронной торговли или аккаунтов в социальных медиа. В рамках WPC методология из предыдущего проекта ESSnet on Big data I была обобщена и расширена с целью ее использования в каждой стране ЕСС, принимая во внимание разнообразие, необходимое для поддержания различных вариантов ее использования (use-cases) в статистической практике.

Национальный статистический институт удостоился чести и привилегии быть ведущей службой и активным партнером в реализации мероприятий в рамках рабочего пакета С.

Эта статья призвана познакомить читателя с достигнутыми результатами в отношении возможностей по производству экспериментальной статистики из источников „больших данных“ для онлайн характеристики предприятий на европейском и национальном уровне.

**Ключевые слова:** характеристики предприятий на основе интернет-данных, экспериментальная статистика, большие данные, статистический Бизнес-регистр, Национальный статистический институт

\* Главен експерт в отдел „Информационни системи и приложен софтуер“, НСИ; e-mail: kgeorgiev@nsi.bg

\*\* Държавен експерт в дирекция „Обща методология, координация и анализ на статистическите изследвания“, НСИ; e-mail: gstateva@nsi.bg.

## PRODUCTION OF EXPERIMENTAL STATISTICS ON THE CHARACTERISTICS OF ENTERPRISES WITH INTERNET DATA

*Kostadin Georgiev\**, *Galya Stateva\*\**

**SUMMARY** With each passing year, online business is becoming even more important to the economy, and with the global pandemic, COVID it is more important than ever. The Work Package C (WPC) for online based enterprise characteristics (OBEC) within the framework of the European project ESSnet on Big data II is related to the understanding of the online economic and business activity of enterprises from the point of view of national statistics.

A key result of the work is the improvement of the quality of the statistical business register in terms of characteristics for the online presence of nationally registered companies such as the availability of websites, e-commerce or social media accounts. Within the WPC, the methodology of the previous ESSnet on Big data I project was summarized and extended for use in each ESS country, taking into account the diversity needed to maintain different use-cases in statistical practice.

The National Statistical Institute had the honour and privilege to be a leading institution and an active partner in the implementation of activities under work package C.

This article aims to acquaint the reader with the results achieved on the opportunities and challenges for the production of experimental statistics from “big data” sources on the online characteristics of enterprises at European and national level.

**Keywords:** online based enterprise characteristics, experimental statistics, big data, statistical Business Register, National Statistical Institute

---

\* Chief Expert in the Information Systems and Applied Software Department, NSI; e-mail: kgeorgiev@nsi.bg

\*\* State expert in the ‘General Methodology Directorate, Coordination and Analysis of Statistical Surveys’, NSI; e-mail: gstateva@nsi.bg





## МЕТОДИ И КОНЦЕПЦИИ ЗА ФУНКЦИОНАЛНО ГЕОГРАФСКО РАЙОНИРАНЕ ЗА СТАТИСТИЧЕСКИ ЦЕЛИ

*Валерия Ангелова\**



### **Въведение**

Необходимостта от функционално райониране за статистически и аналитични цели е безспорна. Широко използваната Класификация на териториалните единици за статистически цели (от френски *La nomenclature des unités territoriales statistiques (NUTS)*) е създадена въз основа на административните структури на държавите - членки на Европейския съюз (ЕС), и се основава на принципа за йерархичност и праг за брой население. Границите често са исторически унаследени и във все по-малка степен съвпадат с ареалите, където хората в действителност живеят и работят, и не отразяват териториалната структура на социалната и икономическата реалност в страната и Европа. Ежедневни трудови пътувания от един NUTS регион в друг или отвъд държавна граница могат да доведат до значителни разлики между общата заетост (основана на работното място) и заетостта на работещото население, което живее в същия регион. Икономически показатели като БВП на жител или показатели за околната среда като

---

\* Държавен експерт в отдел „Регионална статистика и индикатори за мониторинг“, Национален статистически институт; e-mail: VAngelova@nsi.bg.

количество на генерираните отпадъци на жител са изкривени в региони с асиметрични модели на ежедневни трудови миграции. Люксембург, NUTS 3 регионите на Вътрешен Лондон, Брюксел и София (столица) са само някои от примерите за територии, където данните за заетостта, БВП и показателите за околната среда са изкривени, когато са представени спрямо броя на жителите на региона. Използването на концепции, основани на функционален критерий като модели на поведение по отношение на ежедневните трудови миграции, може да предотврати някои потенциални погрешни интерпретации на данни на ниво NUTS 3.

В България границите на NUTS регионите не са актуализирани от 2006 година. Според Регламента за NUTS границите на регионите биха могли да се актуализират на всеки три години, но правният документ не предвижда функционални принципи, които да се вземат предвид при евентуално прерайониране<sup>1</sup>.

Представеният анализ е инспириран от скорошни публикации на Евростат<sup>2</sup> и на Националния статистически институт (НСИ), посветени на районите на пазара на труда (РПТ)<sup>3</sup>. През периода 2016 - 2017 г. статистическите институти на няколко държави членки, между които и България, активно работят по проект за развитие и внедряване на концепцията за РПТ в европейската статистическа практика. Работата по този проект е в изпълнение на договор за субсидия с Европейската комисия, като Евростат координира дейностите. Наред с това вече е налице действаща нормативна база за определяне на редица териториални типологии за статистически цели. Т.нар. Регламент TERCET, който допълва Регламента за NUTS, е публикуван през декември 2017 година<sup>4</sup>. Посредством този законодателен акт различни териториални типологии, като тази на градските и селските райони и региони и на метрополните региони, степента на урбанизация, градовете и техните функционални градски райони (ФГР) (от английски - Functional Urban Areas (FUAs), са включени в Регламента за NUTS. Регламентът за изпълнение на TERCET, публикуван през 2019 г., задава единните условия за хармонизирано прилагане на териториалните типологии<sup>5</sup>. Наличието на хармонизирани дефиниции гарантира

<sup>1</sup> <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:02003R1059-20180118&qid=151913675347>

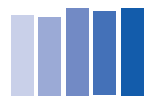
<sup>2</sup> <https://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/ks-tc-20-002>

<sup>3</sup>

<https://www.nsi.bg/bg/content/18760/%D0%BF%D1%83%D0%B1%D0%BB%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F/%D1%80%D0%B0%D0%B9%D0%BE%D0%BD%D0%B8-%D0%BD%D0%B0-%D0%BF%D0%B0%D0%B7%D0%B0%D1%80%D0%B0-%D0%BD%D0%B0-%D1%82%D1%80%D1%83%D0%B4%D0%B0>

<sup>4</sup> <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32017R2391>

<sup>5</sup> <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R1130>



устойчивост на резултатите и международна сравнимост на данните. От 2004 г. насам НСИ работи по проект със съфинансиране от Европейската комисия - „Статистика на европейските градове“, като от 2010 г. използва като основа за докладването на статистическите данни европейските хармонизирани дефиниции за град и ФГР.

Целта на настоящото изследване е да се анализират практическите предимства на различни статистически подходи за функционално географско райониране, основани на данни за ежедневните трудови пътувания, и да се представи обобщение на методите по достъпен начин с помощта на редица примери. Акцентът е поставен върху функционалните градски райони и районите на пазара на труда. Представени са и редица предизвикателства, свързани с дефинирането и използването на тези функционални райони за статистически и аналитични цели, и са очертани възможни насоки за бъдещето, като се отчитат промените в моделите на трудова мобилност и в условията на пазара на труда в резултат на пандемията от COVID-19.

### **1. Принципи за дефиниране на функционални градски райони и райони на пазара на труда**

**В Регламента TERSET функционалните градски райони** са дефинирани като комбинация от градовете и зоните около тях, от които произтичат най-интензивните потоци от ежедневни трудови мигранти към съответните градове. Иначе казано, ФГР се базират на централен тип взаимодействие между града и неговия хинтерланд, тъй като трудовите миграции се наблюдават по посока към централното селище.

Първото необходимо условие за очертаването на ФГР е дефинирането на градовете според правилата на хармонизираната европейска методология. Градовете са една от категориите в т.нар. степен на урбанизация. Тази типология класифицира местните административни единици, в случая на България - общини (LAU), като<sup>6</sup>:

- **Градове или гъсто населени райони**

Това са общини, в които поне 50% от населението живее в градски центрове. Градските центрове са съседни (без диагонални) клетки от един квадратен километър в рамките на „кълъстера от градски тип“ с гъстота на населението от поне 1 500 жители на квадратен километър и с най-малко 50 000 жители в кълъстера след попълване на

---

<sup>6</sup> Авторът на това изследване се придържа към официалния български превод на терминологията в Регламента TERSET.

празнините. Празнините са клетките, попадащи в границите на клъстера, но неотговарящи на условието за гъстота на населението.

- **По-малки градове и предградия или средно населени райони**

Това са общини, в които по-малко от 50% от населението живее в клетки от селски тип и по-малко от 50% от населението живее в градски центрове.

- **Селски райони или слабо населени райони**

Това са общините, в които поне 50% от населението живее в клетки от селски тип. Клетките от селски тип са клетките с площ един квадратен километър с население под 300 жители или с население над 300 жители, но неотговарящи на условията, необходими, за да бъдат включени в клъстерите от градски тип.

Според класификацията в България има 18 града, 134 малки града и предградия, а останалите общини попадат в категорията селски райони. 18-те града (сортирани по брой население към 31.12.2019 г.) са: София, Пловдив, Варна, Бургас, Русе, Стара Загора, Плевен, Сливен, Добрич, Шумен, Перник, Хасково, Благоевград, Велико Търново, Пазарджик, Ямбол, Враца и Видин. Паралелно НСИ поддържа класификацията на степента на урбанизация и на ниво населено място. Тя се използва изключително за производство на данни от социални изследвания с разбивка по степен на урбанизация.

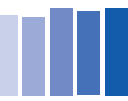
След като са дефинирани границите на градовете, следващият етап е определянето на зоните на регулярно пътуване до работното място в града в три стъпки:

**Стъпка 1:** Проверка за т.нар. свързани градове. В случаите, когато 15% от заетите жители на един град работят в друг град, тези градове се считат за свързани и имат общ ФГР. Не е необходимо свързаните градове да са съседни, т.е. да споделят обща граница.

**Стъпка 2:** Идентифициране на всички LAU, където най-малко 15% от заетите жители работят в прилежащия град.

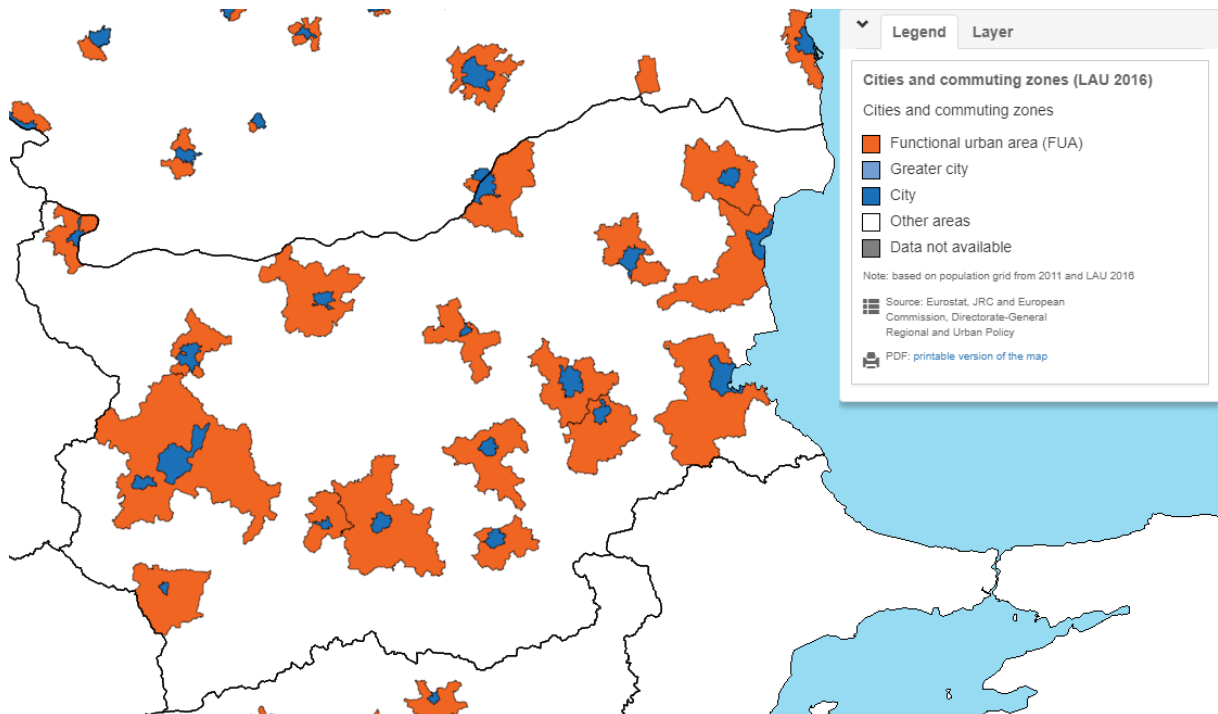
**Стъпка 3:** Проверка за анклави и ексклави. LAU, които не отговарят на условието в стъпка 2, но са заобиколени от такива, които отговарят, се включват в съответния ФГР, а тези, които нямат обща граница с потенциалния ФГР, отпадат.

На фиг. 1 са представени градовете и ФГР в България, произведени съвместно от НСИ и Евростат въз основа на границите на LAU от 2016 г. и последния официален грид



на населението от Преброяване 2011. София и Перник попадат в категорията на т.нар. свързани градове и споделят общ функционален градски район.

**Фиг. 1. Градовете в България и техните ФГР**



Източник: Статистически атлас на Евростат<sup>7</sup>.

ФГР от своя страна служат за дефиниране на още една териториална типология - тази на метрополните региони. Метрополен регион може да бъде един регион на ниво 3 по NUTS или съвкупност от региони на ниво 3 по NUTS, където поне 50% от населението живее в прилежащия функционален градски район, който от своя страна трябва да има най-малко 250 000 жители. С други думи, метрополните региони представляват пространствената NUTS 3 алтернатива на ФГР.

**Районите на пазара на труда (РПТ)**, от английски Labour Market Area (LMA), е утвърдена и широко обсъждана концепция в регионалната география и статистика както в Европа, така и в някои страни отвъд океана като САЩ, Канада, Мексико и Южна Корея<sup>8</sup>. Историята на концепцията за районите на пазара на труда, както и методът за тяхното дефиниране, са подробно описани в посочените публикации на Евростат и НСИ,

<sup>7</sup> <https://ec.europa.eu/eurostat/statistical-atlas/gis/viewer/?config=typologies.json&ch=TYPLOC,CITYCOMMZONE&mid=BKGCNT,BKGNT02013,CITYCOMMZONE2018,CNTOVL&o=1,1,1,0.7&center=42.65322,26.0525,6&lcis=CITYCOMMZONE2018>  
<sup>8</sup> <https://www.oecd.org/publications/delineating-functional-areas-in-all-territories-07970966-en.htm>

затова тук няма да навлизам в детайли, а ще се спра само на основополагащите принципи на метода за този тип функционално райониране.

Според определението, възприето от Евростат и НСИ, РПТ е статистически дефинирана, икономически интегрирана територия, където мнозинството от хората живеят и работят. В резултат на дългогодишната съвместна работа на Евростат, статистическите офиси на държавите членки и академичната общност понастоящем разполагаме с публично достъпен и отворен за подобряване ИТ инструментариум за производство на РПТ под формата на R пакет<sup>9</sup>. Алгоритъмът изисква единствено матрицата на ежедневните трудови миграции и броя на заетите като входни данни. Той е итеративен и работи на принципа „от долу нагоре“ с набор от четири параметъра (минимален и целеви брой заети и минимално и целево ниво на независимост по отношение на работна сила в рамките на самостоятелен РПТ). Четирите параметъра задават ограничителните условия, които дават основание на изследователя да определи коя територия се счита за действителен РПТ. Най-общо казано, алгоритъм изчислява клъстери от съседни градивни компоненти (в случая на България - общини, но в практиката са известни резултати на база NUTS 3 региони, пощенски зони и дори на GRID клетки). Клъстеризацията продължава дотогава, докато проектните РПТ изпълнят следното условие:

$$\frac{\min NRS}{\text{cel}NRS} \leq \left( 1 - \left( 1 - \frac{\min NRS}{\text{cel}NRS} \right) \cdot \text{MAX} \left( \frac{\text{cel}BZ - BZ}{\text{cel}BZ - \min BZ}, 0 \right) \right) \cdot \left( \frac{\text{MIN}(NRS, \text{cel}NRS)}{\text{cel}NRS} \right),$$

където:

BZ е броят на заетите;

NRS е независимостта по отношение на работната сила;

minNRS е минималната независимост по отношение на работната сила;

celNRS е целевата независимост по отношение на работната сила;

minBZ е минималният брой заети;

celBZ е целевият брой заети.

Това условие за валидност се основава на принцип за вътрешна хомогенност на резултативния РПТ и хетерогенност спрямо други РПТ. Наличието на функционалните връзки в рамките на всеки РПТ се потвърждава от интензитета на ежедневните трудови пътувания. В алгоритъма са заложили стойности на четирите параметъра по

<sup>9</sup> R пакет може да бъде изтеглен на следния адрес: <https://CRAN.R-project.org/package=LabourMarketAreas>



подразбиране, които изследователят може да променя ръчно въз основа на опит, познания за територията и ясно поставена изследователска задача относно целевия мащаб на РПТ. Стойностите по подразбиране в алгоритъма за параметъра  $selNSR$  са между 0.75 и 0.8. За параметъра  $minNSR$  стойностите по подразбиране са между 0.6 и 0.6667. Значението на тези стойности е следното. Независимост по отношение на работна сила от 0.8 или 80% означава, че само един от всеки петима заети жители работи извън проектния РПТ, 0.75 означава, че трима от всеки четирима жители, които са заети, работят в границите на РПТ, 0.6667 съответства на двама от трима, 0.6 - на трима от петима и т.н.

Независимостта по отношение на работната сила се измерва както във връзка с търсенето, така и във връзка с предлагането на работна сила. Независимостта по отношение на работна сила от страна на предлагането се изразява в броя на хората, които живеят и работят в даден район, разделен на броя на всички заети жители на района, независимо от това къде е тяхната месторабота. Независимостта по отношение на работна сила от страна на търсенето се изразява в броя на хората, които живеят и работят в даден район, разделен на броя на всички заети работни места в същия район (заети от жители на района или от жители на други райони).

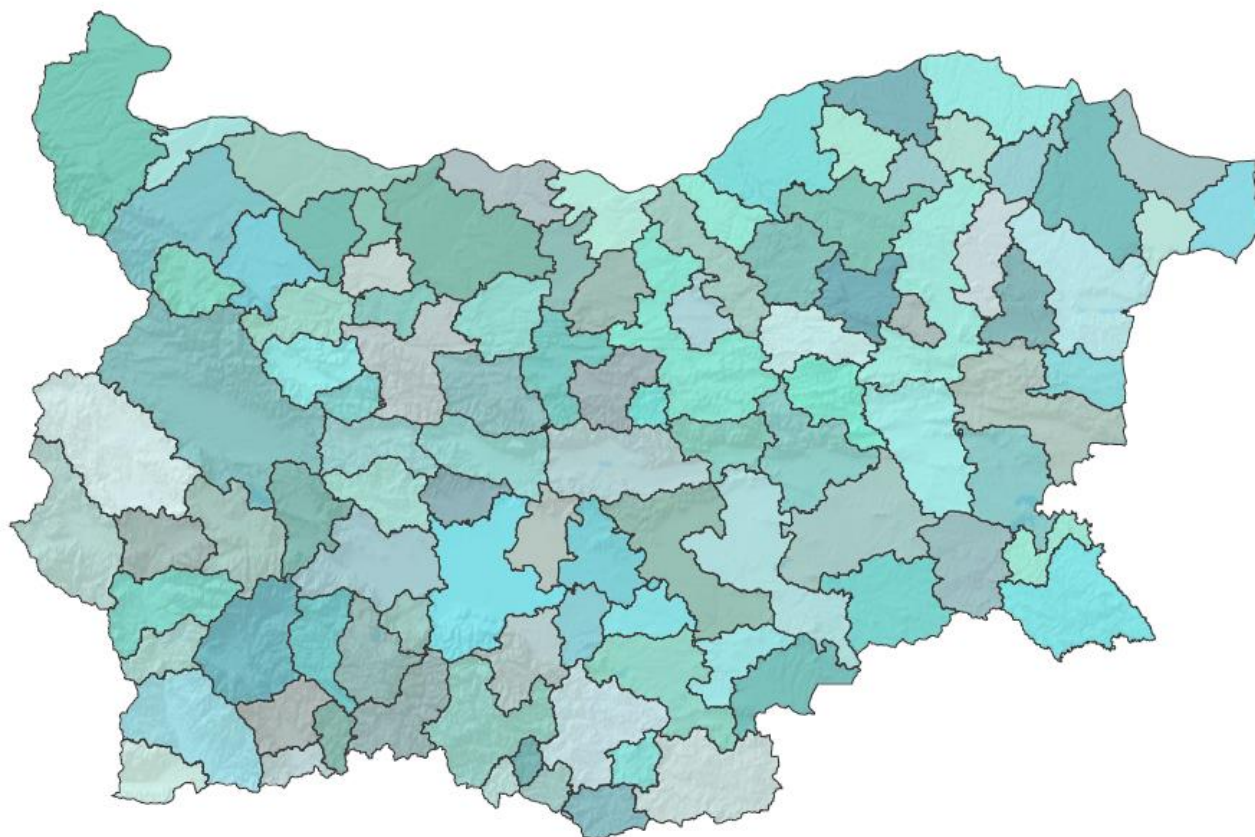
На фиг. 2 са представени РПТ в България, получени въз основа на данни за ежедневните трудови пътувания от Преброяване 2011. Предстои актуализация на границите на РПТ и ФГР с данни от Преброяване 2021, като 2023 г. се счита за реалистичен срок. Практиката в скандинавските страни е границите на РПТ и на ФГР да се актуализират ежегодно на основата на данни за ежедневните трудови пътувания (ЕТП) от регистри. Известни са и опити за приложение на данни от мобилните оператори не само за идентифициране на зоните за регулярни пътувания до работното място, но и за определяне на почасовото териториално разпределение на работещите<sup>10,11</sup>. Предвид нарастващия интерес на политиците към паневропейски грид на заетостта, както и към данни за мобилността с висока резолюция, включително и по цел на пътуването, РПТ, изглежда, имат огромен потенциал във връзка с използването на данни от мобилни телефони и/или интегриране на данни от административни източници с големи данни.

<sup>10</sup> [https://ec.europa.eu/regional\\_policy/en/newsroom/news/2018/11/11-06-2018-enhancing-border-regions-data-collection-final-report-of-a-pilot-project](https://ec.europa.eu/regional_policy/en/newsroom/news/2018/11/11-06-2018-enhancing-border-regions-data-collection-final-report-of-a-pilot-project)

<sup>11</sup> <https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-HA-17-001> (с. 242 и 243).



Фиг. 2. РПТ в България



Източник: НСИ.

Получените въз основа на метода РПТ в България са 103 на брой. Границите на РПТ са получени в резултат от използването на следните параметри: минимален размер от 3 500 заети, целеви размер от 25 000 заети, минимална независимост по отношение на работната сила от 66.67% и целева независимост по отношение на работната сила от 75%. Резултативните райони могат да бъдат интерпретирани като преходно териториално ниво между общини и NUTS 3 региони, но теоретично обосновано на базата на доказателства за съществуването на силни функционални връзки.

Най-малкият РПТ е този на Рудозем с малко над 3 000 заети лица, а най-големият е този на София с близо 700 000 заети лица. Предвид специфичния релеф и конфигурацията на транспортната мрежа в страната и при така избраните стойности на параметрите оценяваме получените конфигурации на РПТ като логични. Табл. 1 съдържа някои генерални характеристики на РПТ в България.



## 1. Основни характеристики на РПТ в България

	Брой LAU в състава на РПТ	Площ (км <sup>2</sup> )	Брой заети жители	Общ брой заети работни места	NRS по отношение на предлагането (%)	NRS по отношение на търсенето (%)
<b>Минимум</b>	1	174.8	3532	3091	77.0	75.6
<b>Медиана</b>	2	907.9	10600	10026	91.8	95.3
<b>Средно аритметично</b>	2.5	1071.6	27505	27505	91.0	94.6
<b>Максимум</b>	11	4210.4	664430	695137	99.1	99.5

Източник: Информация на НСИ, докладвана пред Евростат във връзка с финалните резултати от проекта, реализиран през 2016 - 2017 година.<sup>12</sup>

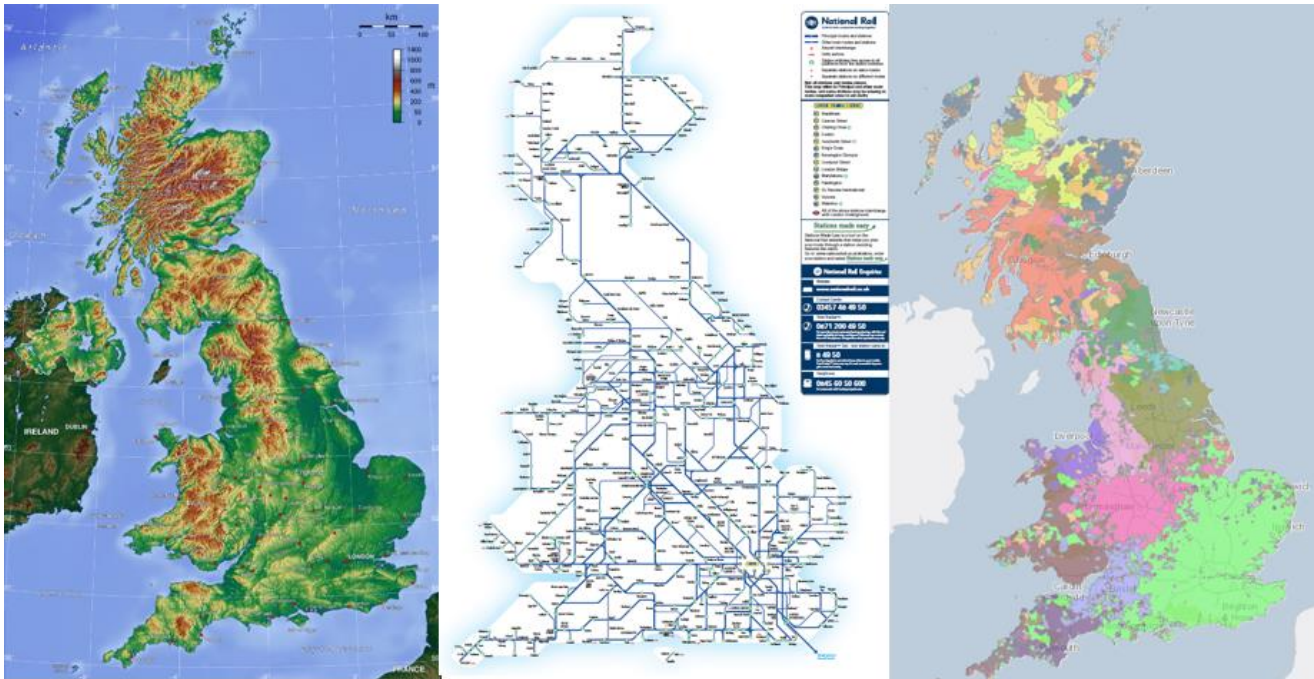
За да потвърдим становището, че методът за очертаване на РПТ дава резултати, които отразяват икономическите връзки в териториален аспект, ще анализираме примера на Обединеното кралство<sup>13</sup>. Същевременно ще разкрием един важен аспект, свързан с РПТ, а именно широката им приложимост и аналитичния им потенциал. Като цяло РПТ представляват резултат от „усредняване“ на моделите на пътуване до работа на различни полове, социално-икономически и професионални групи и транспортните средства, използвани за пътуване до работното място. Освен за очертаване на РПТ за общите потоци от ЕТМ методът е подходящ и за дефиниране на алтернативни, дезагрегирани или дори имагинерни РПТ, например по пол, образование, РПТ за пътуващите с автобус, влак, автомобил и т.н.

Изображенията на фиг. Знапомнят три пръстови отпечатъка. Железопътният транспорт в Обединеното кралство е широко използван за пътуване до работното място поради достъпната му цена. Резултатът от наличието на гъста и високоскоростна жп мрежа са големи по площ РПТ за пътуващите с този вид транспорт.

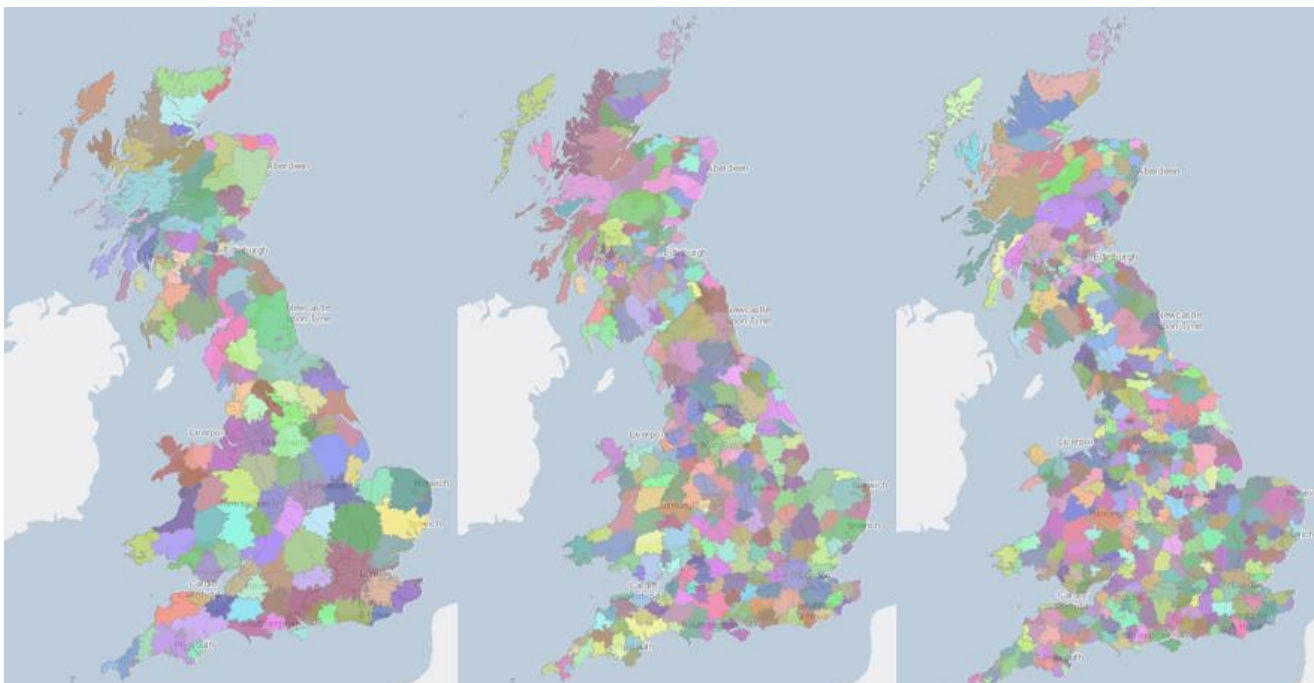
<sup>12</sup> [https://ec.europa.eu/eurostat/cros/content/results-and-future-plans-bulgaria\\_en](https://ec.europa.eu/eurostat/cros/content/results-and-future-plans-bulgaria_en)

<sup>13</sup> <http://ons.maps.arcgis.com/apps/MapSeries/index.html?appid=397ccae5d5c7472e87cf0ca766386cc2>

**Фиг. 3. Релеф, железопътна мрежа и РПТ за пътуващите до работното място с железопътен транспорт в Обединеното кралство**



**Фиг. 4. РПТ за висококвалифицираните (1), средноквалифицираните (2) и нискоквалифицираните работници (3)**

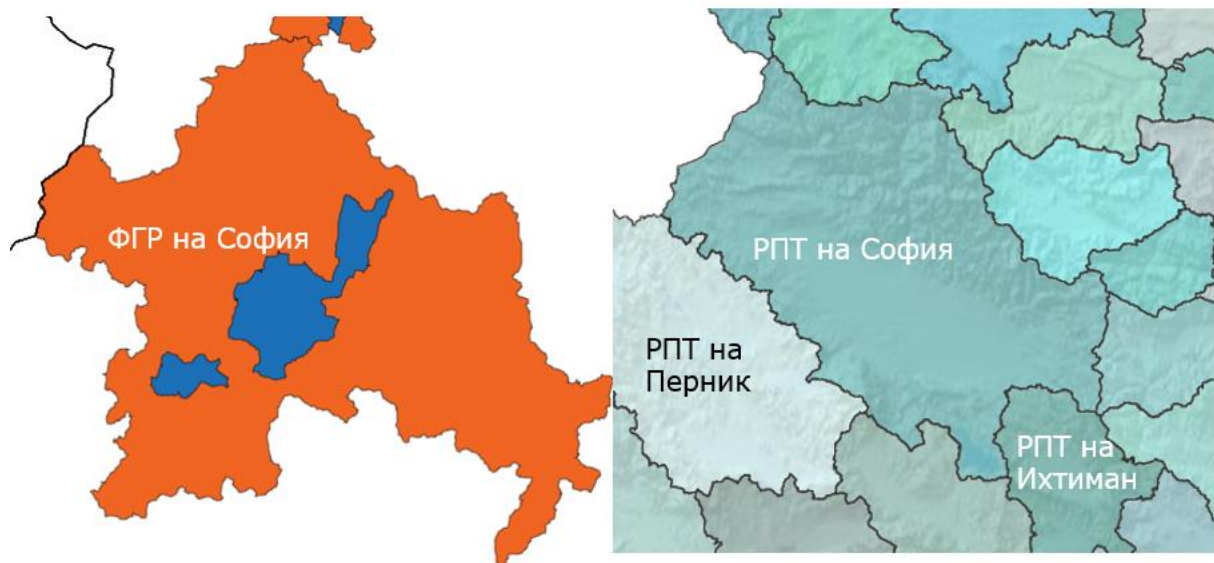


Хората с по-висока квалификация са склонни да пътуват на по-дълги разстояния, за да заемат желаната от тях работа, което логично има за резултат по-големи РПТ в сравнение с тези за средно- и нискоквалифицираните заети лица. Различията в размерите

на този вид алтернативни РПТ в Шотландия не са така забележими както в Англия и Уелс. Това вероятно се дължи на различията в достъпа на населението до жп транспорт, цени, национални особености, както и други фактори, за които не разполагаме с информация.

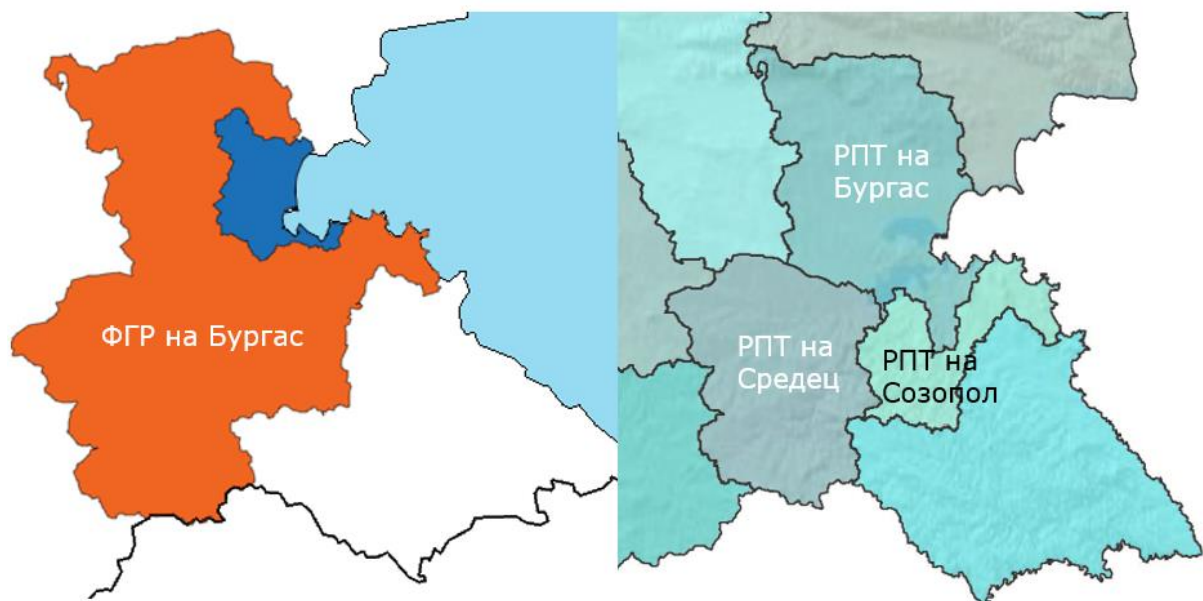
За да обвържем двата типа функционални райони, следва да направим сравнителен анализ. РПТ и ФГР често биват използвани в един контекст, но следва да се отбележи, че РПТ не са проектирани с цел да заменят ФГР. Те се основават на различна логика и са предвидени за различни цели. ФГР са един от елементите на статистиката на градовете, а РПТ са формулирани за целите на оценката и докладването на данни за работната сила и сродни теми. ФГР се базират на миграционни потоци към града и обхващат ограничена територия около градовете, докато РПТ имат за цел да обхванат цялата територия на страната, както и територията на ЕС, и се основават на анализ на всички потоци от ежедневни трудови пътувания между всички LAU или други градивни компоненти. ФГР се дефинират на три сравнително лишени от комплексност стъпки, за изпълнението на които не е необходим специализиран софтуер. За очертаване на границите на РПТ с помощта на ИТ инструмент се изпълнява итеративен алгоритъм с четири параметъра. ФГР обхващат сравнително големи територии, съпоставими с обхвата на NUTS 3 регионите, докато някои островни и изолирани планински РПТ са малки по размер, което затруднява производството на данни и често единственото решение е прилагането на методи за Small Area Estimations.

При сравнение на конфигурациите на ФГР и РПТ в България се установява, че макар и двата вида функционални райони да пресичат административни граници, ФГР са значително по-големи по територия от РПТ. Например ФГР на гр. София е съставен от 12 общини (Божурище, Горна Малина, Драгоман, Елин Пелин, Ихтиман, Костенец, Костинброд, Перник, Радомир, Своге, Сливница и Столична община), докато РПТ, в който попада столичният град, е изграден от 9 общини. В състава на столичния РПТ влиза и община Годеч, докато Перник, Радомир, Ихтиман и Костенец са компоненти на отделно обособени РПТ, съответно тези на Перник и Ихтиман. Това разбираемо се дължи на различните типове взаимодействия, които се изследват с цел дефиниране на двата вида функционални райони - централен тип взаимодействие при формиране на ФГР (трудовете миграции се наблюдават по посока към централното селище) и многопосочни взаимодействия без конкретни ограничения (не е необходимо наличие на централно селище).

**Фиг. 5. ФГР и РПТ в района на София**

Източник: Евростат и НСИ.

Подобна картина се наблюдава и в района на Бургас. Например общините Созопол и Средец участват в конфигурацията на ФГР на Бургас, но са самостоятелни РПТ, съставени единствено от едноименната община. Средец е сравнително малък и изолиран РПТ без силни функционални връзки с други общини по отношение на ежедневните трудови пътувания, докато конфигурацията на РПТ на Созопол вероятно се дължи на притегателната сила на туризма въпреки сезонния му характер.

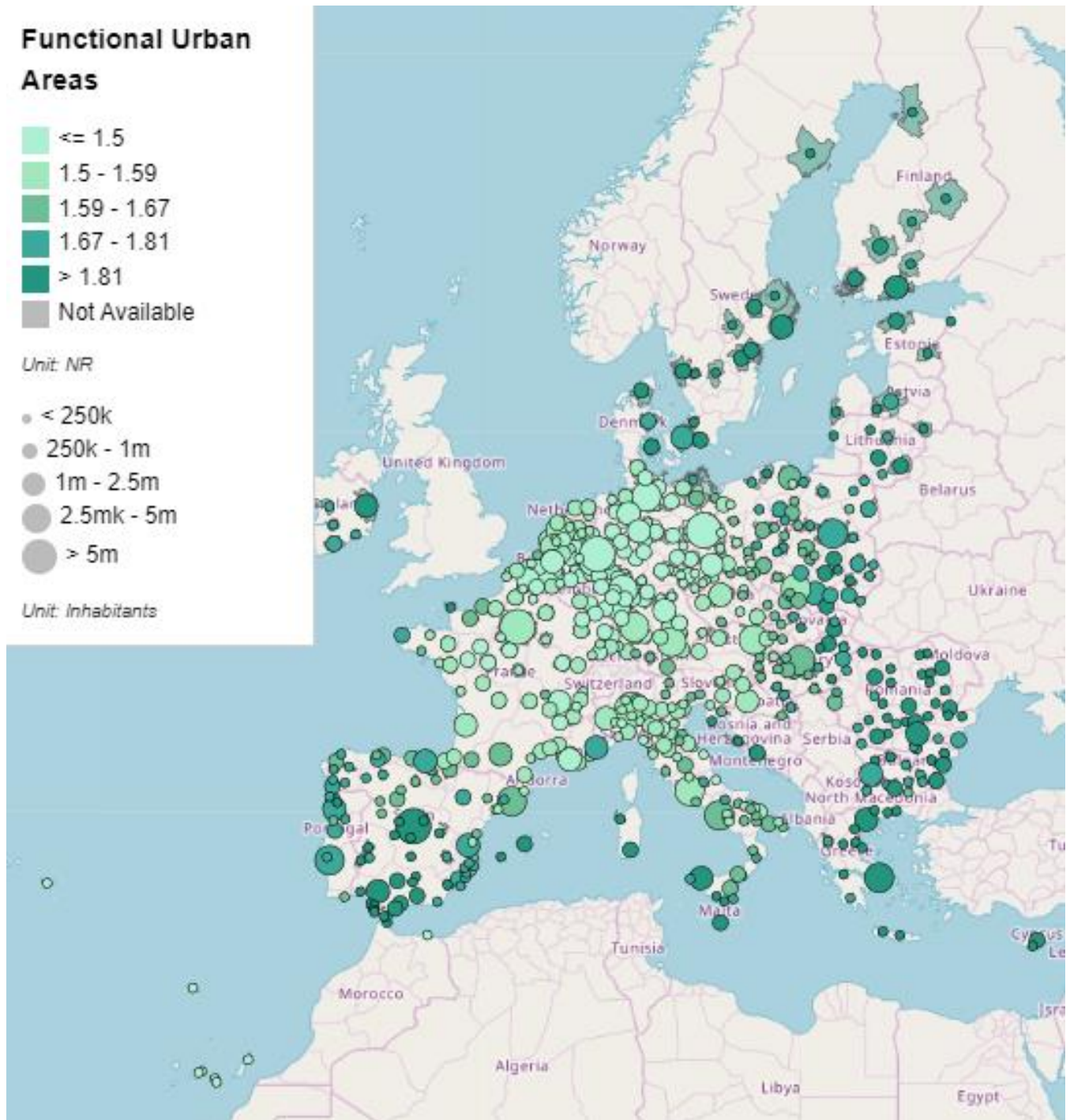
**Фиг. 6. ФГР и РПТ в района на Бургас**

Източник: Евростат и НСИ.



На фиг. 7 и фиг. 8 са показани примери за използването на ФГР и РПТ за производство и представяне на статистически данни.

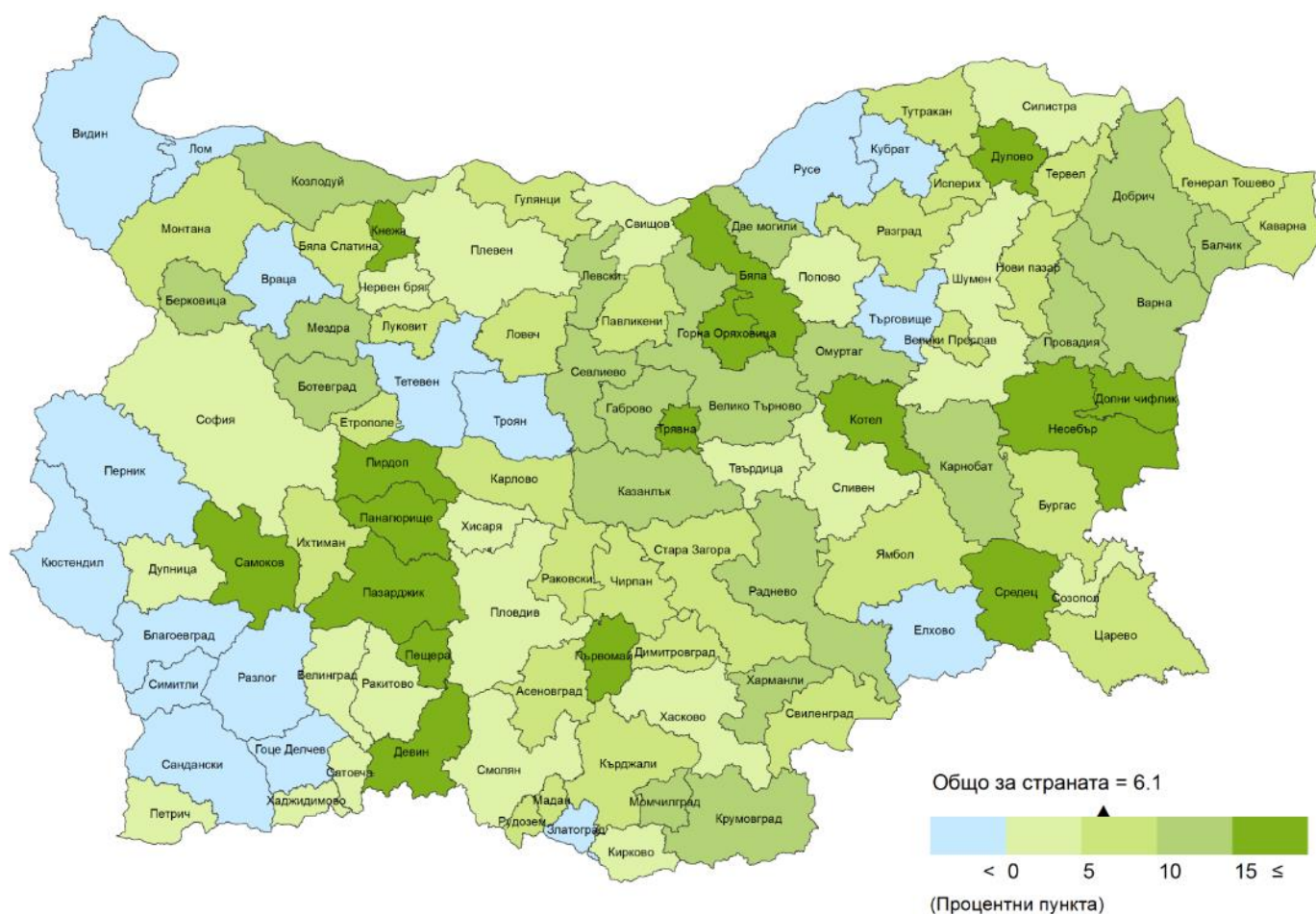
**Фиг. 7. Ефективност на транспортната мрежа във ФГР на Европа  
(Прогнозни данни към 2050 година)**



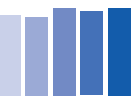
Източник: JRC, DG REGIO - Urban Data Platform Plus<sup>14</sup>.

Като част от проекта, осъществен през 2016 - 2017 г., НСИ произведе някои експериментални данни на ниво РПТ. За да стане възможно използването на информацията от Наблюдението на работната сила, оригиналните тегла на извадката са допълнително калибрирани, за да се осигури съпоставимост с данни за населението по съответните възрастови групи на ниво РПТ. Фиг. 8 дава представа за това какви данни могат да се произведат на ниво РПТ дори без приложение на класическите методи за получаване на данни за малки териториални единици (Small Area Estimations).

**Фиг. 8. Равенство между половете (мъже - жени) по отношение на заетостта в РПТ в България, 2015 г. (като процент от населението на възраст 15 - 64 години)<sup>15</sup>**



Източник: НСИ.



## 2. Предизвикателства при дефинирането на „функционални градски райони“ и „райони на пазара на труда“

### 2.1. Функционални градски райони

По време на дефинирането на използваните понастоящем ФГР на градовете с помощта на данни за ЕТП от Преброяване 2011 в НСИ установихме проявата на т.нар. **парадокс на малките градивни компоненти**. Първоначалната идея беше да се използват данни с по-голям географски детайл, а именно матрицата на ЕТП на ниво населено място. Анализът на получените резултати показва, че в случая използването на входни данни на ниво населено място не е подходящо поради наличието на твърде малки населени места, както и на значителна вариация в броя на населението и броя на заетите. Например според правилата за дефиниране на ФГР населено място със 100 жители и 50 заети, от които 10 работят в съседния град, следва да влезе в състава на съответния ФГР. Заключение ни се потвърди и от значителния брой на получените анклави и ексклави. По този начин за финалните конфигурации на ФГР в България използвахме матрицата на ЕТП на ниво община.

При дефинирането на ФГР и по-специално при определяне на границите на ФГР на т.нар. **свързани градове** установихме следната тънкость, която не е добре документирана в достъпните ни литературни източници. Често срещана заблуда е, че при установяване на наличие на свързани градове потоците от ежедневни трудови мигранти трябва да се изследват само по посока на доминиращия град, като вторичният градски център автоматично се прибавя към границите на ФГР. Правилният подход би бил двата града да се разглеждат като едно цяло и интензитетът на миграционните потоци да се анализира по посока на този въображаем свързан град. В противен случай обхватът на ФГР би бил подценен<sup>16</sup>.

Наличието на свързани градове е свидетелство за **полицентрично градско развитие**. В една още по-комплексна ситуация можем да наблюдаваме три града, които са потенциално свързани въз основа на функционалния критерий „интензитет на ежедневните трудови пътувания“. Тогава подходът е следният. Ако най-малко 15% от заетите жители на град А и най-малко 15% от заетите жители на град Б работят в град В, то трите града биха имали общ ФГР. Ако най-малко 15% от заетите, живеещи в град А,

<sup>16</sup> Авторски наблюдения в резултат на съвместна работа с експерти по регионална статистика от други държави членки.

работят в град Б, а останалите потоци между градовете А и В и Б и В се равняват на по-малко от 15%, то градовете А и Б ще имат общ ФГР, докато град В ще има индивидуален ФГР<sup>17</sup>.

Както вече беше отбелязано, ФГР не покриват цялата територия на страната, респективно на Европа, а включват само онези райони, за които е очевидно, че икономически гравитират към съответния град. Данни на ниво ФГР, както и данните за метрополните региони, често се използват за анализ на демографското и социално-икономическото развитие на гъсто населени територии, тъй като в техните граници по правило има град. Интерпретацията на данните следва да се извършва внимателно и да се има предвид, че в състава на ФГР и метрополните региони влизат също така и села и по-малки градове и предградия, от които произлизат ежедневните трудови миграции към централното селище. Любопитен факт е, че терминът, използван на немски език за зоната на ежедневни трудови пътувания, е „Speckgürtel“, т.е. зоната, която захранва града. От друга страна, в случаите, когато градът няма обособена зона на регулярно пътуване до работното място (няма съседна община, която да отговаря на описаните условия), границите на града и ФГР съвпадат. В България такива са Велико Търново, Враца, Пазарджик, Сливен и Хасково.

Считам, че изброените особености не представляват ограничителни условия при използването на типологията на ФГР за статистически и аналитични цели, но следва да се вземат предвид.

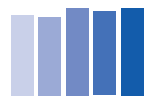
## 2.2. Райони на пазара на труда

В този случай предизвикателствата от теоретично и методологично естество са по-сложни, което се дължи на далеч по-комплексния характер на метода за дефиниране на РПТ в сравнение с подхода за определяне на границите на ФГР.

**Изборът на градивни компоненти** има комплексно влияние върху размера и характеристиките на резултативните РПТ и е свързан с вземането на решение относно целевия мащаб на РПТ. Решението относно градивните компоненти на първо място зависи от наличието на входни данни на съответното географско ниво. Така например в Германия се използват входни данни за ЕТМ на ниво NUTS 3 региони, в България и

<sup>17</sup> Авторска интерпретация на текст от Методологичния наръчник на Евростат от 2018 г., посветен на териториалните типологии, с. 52, <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-008>





Португалия - на ниво общини, а в редица страни като Полша, Унгария, Италия и Швейцария се борави с данни на ниво населено място. Финландия експериментира с данни на ниво пощенски зони, а Италия - с данни на ниво грид със страна един километър.

Грид клетките със страна един километър са устойчиви във времето и не зависят от промените в административните и административно-териториалните граници. Следва да се има предвид, че при използването на входни данни на ниво грид изпълнението на алгоритъма ще отнеме значително повече време в сравнение с използването на по-големи и по-малко на брой градивни компоненти. Изследвателят ще се сблъска с клетки с нулев или близък до нулата брой заети работни места, работещи жители или с незначителни потоци от ежедневни трудови пътувания. Редица клетки са покрити от природни обекти като езера, големи реки, глетчери, високопланински територии и други. В тези случаи очакваният резултат би бил голям брой клетки в списъка с резервни компоненти, които могат да бъдат само фиктивно причислени към прилежащ РПТ, чиито компоненти са удовлетворили условието за валидност. Въпреки съществуващите ограничения считаме грида за универсален градивен компонент за бъдещо дефиниране на РПТ.

Освен че РПТ, дефинирани въз основа на входни данни на ниво грид, ще бъдат по-прецизни в сравнение с конфигурации, получени с по-големи градивни компоненти, данните на ниво клетка сами по себе си могат да бъдат агрегирани до всякакви функционални райони, което отваря нови възможности в статистическата наука и практика. Наличието на координатите на сградите за целите на Преброяване 2021 означава наличие на координатите на адресите на респондентите, включени в социални изследвания като Наблюдението на работната сила, Изследването на доходите и условията на живот, Изследването на домакинските бюджети, Европейското здравно интервю и редица други. Наличието на географските координати на наблюденията или идентификационния номер на грид клетката, в която е регистрирано наблюдението, предлага неограничен ресурс за преизчисление на динамичните редове до всевъзможни територии и функционални райони от потребителски/изследователски интерес, чиито конфигурации не са непременно имплементирани в дизайна на съответната извадка, както и за интегрирането на данните от различни социални изследвания въз основа на местоположението като ключ. Този ключ би осигурил и независимост на производствения процес от евентуални промени в границите на NUTS и още по-често случващите се промени в границите на LAU. Не на последно място гридът може да

намери приложение в оптимизирането на извадките посредством по-рационално покритие на територията на страната и по-добър обхват на население с различни характеристики.

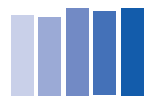
Другото основно предизвикателство е свързано с избора на **стойности на параметрите**. Проведените през годините анализи на чувствителността на параметрите показват, че резултатите за различни държави реагират с различна степен на чувствителност на промяната в стойностите на параметрите<sup>18</sup>. Например броят на РПТ в Унгария значително се увеличава при залагане на минималния брой заети на 1 000 души, докато в други страни броят на РПТ остава доста стабилен при този вариант. Резултатите за РПТ в Нидерландия реагират чувствително на увеличението на минималната независимост по отношение на работната сила, докато РПТ в Швеция - на увеличаването на целевата независимост по отношение на работна сила. НСИ също проведе серия от тестове, преди да се спре на стойностите на параметрите, поместени по-горе.

През периода 2018 - 2019 г. в рамките на специализирана работна група, организирана от Евростат, държавите членки с опит в дефинирането на РПТ широко дискутират възможностите за въвеждане на две йерархични нива на РПТ (по подобие на йерархичната структура на NUTS)<sup>19</sup>. Споделям идеята, че национални РПТ, произведени на базата на хармонизирания метод и критерии, заложили според националните нужди, са също толкова необходими, колкото и РПТ в европейски мащаб. Националните РПТ са обикновено по-малки по отношение на броя на заетите лица. Разбираемо, европейските заинтересовани страни се нуждаят от сравними РПТ в цяла Европа. Те следва да са по-големи, което наред с един съдържателен сравнителен анализ в рамките на Европа ще улесни и изготвянето на статистически данни за заетостта, както и други показатели от социални изследвания. Според мен подходът на Португалия, адресиращ конкретната специфика на пространственото развитие на страната, дава възможно решение относно подходящ набор от параметри за производство на РПТ в европейски мащаб.

Наличието на асиметрия в селищното развитие на Португалия (север - юг; крайбрежна - континентална част) затруднява дефинирането на функционални райони, хомогенни от гледна точка на население, заети лица и площ. Двата най-големи метрополни региона (Лисабон и Порто) имат голям капацитет за привличане на

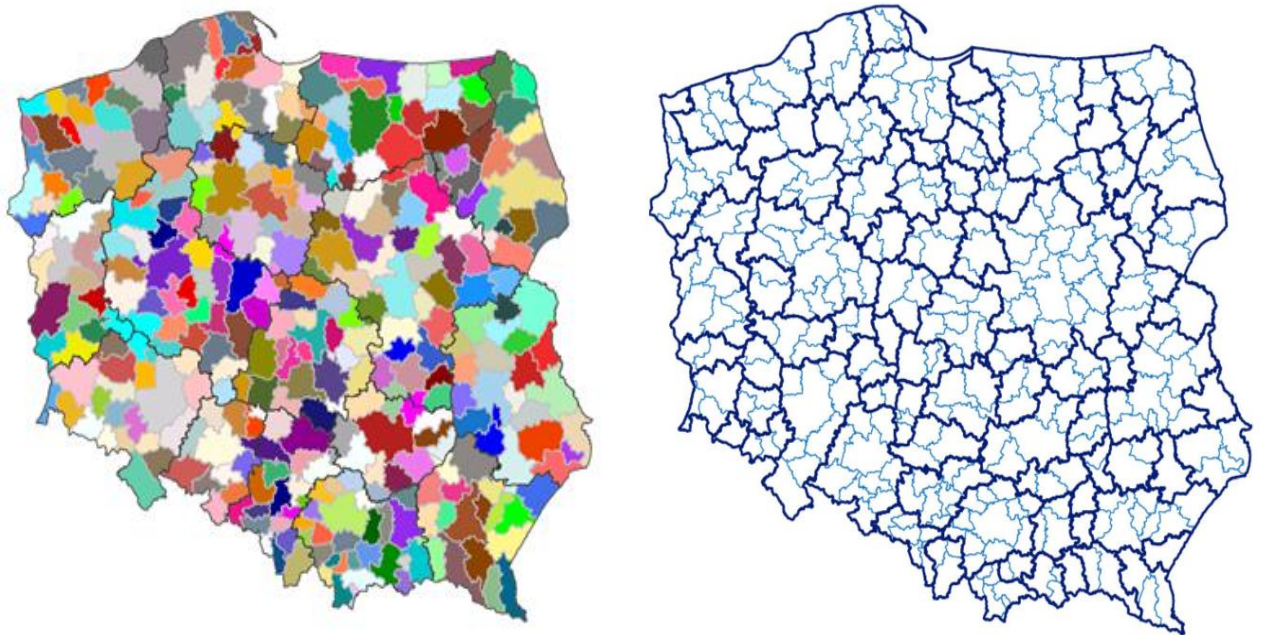
<sup>18</sup> [https://ec.europa.eu/eurostat/cros/content/task-force-lmas-final-report\\_en](https://ec.europa.eu/eurostat/cros/content/task-force-lmas-final-report_en)

<sup>19</sup> [https://ec.europa.eu/eurostat/cros/content/conclusions-tf-lmas-european-scale-eurostat\\_en](https://ec.europa.eu/eurostat/cros/content/conclusions-tf-lmas-european-scale-eurostat_en)



пътуващите от други общини (40% от португалското население е заето в тях). С цел преодоляване на тези проблеми статистиците от Португалия са избрали да работят с високи прагове на независимост по отношение на работна сила и високи стойности на минимален и целеви брой на заетите (minNRS 0.80, celNRS 0.85, minBZ 35 000, celBZ 100 000). Като резултат 25-те РПТ, дефинирани в Португалия, се считат за отговарящи както на националните, така и на европейските нужди, тъй като националните заинтересовани страни възприемат така изготвената класификация като логична и практически приложима, а мащабът на РПТ съответства на този на регионите от ниво NUTS 3. След поредица от тестове този подход дава добри резултати в държави с много различни морфологични особености. Следва да се направи уговорката, че са допустими някои корекции в стойностите на параметрите, за да бъдат отразени социално-икономическите особености на отделните страни. За разлика от параметрите за генериране на европейски РПТ тези за създаване на РПТ в национален мащаб трудно биха могли да се унифицират.

Считам, че с цел поддържане на йерархична структура правилното решение е като градивен компонент на европейските РПТ да се използват националните РПТ, а не изначалните компоненти за дефиниране на национални РПТ като LAU, грид клетки или други пространствени единици, за които входните данни са налични. Фиг. 9 показва резултатите за РПТ в Полша в национален и европейски мащаб. Националният набор от параметри (minNRS 0.667, celNRS 0.80, minBZ 4 000, celBZ 30 000) дава резултат от 339 РПТ, докато т.нар. „португалски параметри“, приложени към националните РПТ като градивен компонент, дават резултат от 83 РПТ, съобразени с целевия европейски мащаб.

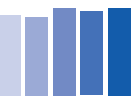
Фиг. 9. Национални (вляво) и европейски РПТ в Полша<sup>20</sup>

Източник: НСИ на Полша.

### 3. Трансгранични функционални райони, дигитализация, пандемия от COVID-19 и промяна в моделите на трудова мобилност

Наскоро проведено изследване на статистическия офис на Италия, представено на среща на работната група на Евростат по регионална статистика, градско и селско развитие, показва възможностите на РПТ за анализ на моделите на разпространение на пандемията от COVID-19<sup>21</sup>. За целите на изследването РПТ в Италия са класифицирани като градски, селски или междинни въз основа на гъстотата на населението и като антропогенни, земеделски или натурални според водещия тип земно покритие. Анализът показва нулева смъртност от COVID-19 в планинските и вътрешните изолирани РПТ, отдалечени от големи градове и от основни услуги като образование и здравеопазване. Т.е. не достъпът до здравни услуги, а слабата мобилност на населението е решаваща за ограничаване на пандемията. Следва да се има предвид, че границите на РПТ в Италия са произведени на базата на ЕТМ матрицата от Преброяване 2011. Пандемията от COVID-19 и променящият се характер на пазара на труда в условията на глобализация и дигитализация водят до изменение в представите ни за традиционните РПТ. Все повече

<sup>20</sup> Тъмносините граници на фигурата вдясно маркират границите на европейските РПТ, а светлосините обозначават границите на националните РПТ, използвани като градивни компоненти.



хора работят от дома си, дори и в България, където преди пандемията работата онлайн не беше толкова популярна.

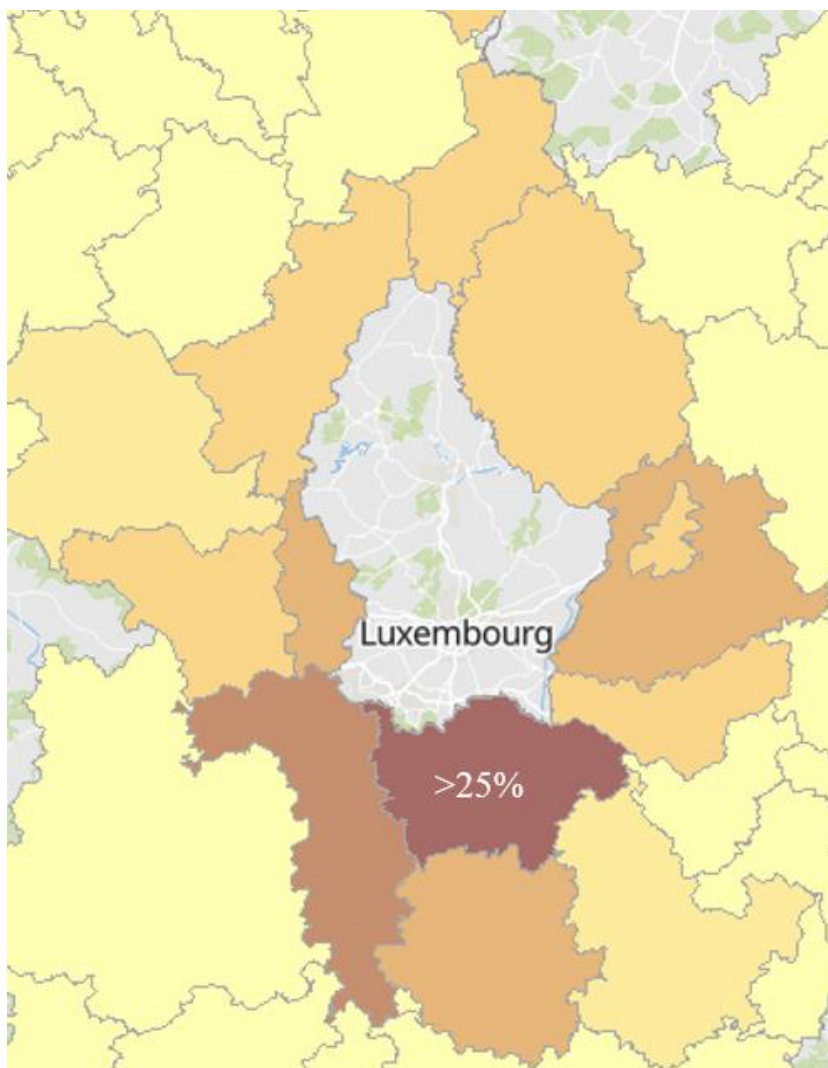
Дори в условията на пандемия **трансграничните ФГР и РПТ** традиционно са от специален интерес както на европейско, така и на национално равнище в държавите, в които феноменът на ЕТП през граница е в значителни мащаби. За целите на изследването „Статистика на градовете“ Евростат поддържа класификацията на трансграничните ФГР на Женева и Базел в Швейцария, като данни на това географско ниво са налични.

Трансграничният ФГР на Люксембург понастоящем не е включен в обхвата на изследването „Статистика на градовете“, но данни и географски граници са обществено достъпни на геопортала на Люксембург<sup>22</sup>. Дефинирането на транснационален ФГР е от съществено значение както за икономически гравитиращите към Люксембург части от Валония, Рейнланд-Пфалц и Саарланд, така и за самия Люксембург, тъй като без данни за ежедневните трудови пътувания през граница националният ФГР се очертава от държавната граница и съвпада с единствения NUTS 3 регион, съставляващ Люксембург. Фиг. 10 показва, че през 2019 г. една четвърт от транснационалните трудови пътувания по посока Люксембург са генерирани от региона на Тионвил.

---

<sup>22</sup> [https://map.sig-gr.eu/theme/marche\\_du\\_travail?version=3&zoom=9&X=697575&Y=6390813&lang=en&layers=2018&opacities=1&bgLayer=basemap\\_2015\\_global&crosshair=false](https://map.sig-gr.eu/theme/marche_du_travail?version=3&zoom=9&X=697575&Y=6390813&lang=en&layers=2018&opacities=1&bgLayer=basemap_2015_global&crosshair=false)

**Фиг. 10. Принос на съседни NUTS 3 региони към общия брой на трансграничните трудови пътуващи до Люксембург, 2019 година**



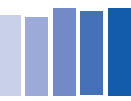
Източник: Геопортал на Люксембург.

Анализ на ESPON (European Spatial Planning and Observation Network) от 2018 г. разглежда региона на Залцбург като транснационален ФГР въз основа на силни икономически връзки със съседни територии в Германия и през призмата на трансграничното сътрудничество по отношение на обществени услуги, третиране на трансгранични отпадъци и трансгранични пътувания не само с цел работа, но и за обучение, пазаруване, туризъм и свободно време<sup>23</sup>.

<sup>23</sup>

<https://www.espon.eu/sites/default/files/attachments/ESPON%20CPS%2007%20Scientific%20Report%20Annex%20IV%20Salzburg.pdf>



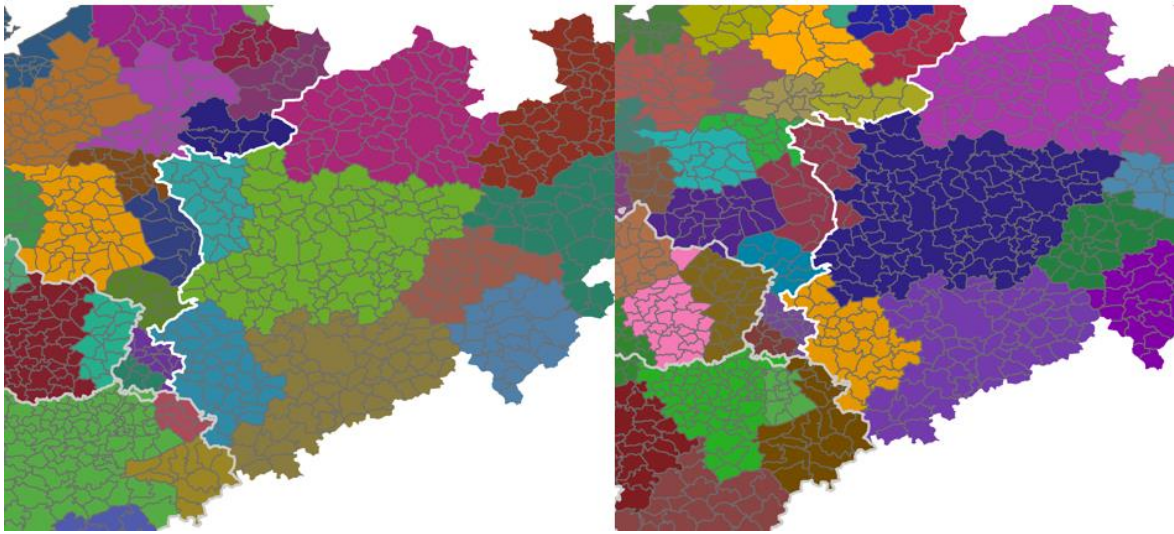


Що се отнася до трансграничните РПТ, основното предизвикателство при дефинирането им е да се гарантира, че входните данни са съгласувани в съседните страни и съдържат потоците от ЕТМ в двете посоки. Данните често дават само информация за страната по месторабота, но не и идентификация на LAU дестинацията, към която са насочени ЕТМ (в случай на данни от преброяването), или обратното (в случай на административни данни) - местоживеенето на приходящите работници не е известно.

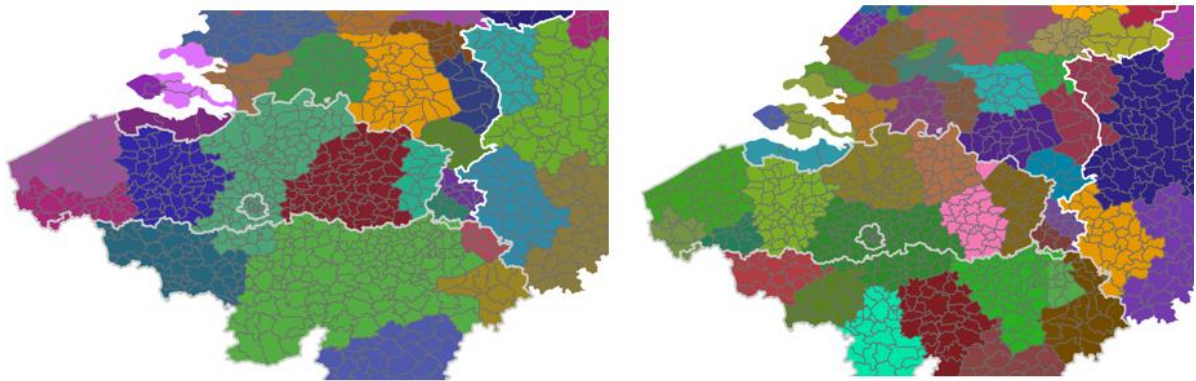
В рамките на проекта със съфинансиране от Европейската комисия, реализиран през периода 2016 - 2017 г., в който взе участие и България, статистиците от офиса на Нидерландия в сътрудничество с колеги от Белгия и Северен Рейн-Вестфалия (NRW) в Германия произвеждат трансгранични РПТ за трите държави<sup>24</sup>. РПТ на нидерландско-белгийската граница и на границата между Нидерландия и NRW са конструирани въз основа на национални данни за ЕТМ и като втори експеримент - въз основа на оценените трансгранични потоци. Резултатът показва, че включването на данни за трансграничните пътувания не оказва толкова голямо влияние в близост до държавните граници, но има голям ефект върху границите на РПТ във вътрешността на територията на държавите, като наблюдаваният ефект е по-силен на територията на Нидерландия. Заключение е, че включването на информацията за трансграничните ЕТМ води до по-логични клъстери. Фиг. 11 и фиг. 12 съдържат обобщение на резултатите.

---

<sup>24</sup> [https://ec.europa.eu/eurostat/cros/content/final-report-netherlands-cross-border-lmas\\_en](https://ec.europa.eu/eurostat/cros/content/final-report-netherlands-cross-border-lmas_en)

**Фиг. 11. РПТ на границата между Нидерландия и NRW<sup>25</sup>**

Източник: НСИ на Нидерландия.

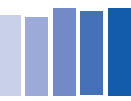
**Фиг. 12. РПТ на границата между Нидерландия и Белгия**

Източник: НСИ на Нидерландия.

И за трите държави е установено, че матрицата на ЕТМ е непълна и изисква извършването на някои оценки. Тази ситуация не е единствено проблем за Нидерландия и съседите ѝ. Това е често срещаната ситуация, при която се налага обмен на микроданни между държавите, както и интегриране на административни данни за лицата. За целите на този проект липсващите данни са импутирани посредством изчисляване на вероятността за осъществяване на ЕТМ през граница въз основа на предполагаемото разстояние за пътуване и данни за националността на зетите лица. Въпреки това експертите от Нидерландия, както и Евростат, препоръчват разработването в бъдеще на по-прецизна методология за попълване на липсващите елементи от матрицата. Въпреки

<sup>25</sup> Резултатите отляво са получени без данни за трансграничните ЕТМ; резултатите отдясно - при включване на данни за трансграничните ЕТМ в модела при използване на същия набор от параметри. Забележката важи и за следващата фигура.





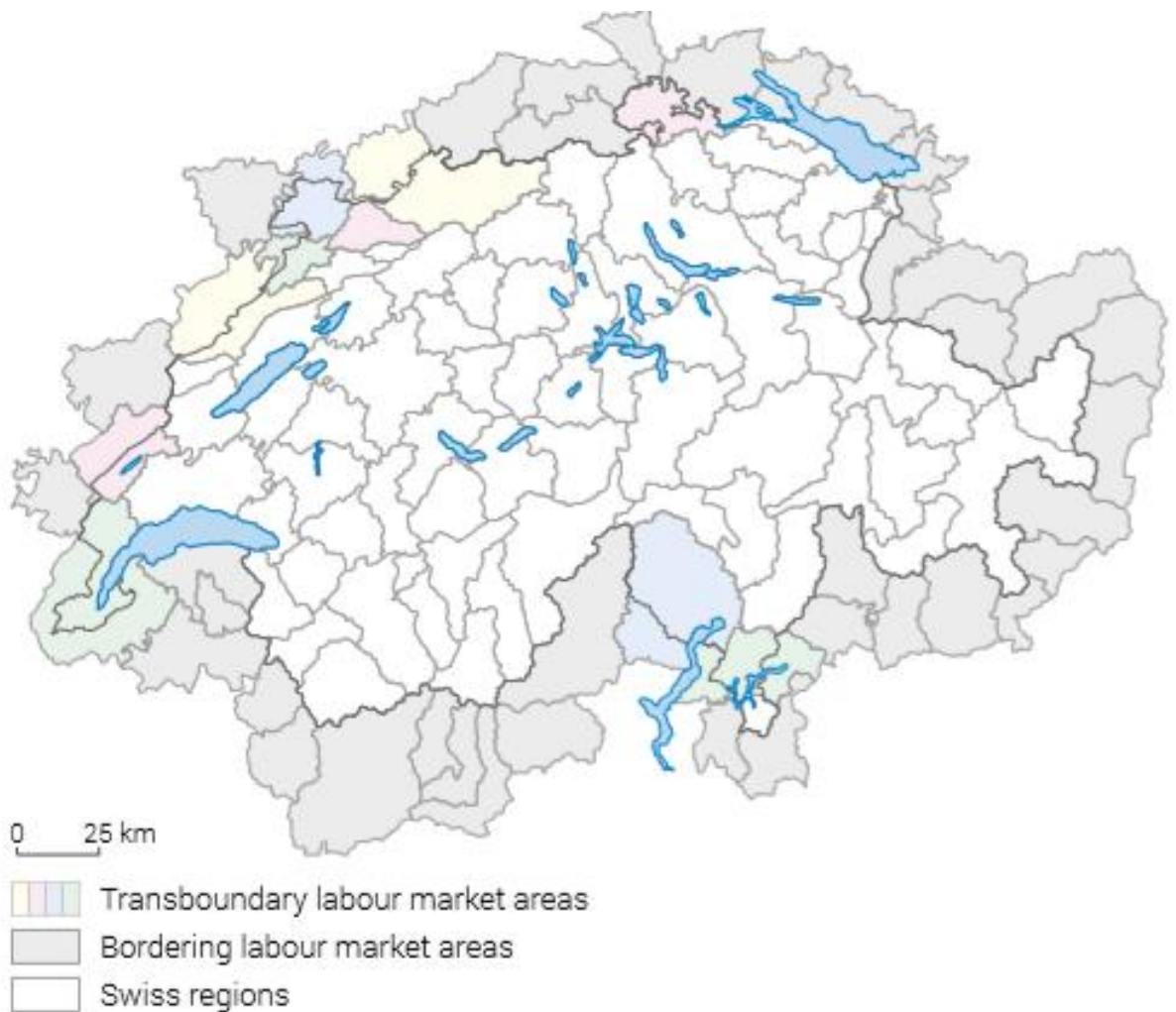
че феноменът на трансграничните трудови пътувания не е толкова характерен за България, особено в условията на COVID-19 кризата, създаването на една такава методология би могло да е предмет на бъдещо изследване на НСИ.

В публикация от 2019 г. швейцарските експерти повдигат допълнителни методологични и технически въпроси, свързани с трансграничните РПТ. Тествани са няколко възможности за дефиниране на трансгранични РПТ. Тъй като целта е да се поддържат РПТ, които да са разпознаваеми за швейцарските граждани, матрицата на ЕТМ е предварително зададена, като включва избрани местни административни единици в страната и зад граница. Анклавите се елиминират. Освен това трансграничните РПТ, които имат само една или две чужди „изолирани“ общини, се премахват, като тези общини се преразпределят към съседни „национални“ РПТ. Методът изисква значителна предварителна работа по събирането на данните за ЕТМ от всички съседни страни. Така дефинираните 10 трансгранични РПТ в районите на Женева, района на Юра, Базел, Шафхаузен и Тичино са представени на фиг. 13<sup>26</sup>.

---

<sup>26</sup> <https://www.bfs.admin.ch/bfs/en/home/news/whats-new.assetdetail.8948838.html>

Фиг. 13. Трансгранични РПТ за Швейцария



Source: FSO – Territorial typologies of Switzerland

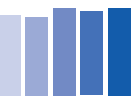
© FSO 2019

Източник: НСИ на Швейцария.

Въпреки че не разполагаме с конкретни данни, за да измерим ефекта от пандемията от COVID-19, можем да твърдим, че след пролетта на 2020 г. транснационалните ФГР и РПТ са силно видоизменени наред с националните поради принудително променените модели на мобилност на заетите лица. За период от една година работата и пазаруването онлайн, както и доставките до дома, се превърнаха в траен бихевиористичен модел.

#### 4. Заключение и визия за бъдещето

Предлагам в бъдеще административно-териториалното устройство на страната и другите европейски държави да се основава на функционални райони от типа на РПТ,



унаследените граници на NUTS имат върху регионалните данни. За функционални райони като РПТ също могат да се приложат критерии за йерархичност и праг за брой население по подобие на NUTS класификацията. Редица опити за дефиниране на национални и европейски РПТ са широко обсъждани на различни форуми, организирани от Евростат. Разбира се, в годината на евентуално влизане в сила на граници на административно-териториални единици, основани на функционален критерий, коефициентът на вариация относно броя на населението се очаква да бъде значителен. Добавената стойност от граници, отговарящи на социално-икономическата реалност, наред с преизчисление на динамичните редове от данни назад във времето по-скоро би минимизирала неудобството, което причинява всяко мащабно прерайониране.

Считам, че бъдещето е в грида както за дефиниране на границите на РПТ, така и за оптимизиране на извадките, използвани за социални изследвания с цел по-дълги динамични редове и подпомагане на производството на данни не само на ниво РПТ, но и за всевъзможни функционални райони от интерес за потребители и изследователи, в т.ч. и ФГР. Това е свързано и с по-широкото използване на регистри и възможностите на големите данни. Налице са и редица други аргументи в подкрепа на използването на грида като универсално решение в отговор на редица предизвикателства в статистическата теория и практика.

По подобие на разширената регионална класификация на база отдалеченост от метрополни региони, изготвена от ОИСР, като предмет на бъдещо изследване РПТ могат да бъдат класифицирани според достъпа им до транспортна мрежа от висок клас (магистрала и първокласни пътища) на основата на TomTom или алтернативни източници на географски данни. Ценно би било и по-нататъшното обвързване на РПТ с други териториални типологии, като ФГР, типологията на градските, селските и междинните региони, със степента на урбанизация и метрополните региони.

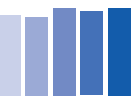
Намирам, че допълването на променливите на ниво град и ФГР, публикувани от НСИ и Евростат, с данни на ниво зони на регулярно пътуване до работното място в града ще увеличи аналитичния потенциал на информацията, събирана в рамките на проекта „Статистика на градовете“. Това е познато от практиката на колегите от статистическия офис на Швейцария<sup>27</sup>. Такава информация ще позволи да се направи анализ на реалния принос на хинтерланда на града към икономическите резултати и демографското

<sup>27</sup> [https://www.atlas.bfs.admin.ch/maps/162/de/15778\\_12157\\_12155\\_12154/24663.html](https://www.atlas.bfs.admin.ch/maps/162/de/15778_12157_12155_12154/24663.html)

развитие на целокупния функционален градски район. Също така би било интересно да се изследва интензитетът на преселванията от хинтерланда към града, както и в обратната посока.

Вероятно се налага актуализация на дефиницията за ежедневно трудово пътуване. Например според действащата дефиниция, ако поради пандемията и въведения дистанционен режим на работа лицето дългосрочно работи онлайн (telework) в Севлиево, където е адресът му по местоживееене, а офиса му физически е разположен в Габрово, то технически погледнато, лицето не извършва ежедневно трудово пътуване.

Не на последно място препоръчвам включването на методиката и алгоритъма за дефиниране на ФГР и РПТ в обучението по специалностите „География“, „Статистика“ и „Икономика на труда“ във висшите училища. Методите са широко приложими в практиката, а в двете концепции се съчетават регионална статистика, икономика на труда и транспорта и социално-икономическа география. Познаването на метода и теоретичната постановка на изложените проблеми би разширило кръгозора на студентите към сродни области на познанието и би подобрило пространственото им мислене.



## Използвани съкращения

БВП - брутен вътрешен продукт

ЕТП - ежедневни трудови пътувания

ЕТМ -

НСИ - Национален статистически институт

ОИСР - Организация за икономическо сътрудничество и развитие

РПТ - райони на пазара на труда

ФГР - функционални географски райони

BZ - брой на заети

celBZ - целеви брой заети

celNRS - целева независимост по отношение на работната сила

DG REGIO - Directorate-General for Regional and Urban Policy

FUA - Functional Urban Area

JRC - Joint Research Centre

LAU - Local Administrative Units

minBZ - минимален брой заети

minNRS - минимална независимост по отношение на работната сила

NRS - независимост по отношение на работната сила

NRW - Северен Рейн-Вестфалия

NUTS - La nomenclature des unités territoriales statistiques

**ЦИТИРАНА ЛИТЕРАТУРА:**

**НСИ** (2020). Райони на пазара на труда. София

**Directorate-General for Regional and Urban Policy** (2018), Border regiona data collection,  
[https://ec.europa.eu/regional\\_policy/en/information/publications/studies/2018/border-region-data-collection](https://ec.europa.eu/regional_policy/en/information/publications/studies/2018/border-region-data-collection)

**ESPON** (2018), Cross-border Public Services, Targeted Analysis Final Report (Scientific Report – Annex IV Case study report – EuRegio SalzburgBerchtesgadener Land-Traunstein), <https://www.espon.eu/CPS>

**Eurostat** (2017), Eurostat Regional Yearbook (242-243), <https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-HA-17-001>

**Eurostat** (2019), Methodological manual on territorial typologies, Luxembourg <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-008>

**Eurostat** (2020), European harmonised Labour Market Areas — Methodology on functional geographies with potential, Luxembourg <https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/ks-tc-20-002>

**Eurostat, Collaboration in Research and Methodology for Official Statistics, Groups, General interest groups, Labour Market Areas**, достъпен към февруари 2021 г., <https://ec.europa.eu/eurostat/cros/content/labour-market-areas>

**Fadic, M., et al.** (2019), Classifying small (TL3) regions based on metropolitan population, low density and remoteness, OECD Regional Development Working Papers, No. 2019/06, OECD Publishing, Paris, <https://doi.org/10.1787/b902cc00-en>.

**Federal Statistical Office (FSO)** (2019), Labour Market Areas 2018, <https://www.bfs.admin.ch/bfs/en/home/news/whats-new.assetdetail.8948838.html>

**OECD** (2020), Delineating Functional Areas in All Territories, OECD Territorial Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/07970966-en>.

## МЕТОДИ И КОНЦЕПЦИИ ЗА ФУНКЦИОНАЛНО ГЕОГРАФСКО РАЙОНИРАНЕ ЗА СТАТИСТИЧЕСКИ ЦЕЛИ

*Валерия Ангелова\**

**РЕЗЮМЕ** Целта на настоящото изследване е да се анализират практическите предимства на различни статистически подходи за функционално географско райониране, основани на данни за ежедневните трудови пътувания, и да се направи обобщение на методите по достъпен начин с помощта на редица примери. Акцентът е поставен върху функционалните градски райони и районите на пазара на труда. Авторът представя и редица предизвикателства, свързани с дефинирането и използването на тези функционални райони за статистически и аналитични цели, и очертава възможни насоки за бъдещето, като отчита промените в моделите на трудова мобилност и в условията на пазара на труда в резултат на пандемията от COVID-19.

**Ключови думи:** функционални градски райони; райони на пазара на труда; ежедневни трудови пътувания; работна сила; икономическа интеграция; териториални типологии; гри Държавен експерт в отдел „Регионална статистика и индикатори за мониторинг“ в Националния статистически институт; e-mail: VAngelova@nsi.bg.

---

\* Държавен експерт в отдел „Регионална статистика и индикатори за мониторинг“ в Националния статистически институт; e-mail: VAngelova@nsi.bg.

## МЕТОДЫ И КОНЦЕПЦИИ ФУНКЦИОНАЛЬНОГО ГЕОГРАФИЧЕСКОГО РАЙОНИРОВАНИЯ В СТАТИСТИЧЕСКИХ ЦЕЛЯХ

*Валерия Ангелова\**

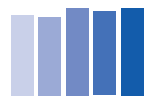
**РЕЗЮМЕ** Целью настоящего исследования является анализ практических преимуществ различных статистических подходов относительно функционального географического районирования на основе данных о ежедневных трудовых поездках и обобщение методов в доступной форме с помощью ряда примеров. Акцент делается на функциональных городских районах и районах рынка труда. Автор также представляет ряд проблем, связанных с дефинированием и использованием этих функциональных районов в статистических и аналитических целях, и намечает возможные направления на будущее с учетом изменений в моделях трудовой мобильности и условиях на рынке труда в результате влияния пандемии COVID-19.

**Ключевые слова:** функциональные городские районы; районы рынка труда; ежедневные трудовые поездки; рабочая сила; экономическая интеграция; территориальные типологии; грид

---

\* Государственный эксперт отдела „Региональная статистика и мониторинговые индикаторы“ в Национальном статистическом институте; e-mail: VAngelova@nsi.bg.





## METHODS AND CONCEPTS FOR FUNCTIONAL GEOGRAPHICAL ZONING FOR STATISTICAL PURPOSES

*Valeria Angelova\**

**SUMMARY** The aim of this study is to analyze the practical advantages of various statistical approaches to functional geographical zoning based on data from daily business trips and to summarize the methods in an accessible way using a number of examples.

The emphasis is on functional urban areas and labor market areas. The author presents a number of challenges related to the definition and use of these functional areas for statistical and analytical purposes, and possible outlines for the future, taking into account changes in labor mobility patterns and labor market conditions as a result of the COVID-19 pandemic.

**Keywords:** functional urban areas; labor market areas; daily business trips; work force; economic integration; territorial typologies; grid

---

\* State expert in the Regional Statistics and Monitoring Indicators Department at the National Statistical Institute; e-mail: VAngelova@nsi.bg.



**СТАТИСТИЧЕСКИ ИЗСЛЕДВАНИЯ И  
АНАЛИЗИ**





## ОРГАНИЗАЦИОННАТА СТРУКТУРА НА НАЦИОНАЛНИЯ СТАТИСТИЧЕСКИ ИНСТИТУТ - СИНТЕЗ МЕЖДУ ТРАДИЦИИ И СЪВРЕМЕННОСТ

*Цветозария Гатева\**



### **Въведение**

В глобалното информационно пространство необходимостта от анализи на икономическото, социалното и демографското състояние на страните, тенденциите на развитие и на факторите, от които зависят, непрекъснато нараства, поради което се увеличава и необходимостта от статистическа информация. Осигуряването на надеждна, достатъчна по обем, разнообразна, детайлизирана и сравнима информация е основна задача и отговорност на статистическия орган на страната. Основен източник на информация за протичащите икономически и социални явления и процеси в България са провежданите от Националния статистически институт (НСИ) и органите на статистиката статистически изследвания. За своята 136-годишна история българската държавна статистика е постигнала значими

---

\* Началник на отдел „Статистически изследвания - Ловеч“, ТСБ - Северозапад, докторант в Стопанска академия „Д. А. Ценов“ - Свищов; e-mail: CGateva@NSI.bg.

резултати, а в организационната структура на държавния статистически орган настъпват редица промени - в наименованието, подчинеността, ресурсното осигуряване и други.

**Обект на изследване** е организационната структура на държавния статистически орган от неговото създаване до днес.

**Целта на статията** е да се изследват промените в организационната структура на НСИ в хронологичен план и да се докаже, че те са допринесли за подобряване на статистическата дейност и за повишаване на качеството на статистическата информация.

За изпълнението на поставената цел са формулирани следните **задачи**:

1. Да се извърши ретроспективен анализ на организационната структура на националния статистически орган на България от неговото създаване до настоящия момент.

2. Да се разкрият сходни белези и характеристики в организационната структура през отделните етапи на развитие.

3. Да се направи анализ на ресурсното осигуряване и ефективността на промените в организационната структура на НСИ за периода 2000 - 2019 година.

Изследването се основава на проучване на официални статистически документи - укази, закони, правилници и други документи, публикации, посветени на годишнини от създаването на българската държавна статистика, Стратегията за развитие на Националната статистическа система на Република България и отчетите за изпълнението на Националната статистическа програма и за дейността на НСИ.

### **1. Ретроспективен анализ на създаването и функционирането на държавния статистически орган на България**

Развитието на българската държавна статистика е белязано от много промени, повечето от които в наименованието и подчинеността на институцията и в по-малка степен свързани с промени в нейната организация и структура. Същевременно промените в социално-икономическото развитие на страната са в основата на извършените промени в подчинеността и структурата на статистическата институция. Въз основа на това в развитието на българската държавна статистика могат да бъдат открити три етапа:

• I етап. Създаване и утвърждаване на българската държавна статистика от 1880 до 1944 година;



- II етап. Българската държавна статистика в условията на планова икономика;
- III етап. Българската държавна статистика в условията на преход към пазарна икономика и членство в Европейския съюз (ЕС).

### **I етап. Създаване и утвърждаване на българската държавна статистика от 1880 до 1944 година**

Началото на българската статистика е поставено след Освобождението през 1878 г. с цел задоволяване на първите потребности на Княжество България от установяване на броя на населението за провеждане на първите парламентарни избори за учредително събрание и за извършване на териториално разпределение на страната. Така първата статистическа служба се създава през 1879 г. към Отдела за вътрешните работи на Руския императорски комисариат. На 21 януари 1880 г. е учредено Статистическо отделение към Министерството на правосъдието, последвано от образуване на Статистическо организационно отделение към Министерството на финансите на 25 юни 1880 година (Стефанов, 1980, с. 9).

Впоследствие отделението се прехвърля към Министерството на просвещението, като от 19 август 1881 г. с Указ № 712 на княз Александър I „се възвисява в самостоятелно статистическо бюро със задачата да събира, обработва и обнародва всяка година статистически данни по всички клонове на управлението и всички явления, отнасящи се до физическото, икономическото, интелектуалното и моралното състояние на държавата“ (Мичев, 1984, с. 14).

През 1897 г. е приет Закон за Дирекцията на статистиката на Българското Княжество, с който статистическото учреждение се издига на много високо равнище в системата на администрацията - задачите на дирекцията се увеличават и разширяват. В организационно отношение, съобразно потребностите на държавата, българската статистическа институция е била подчинена на различни министерства (табл. 1). От 1898 г. прераства в самостоятелно учреждение като Дирекция на статистиката.

Периодът на създаването на българската държавна статистика завършва с подробна анкета от правителствена комисия, включваща осем именити учени и висши чиновници. От анкетата се установява липса на достатъчна теоретическа и методологическа обосновааност на много статистики, започнати и проведени през



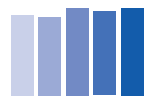
периода от август 1903 г. до февруари 1907 година. През този период ръководството на Дирекцията на статистиката се придържа към съществуващата практика в други страни и разширява обхвата на изследванията до 12 области на обществения живот с 96 отделни статистики. Пренасянето на чуждестранния опит в дейността на статистическия орган е добро начало, но основният недостатък е, че не се държи сметка за теоретичното и методологическото обосноваване, за липсата на условия и възможности за всяка нова статистика. С доклада на анкетната комисия се отхвърлят всички статистики, които не са достатъчно полезни за обществото и за които няма обективни условия и възможности за водене (Стефанов, 1980, с. 10).

От 1908 г. започва периодът на творчески подем в развитието на българската държавна статистика под ръководството на Кирил Г. Попов. Главна дирекция на статистиката е издигната на равнището на автономна държавна институция, която административно се намира под ведомството на Министерството на търговията, промишлеността и труда. Ръководителите на това министерство не са могли да се противопоставят на авторитета - служебен и обществен - на Кирил Г. Попов, който успява да защити кадрите на Главната дирекция на статистиката.

Първите български статистики са били само четирима души, а броят на щатните сътрудници, макар и с бавни темпове, непрекъснато се увеличава. През този период на държавната статистика се предоставя временен персонал от щата на другите заинтересовани ведомства, който участва в провеждането на преброявания, анкети и други статистически изследвания.

От 1910 г. българската държавна статистика получава високо положение в системата на администрацията. Приет е нов закон за Главната дирекция на статистиката, който издига йерархическото значение на централното статистическо учреждение до ранг на министерство. Главната дирекция на статистиката изпълнява своите функции самостоятелно и поддържа връзки непосредствено в страната и чужбина.

Ранното обособяване на Главната дирекция на статистиката като специална организация с относителна самостоятелност в ръководенето и извършването на статистическа работа в страната се оказва от решаващо значение за по-нататъшното развитие на статистическата дейност, за разширяване на обхвата на



изследванията и постигането на сравнително високо ниво за времето и условията тогава.

Основните трудности във функционирането на учреждението и в извършваната статистическа дейност са свързани с решаването на нови задачи и с необходимостта от разработване на специални конкретни методики при липсата на собствен опит и на достатъчно изпълнителски кадри с нужната подготовка и квалификация при ниска или почти отсъстваща механизированост на работата и при големи финансови ограничения.

Въпреки посочените трудности статистическите изследвания **през периода 1900 - 1944 г. са проведени на международно равнище**. Статистическата информация се произвежда на основата на система от месечни, тримесечни, годишни и еднократни статистически наблюдения. Голяма част от произведените статистически данни са съобразени както с решенията и препоръките на международните статистически конгреси и конференции, така и със сключените международни конвенции по методологически статистически въпроси.

## **II етап. Българската държавна статистика в условията на планова икономика**

През 1946 г. Народното събрание приема Закон за организиране на статистиката в България, като Главната дирекция на статистиката преминава на пряко подчинение на Министерския съвет. На основата на този закон за първи път се изграждат местни статистически органи като структури на Главната дирекция на статистиката. Всички градски общини с население над 15 хил. души са задължени да имат своя специална статистическа служба. Функциите и задачите на местните статистически органи са дефинирани подробно в Правилника за приложението на закона. Първото голямо изследване, в което вземат участие местните статистически служби, е преброяването на населението и жилищния фонд към 31 декември 1946 година.

В края на 1947 г. Главна дирекция на статистиката преминава на подчинение на Държавната планова комисия с оглед статистическата дейност да се насочи предимно към отчитане на изпълнението на държавния народностопански план. Във връзка с това се определят и изследванията, които се провеждат и които имат за цел получаване на числови стойности,

характеризиращи социално-икономическото развитие на страната през този период.

С постановление на Министерския съвет от 3 март 1953 г. се създава Централното статистическо управление (ЦСУ) като общодържавен орган, осъществяващ ръководството на цялостната статистическа дейност в страната, който е на пряко подчинение на Министерския съвет.

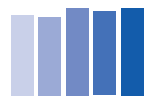
След реорганизация на статистическата служба и в резултат на новото административно деление на страната през 1959 г. са сформирани 30 окръжни отдела на статистиката, намиращи се в съответните окръжни центрове. Едновременно с това се прекратява дейността на другите структури по места - градски и районни статистически инспекторати и участъкови инспектори.

С постановление на Министерския съвет през 1969 г. е основано Държавно управление за информация със задача да ръководи изграждането на единна система за социална информация в страната. В състава на създаденото през 1971 г. Министерство на информацията и съобщенията влизат Централното статистическо управление и Управление „Изследвания, анализи и доклади“, които се занимават само със статистическа дейност.

През 1976 г. Централното статистическо управление се обособява като самостоятелно ведомство, пряко подчинено на Министерския съвет. С цел интегриране на статистическата дейност с дейността по изграждане на единна система за социална информация и колективно ползване на електронноизчислителните центрове през 1977 г. се създава Комитет по единна система за социална информация, в чийто състав влиза и Централното статистическо управление.

### **III етап. Българската държавна статистика в условията на преход към пазарна икономика и членство в Европейския съюз**

Динамичните процеси в прехода от планова към пазарна икономика изискват изграждането на качествено нова статистическа система. През 1991 г. Великото народно събрание приема Закон за статистиката, с който се създава Националният статистически институт с местни органи - 28 териториални статистически бюра (ТСБ). Националният статистически институт е на подчинение на Народното събрание. Започва процес на разработване на нови и усъвършенстване на действащите методологии в съответствие с международните



статистически стандарти, регламенти, препоръки и споразумения на Евростат, МВФ, ООН и други международни организации.

През 1999 г. е приет нов Закон за статистиката, разработен в съответствие с нормативната база на Европейския съюз - Регламент на Съвета на Европа № 322 от 17 февруари 1997 г., и с фундаменталните принципи на официалната статистика, утвърдени от Статистическата комисия на ООН. Целта на закона е да създаде нормативната рамка на организацията и провеждането на статистическата дейност в страната. Дефинирани са понятия като: основни принципи на статистическата дейност - безпристрастност, надеждност, ефективност, адекватност, публичност, поверителност; органи на статистиката; статистическо изследване; статистическа информация. Законът урежда статута на НСИ като държавна агенция на пряко подчинение на Министерския съвет. В процеса на преговори за членство в ЕС НСИ предоставя необходимата статистическа информация както по глава 12 „Статистика“, така и по други глави от преговорния процес.

### 1. Наименование и подчиненост на държавния статистически орган на България

Година на създаване/промяна	Наименование	Подчиненост/в структурата на
1879	Отдел за вътрешните работи - не е действал	Руски императорски комисариат
1880	Статистическо отделение	Министерство на правосъдието
1880	Статистическо организационно отделение	Министерство на финансите
1880	Статистическо организационно отделение	Министерство на народното просвещение
1881	Статистическо бюро	Министерство на народното просвещение
1894	Статистическо бюро	Министерство на търговията и земеделието
1897	Дирекция на статистиката на Българското Княжество	Министерство на търговията и земеделието
1910	Главна дирекция на статистиката на Българското царство	Министерство на търговията, промишлеността и труда
1946	Главна дирекция на статистиката	Министерски съвет
1947	Главна дирекция на статистиката	Държавна планова комисия
1953	Централно статистическо управление	Министерски съвет
1971	Централно статистическо управление	Министерство на информацията и съобщенията
1976	Централно статистическо управление	Министерски съвет
1977	Централно статистическо управление	Комитет по единна система за социална информация (без ранг на министерство) при Министерския съвет
1991	Национален статистически институт	Народно събрание
1999 до момента	Национален статистически институт	Министерски съвет

От систематизирането на подчинеността на статистическата институция прави впечатление фактът, че промените са извършвани главно в по-ранните

периоди от създаването ѝ. През последните 60 години статистическата институция е подчинена на Министерския съвет с изключение на годините от 1991 до 1999, когато е подчинена на Народното събрание. Според Т. Къналиев подчинеността на НСИ на Народното събрание през този период определя неговия достатъчно висок ранг и по принцип осигурява по-голяма безпристрастност на статистическата информационна дейност в страната, като същевременно тя подлежи на по-широк обществен контрол, а понижаването на статута на НСИ и подчиняването му непосредствено на Министерския съвет не намалява възможната опасност от политическо и административно влияние и евентуален натиск и намеса на управляващите органи в дейността му (Къналиев, 2005, с. 9).

Националният статистически институт остава на подчинение на Министерския съвет и след приемането на България за член на Европейския съюз. Тази устойчивост е важна предпоставка за правилно формулиране на мисията, ценностните характеристики, визията и основните задачи на българската статистика.

## **2. Анализ на промените в организационната структура на българската държавна статистика**

Още в годините на създаване и утвърждаване на българската държавна статистика с нормативните актове ясно се посочват задачите на статистическата институция - „да събира, обработва и обнародва всяка година статистически данни по всички клонове на управлението и всички явления, отнасящи се до физическото, икономическото, интелектуалното и моралното състояние на държавата“ (Мичев, 1984, с. 14). В Закона за Главната дирекция на статистиката на Българското царство много прецизно е представена нейната структура, която се е изчерпвала с наличието на три отдела - демографски; стопански; културен и политически, като към всеки от тях има по две отделения, включващи в състава си по две-три бюра. В съответните документи много точно са посочени изискванията към наетия за работа персонал, както и условията, на които трябва да отговарят заемащите съответната длъжност независимо от подчинеността на институцията.

Въвеждането на машинна обработка на данните дава възможност да се съкрати времето между събирането, обработката и публикуването им, а това в значителна степен променя изискванията към работещите в системата на

статистиката, към организацията и функционирането на статистическата институция и нейната организационна структура.

През различните периоди на социално-икономическо развитие пред националната статистика са поставяни различни задачи, които тя е изпълнявала отговорно, компетентно и спазвайки предварително определените срокове за предоставяне и публикуване на статистическите данни. В условията на масово навлизане на информационните и комуникационните технологии във всички области изискванията към количеството и качеството на статистическата информация непрекъснато нарастват. По този начин националната статистика се изправя пред нови предизвикателства - да се развива в крак с времето и съвременните технологии, да следва европейските и световните стандарти за производство на статистическа информация, но да запази традициите в своята дейност, както и изградената структура, която да рационализира и оптимизира.

Прави впечатление относително стабилната структура на статистическата институция на България - променят се наименованието и подчинеността на институцията, но създадената през 1946 г. структура, състояща се от централно управление (ЦУ) и местни структури, съществува и към настоящия момент (табл. 2). От 1946 до 1953 г. към Главна дирекция на статистиката функционират градски статистически служби в градските общини с население над 50 хил. души и областни статистически служби. Районни статистики има в селските и градските общини и в най-големите градове. В структурата на статистическата институция настъпват промени през 1953 г., когато към Централното статистическо управление функционират окръжни статистически управления; околийски, градски и районни статистически инспекторати; участъкови инспектори.

В резултат на настъпилите промени в административното деление на страната през 1959 г. към ЦСУ се създават 30 окръжни отдела на статистиката в съответните окръжни центрове. В структурата на статистическата институция настъпва значителна промяна, тъй като е прекратена дейността на останалите структури по места - градски и районни статистически инспекторати и участъкови инспектори. През следващите години към ЦСУ са обособени окръжни отдели по статистика и окръжни статистически управления, чиято дейност и отговорности не се различават особено. От 1987 до 1991 г. към ЦСУ функционират районни статистически управления.



## 2. Наименование и структура на статистическата институция на България през периода 1946 - 2015 година

Година на създаване/промяна	Наименование	Структура
1946	Главна дирекция на статистиката - за първи път се създават местни органи	1. Районни статистики - в селските и градските общини и в най-големите градове. В градските общини с население над 50 хил. души - Градска статистическа служба 2. Околийски статистик 3. Областни статистически служби
1953	Централно статистическо управление	1. Окръжни статистически управления 2. Околийски, градски и районни статистически инспекторати 3. Участъкови инспектори
1959	Централно статистическо управление	Окръжни отдели „Статистика“ - с двойно подчинение
1959	Централно статистическо управление	Окръжни отдели по статистика
1976	Централно статистическо управление	Окръжни статистически управления
1977	Централно статистическо управление	Отдели към териториалните информационно-изчислителни центрове (ТИИЦ)
1987	Централно статистическо управление	Районни статистически управления
1991	Национален статистически институт	28 териториални статистически бюра
2015	Национален статистически институт	6 териториални статистически бюра (28 отдела)

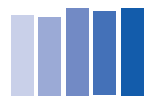
Законодателната регламентация на дейността на НСИ го определя като водеща институция при осигуряването на икономическа и социална информация. НСИ използва методики, синхронизирани с Евростат, гарантиращи съпоставимост на информацията по основни показатели; разполага с богата база данни, даваща

възможност за проследяване на тенденции; има утвърдени експерти и надеждни анкетьори.

Съгласно първия Закон за статистиката (1991) Националният статистически институт е държавно учреждение, което осъществява статистическа дейност в страната и осигурява обществото със статистическа информация. Националният статистически институт се състои от централно управление и 28 териториални статистически бюра, разположени в съответните областни центрове на страната. Териториалните статистически бюра осъществяват функциите и задачите на Националния статистически институт по производството и разпространението на статистическа информация на регионално ниво. Тази структура на НСИ се запазва в продължение на почти 25 години.

Дейността на НСИ е доста специфична и не обслужва административно граждани. Въпреки това статистическата институция се определя като държавна администрация и участва във всички действия, свързани с провеждането на реформите в държавната администрация и най-вече с оптимизацията на персонала. Развитието на информационните и комуникационните технологии и навлизането им в сферата на администрацията и управлението поставят нови предизвикателства пред НСИ и Националната статистическа система - изграждане на електронно управление, сайтове на отделните администрации, компютризация на работните места в публичната сфера, електронни източници на информация, разработване и поддържане на интегрирани информационни системи и други. Националният статистически институт обаче произвежда информационен продукт, дейност, която по-скоро е производствена и коренно се различава от административната. За периода от 2005 до 2014 г. се правят три оптимизации на персонала в системата на НСИ. През същия период завършва и работата на два екипа външни и международни експерти, работещи по проблематиката за структурата на Националния статистически институт в рамките на проекти по програма ФАР 2003. Първият проект приключва през 2006 г. с разработване на „Концепция за реорганизация на регионалната структура на Националния статистически институт“. В нея екипът предлага:

- Да се създаде междинна структура Регионален статистически офис, включващ в себе си няколко области, например областите от съответния район за планиране;



- ТСБ, включително и там, където ще има Регионален офис, да се преобразуват в областни статистически офиси, като им се отнемат функциите, свързани с финанси и човешки ресурси, които преминават към Регионалния офис;
- Да се намали персоналът в ТСБ.

Посочените предложения са обосновани с очакваното увеличаване на използването на информационните технологии, опита на някои страни, в които са направени промени, и приложената подобна структура в други администрации в страната.

Вторият проект, и по-конкретно дейност 4.1.3 „Устойчиво развитие на Националната статистическа система: Реорганизация на регионалните структури на НСИ“, приключва през 2009 г. и извежда коренно различна теза. Резултатите от него не поставят под съмнение съществуването на териториалните статистически бюра, а по-скоро се анализира тяхното състояние и се открояват проблемите им, набелязват се мерки за развитието и утвърждаването им като действени местни центрове.

Във фокуса на проекта попадат и теми, които са свързани с отделни елементи на политиката по качеството, като: обучение на ръководния състав на института и останалия персонал; въвеждане на система за отчитане и контрол; подобряване на качеството на работния поток между ЦУ на НСИ и ТСБ, от една страна, и между отделните ТСБ, от друга; оптимизиране на разпределението на задачите и компетенциите между централния офис и регионалните звена; намаляване на тежестта върху респондентите и анкетъорите.

В резултат на направените анализи и оценки в рамките на период от три години се срещат две коренно противоположни мнения. През 2015 г. приключва проект „Модернизиране на организацията и функционирането на НСИ“ с финансовата подкрепа на Оперативна програма „Административен капацитет“. Едновременно с това пилотно се извършва реорганизация на ТСБ, които влизат в състава на ТСБ - Югозапад. Целта е създаване на нова действаща структура, устойчива на сътресения, успяваща да се справи с предизвикателствата пред нея.

От 1 октомври 2015 г. в системата на НСИ е извършено реструктуриране с цел подобряване на административния капацитет, гъвкавост при разпределение на задачите, оптимизиране на работата и ефективна организация на ресурсите (НСИ, Отчет за изпълнението на Националната статистическа програма и за

дейността на НСИ през 2015 г., с. 56). Териториалната структура на НСИ е определена с ПМС № 272 от 8.10.2015 г., според което остават следните шест териториални статистически бюра:

- ТСБ - Северозапад с център гр. Плевен
- ТСБ - Север с център гр. Русе
- ТСБ - Североизток с център гр. Варна
- ТСБ - Югоизток с център гр. Бургас
- ТСБ - Юг с център гр. Пловдив
- ТСБ - Югозапад с център гр. София.

Посочените териториални статистически бюра са с единна обща администрация и отдели „Статистически изследвания“ в областните градове, в които през предходните години е функционирало ТСБ.

В отговор на обществените очаквания НСИ и териториалните му поделения се стремят да осигуряват сравними, надеждни и отговарящи на изискванията за конфиденциалност и защита на информацията данни, произведени в съответствие с международните стандарти.

Цялостната дейност на НСИ е насочена към практическата реализация на формулираните в *Стратегията за развитието на Националната статистическа система на Република България, 2013 - 2017 година* **визия** („Националната статистическа система да произвежда и предоставя на всички групи потребители статистически продукти и услуги с високо качество, произведени по хармонизирани европейски методологии и стандарти“) и **главна цел** („... да се усъвършенства разработването, производството и разпространението на качествена статистическа информация за всички групи потребители, съобразно техните информационни нужди“) (НСИ, Стратегия за развитие на Националната статистическа система на Република България, 2013 - 2017 г., 2013, с. 4).

### **3. Анализ на ресурсната осигуреност на НСИ**

Успешното интегриране в общото европейско административно пространство е свързано както с необходимостта от подобряване на професионалните умения за формиране на вътрешни политики, така и с подходящи квалификационни инициативи за адекватна работа в отделните дирекции. Опитът в тази посока на другите страни може да бъде полезен на българската статистика. Прилагането на европейските изисквания в областта на

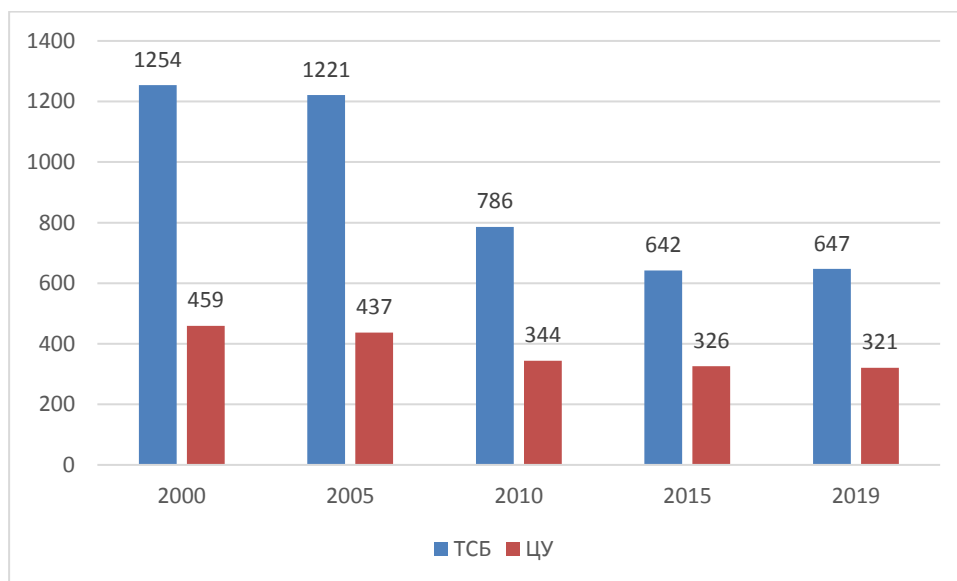


статистиката може да бъде ускорено чрез провеждане на специализирано обучение за запознаване с добрите европейски практики, което да обхване и служителите на ТСБ. По този начин експертите на различни нива в институцията ще натрупат ценен професионален опит, който да използват в своята работа.

Изменението в числеността и структурата на персонала на НСИ в значителна степен е в резултат на въведените нови организационни подходи в управлението на процесите по събиране, обработка и производство на статистическата информация - онлайн системи за въвеждане на данни по различните статистически наблюдения (в т.ч. от респонденти); информационна система за пълен (безплатен и платен) достъп до статистически данни за различни години и в различни териториални обхвати. В системата на НСИ делът на администрацията е много по-малък от нормативно определения, тъй като структурните звена изпълняват по-скоро производствени, а не толкова административни и управленски функции и задачи. За периода 2000 - 2019 г. е налице тенденция към намаляване на броя на заетите лица в системата на НСИ. За периода от 2005 до 2014 г. се правят три оптимизации на персонала, който от 1 658 се съкращава на 1 130 души, или с 32%. През 2015 г. броят на заетите продължава да намалява, като вече е 968 и се задържа такъв през следващите години.

Заетите лица в системата на НСИ са експерти с висока квалификация и образование - през 2019 г. относителният дял на висшистите е над 86%. От специалистите с висше образование около 70% са с икономически специалности, а 42% от всички експерти по щат са статистици. За възрастовата структура на заетите лица в системата на НСИ е характерно, че преобладават служителите от по-високите възрастови групи, които са натрупали значителен практически опит.

За периода 2000 - 2019 г. броят на заетите лица в ЦУ на НСИ спада от 459 на 321. Много по-силно изразено е намалението в броя на заетите лица в ТСБ - от 1 254 през 2000 г. на 647 през 2019 година (фиг. 1). Регистрираното значително съкращаване на заетите в ТСБ (с 48.4%) за изследвания период е резултат от проведените реформи, реструктуриране и оптимизиране на дейността на ТСБ.

**Фиг. 1. Заети лица в системата на НСИ за периода 2000 - 2019 година**

### Заклучение

За своята 140-годишна история българската държавна статистика е постигнала значими резултати - проведени множество статистически изследвания на европейско и световно равнище, осигурен достъп и публичност на резултатите от тях. През този период в организационната структура на държавния статистически орган настъпват редица промени както в наименованието и подчинеността, така и в ресурсното осигуряване.

Доверието към официалната статистика е резултат от цялостната дейност и условията, в които се произвежда статистическата информация. Това се постига чрез строго спазване на методологията на статистическите изследвания, принципите на официалната статистика, Кодекса на европейската статистическа практика и прилагането на ефективна политика по разпространение на информацията и обслужване на потребителите.

Внедряването на постиженията на информационните технологии в производството и разпространението на статистическа информация поставя нови предизвикателства пред Националната статистическа система, свързани с нейното реструктуриране, оптимизиране и функциониране. Извършваните организационни, структурни, териториални и други промени създават условия за



повишаване на качеството на статистическата информация по стандартите на Евростат.

Новата организационна структура на НСИ отговаря на съвременните условия, кореспондира с визията и главната цел за развитие на Националната статистическа система и допринася за производството на по-качествена информация с по-малък обем ресурси - човешки и финансови. Извършването на проследяващ мониторинг на функционирането на системата след направените промени и въведените нови регионални структури на НСИ по отношение на системите за отчетност и статистическите услуги и дейности е необходимо условие, в резултат на което да се оценят ефектите от тях и готовността на системата да посрещне нови предизвикателства.



**ЦИТИРАНА ЛИТЕРАТУРА:**

**Балев, И.** (2005). 125 години българска държавна статистика. Статистика, кн. 2.

**Балевски, Д.** (1981). Сто години от създаването на българската статистика. Статистика, кн. 4.

**Къналиев, Т.** (2005). Законови, организационни и кадрови проблеми на статистическата информационна система у нас. Статистика, кн. 4.

**Мичев, Д.** (1984). Основни жалони в развитието на българската държавна статистика. В: *Сто години българска държавна статистика (1881 - 1981)*. София: КЕССИ.

**НСИ.** Устойчиво развитие на регионалната структура на Националната статистическа система: Реорганизация на регионалната структура на НСИ - туининг проект BG06/IB/F1/01, дейност 4.1.3. София: НСИ.

**НСИ** (2002, 2005, 2010, 2015). Годишен отчет на НСИ. София.

**НСИ** (2003). Концепция за реорганизация на регионалната структура на Националния статистически институт (проект BG2003/2004-937.12.02-CN1 в рамките на програма ФАР 2003). София: НСИ.

**НСИ** (2013). Стратегия за развитие на Националната статистическа система на Република България, 2013 - 2017 година. София: НСИ. Извлечено от [www.nsi.bg](http://www.nsi.bg).

**НСИ** (2015). Модернизиране на организацията и функционирането на НСИ.

**НСИ** (2015). Отчет за изпълнението на Националната статистическа програма и за дейността на НСИ. Извлечено от [www.nsi.bg](http://www.nsi.bg).

**НСИ** (2015). Отчет за изпълнението на Националната статистическа програма и за дейността на НСИ. София: НСИ. Извлечено от [www.nsi.bg](http://www.nsi.bg).

**Стефанов, И.** (1980). Основни жалони в развитието на българската статистика. Материали по история на българската статистика. София: КЕССИ.

## ОРГАНИЗАЦИОННАТА СТРУКТУРА НА НАЦИОНАЛНИЯ СТАТИСТИЧЕСКИ ИНСТИТУТ - СИНТЕЗ МЕЖДУ ТРАДИЦИИ И СЪВРЕМЕННОСТ

*Цветозария Гатева\**

**РЕЗЮМЕ** Управлението на обществото, икономиката и социалната сфера се основава на достатъчна по обем, обективна, разнообразна и детайлизирана информация. За да отговори на нарасналите потребности от надеждна, навременна и сравнима статистическа информация, статистическата институция извършва редица промени както в методологиите на изследванията, така и в организационната си структура. Статията има за цел да представи развитието на статистическата институция от създаването ѝ през

1880 г. до настоящия момент. Акцентира се върху основните проблеми и предизвикателства пред държавния статистически орган през различните периоди, неговата подчиненост, значимостта на извършваните промени в структурно и организационно отношение и постигнатите резултати.

**Ключови думи:** статистическа институция, статистическа система, Национален статистически институт, организационна структура, реорганизация, ресурсно осигуряване

**JEL:** C10, C81, C82

---

\* Началник на отдел „Статистически изследвания - Ловеч“, ТСБ - Северозапад, докторант в Стопанска академия „Д. А. Ценов“ - Свищов; e-mail: CGateva@NSI.bg.

## ОРГАНИЗАЦИОННАЯ СТРУКТУРА НАЦИОНАЛЬНОГО СТАТИСТИЧЕСКОГО ИНСТИТУТА - СИНТЕЗ ТРАДИЦИЙ И СОВРЕМЕННОСТИ

*Цветозария Гатева\**

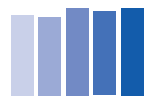
**РЕЗЮМЕ** Управление обществом, экономикой и социальной сферой основывается на объективной, разнообразной и подробной информации в достаточном объеме. Для удовлетворения растущих потребностей в надежной, своевременной и сопоставимой статистической информации статистическое учреждение вносит ряд изменений как в методологии исследований, так и в своей организационной структуре. Статья направлена на представление развития статистического учреждения с момента его создания в 1880 г. до настоящего времени. Акцентируется внимание на основных проблемах и задачах, стоящих перед органом государственной статистики в разные периоды, изменениях в его подчиненности, значимости перемен в структурном и организационном плане и достигнутых результатах.

**Ключевые слова:** статистическое учреждение, статистическая система, Национальный статистический институт, организационная структура, реорганизация, ресурсное обеспечение

**JEL:** C10, C81, C82

---

\* Начальник отдела „Статистические исследования - Ловеч“, ТСБ - Северозапад, докторант Академии народного хозяйства „Д. А. Ценов“ - Свищов; e-mail: CGateva@NSI.bg.



## THE ORGANIZATIONAL STRUCTURE OF THE NATIONAL STATISTICAL INSTITUTE - A SYNTHESIS BETWEEN TRADITION AND MODERNITY

*Tsvetozaria Gateva\**

**SUMMARY** The management of the society, economy and social sphere is based on sufficient in volume, objective, varied and detailed information. To meet the growing demands for reliable, timely and comparable statistical information, the statistical institution makes a number of changes both in research methodologies and organizational structure. This article aims to present the development of the statistical authority since its establishment in 1880 to the present. Emphasis is placed on the main issues and challenges facing the state statistical body during different periods, changes in its subordination, the significance of the changes in structural and organizational aspect and achieved results.

**Key words:** statistical authority, statistical system, National Statistical Institute, organizational structure, reorganization, resourcing.

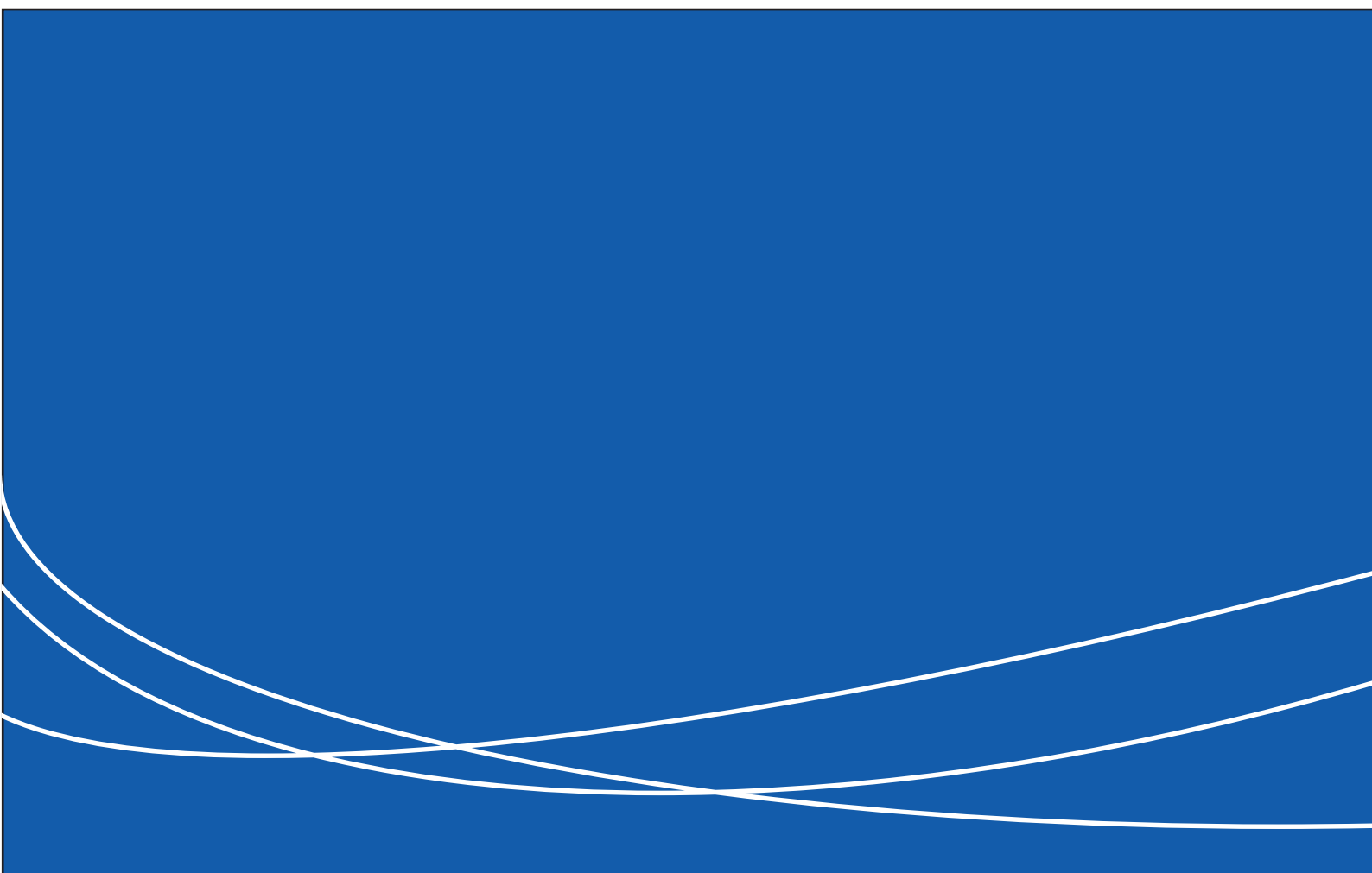
**JEL:** C10, C81, C82

---

\* Head of 'Statistical Surveys Department - Lovech', RSO - Severozapad, Ph.D. student at the Academy of Economics 'D. A. Tsenov' - Svishtov; e-mail: CGateva@NSI.bg.



**ИНФОРМАЦИИ, РЕЦЕНЗИИ,  
КОНСУЛТАЦИИ**







# МАТЕМАТИКАТА С ПОГЛЕД КЪМ СТАТИСТИКАТА

*Богдан Богданов\**



В архивите на сп. „Статистика“ се намира едно забележително есе на известния наш математик дмн. проф. Николай Янев<sup>1</sup>. Това есе е написано по повод Международната година на статистиката през 2013 година. Считам, че тази година по повод 141 години от създаването на българската държавна статистика полезно ще бъде отново да публикуваме това забележително есе. Причините и обясненията за това са накратко следните:

1. Да си припомним **съществуващата взаимозависимост** между математиката и статистиката, присъстващи в живота на стотици поколения от древността до наши дни.
2. Отново да акцентираме върху **фундамента на научни дисциплини**, важни за успешното развитие на всяка държава.
3. Да си припомним, че **успешното управление се основава на знанието** в основата, на което по принцип, са точните науки. В този смисъл математиката и статистиката са част от тях.
4. Да **проектираме бъдещето** като щрих на хоризонт, до който се достига с усвояване на инструментариума на точните науки.
5. Да отворим широко **прозорците на познанието** за идващите след нас.

\* Д-р, заместник-председател на Националния статистически институт; e-mail: [BBogdanov@nsi.bg](mailto:BBogdanov@nsi.bg).

<sup>1</sup> Дмн, професор, Институт по информатика и математика към Българската академия на науките; e-mail: [yanev@math.bas.bg](mailto:yanev@math.bas.bg).

Нямам съмнение, че младото поколение от представители на науката ги очакват предизвикателствата на резултатите от бурното развитие на информационните технологии. Тези от тях, които поемат по трънливия път на изследователи на социалните явления и процеси, ги очакват вълнуващи моменти. Тези моменти ще бъдат свързани с усвояването на нови знания, както и със стремежа им да оставят своя следа в света на науката, където информационните иновации ще бъдат неотменима част от живота на обществото. Най-добрите ще станат и бъдат уникални в един нов дигитален свят, където мечтите се сбъдват и стават реалност. Те ще бъдат и съществена част от новите лидери, които ще построят един по-добър и справедлив свят.

*Богдан Богданов*

Сп. „Статистика“, бр. 3 - 4/2012

## МЕЖДУНАРОДНА ГОДИНА НА СТАТИСТИКАТА - 2013<sup>2</sup>

*Николай Янев*

Обявяването на 2013 година от ООН за Международна година на статистиката (МГС) ни дава повод отново да се върнем към обсъждане на въпросите, отнасящи се до същността на Статистиката, нейния предмет, основни цели и задачи и обхвата на нейната приложимост. Това обяснява желанието на автора да стъпи на една по-широка платформа, като се постарее с известна доза атрактивност да привлече вниманието към проблемите и значението на Статистиката. Така че на всичко това би могло да се даде и подзаглавие „МГС през погледа на един математик“. Надявам се, че направените разсъждения ще бъдат интересни и полезни и за читателите на списание „Статистика“.

Във връзка с това да припомним, че 2000 година бе обявена за Международна година на Математиката. Това, че 21 век и въобще третото хилядолетие бяха посветени на Математиката, е не само признание за „царицата на науките“, но и изразява надеждата на човечеството за едно по-добро и справедливо бъдеще, основано на науката, в която Математиката е най-светлият символ на истинност, честност, справедливост и универсалност.

Последваха години, посветени и на други клонове на науката, дали своя принос в еволюцията на човешката цивилизация, за да стигнем до датата 20.10.2010 г., когато тържествено бе отбелязан Световният ден на Статистиката. Двадесет и първи век беше наречен век на информационното общество, а Статистиката играе много важна роля в него. Видя се, че един ден е малко, за да се разгледа всестранно тази роля и да се отдаде нужното внимание на Статистиката като част от Математиката и на нейните многобройни приложения, станали основа на много различни научни направления и всекидневни важни изводи, свързани с управленски прогнози и решения.

---

<sup>2</sup> Статията е подготвена въз основа на пленарен доклад, изнесен пред 42-рата пролетна конференция на Съюза на математиците в България. Бел. на редактора: В текста съзнателно е допуснато изписването с главни букви на наименованията на научните дисциплини.

Така естествено се стигна до определянето на 2013 година като Международна година на Статистиката (International Year of Statistics). Може би защото според Статистиката 13 не е фатално число?! Във всеки случай, ще имаме на разположение цяла година, през която се организират многобройни и разнообразни прояви за осмисляне на значението и ролята на Статистиката, на нейните постижения и приложения, за популяризиране на интереса към нейното по-задълбочено изучаване и по-всеобхватно използване, за нейното по-нататъшно развитие, свързано с развиващото се информационно общество. Разбира се, Съюзът на математиците в България (СМБ) не може да остане настрана от световните тенденции и този доклад е една от набелязаните прояви. И все пак, какво е това Статистика? Какъв е нейният предмет на дейност, нейните основни цели и задачи, обхватът на нейната приложимост? Кога, как и какво да се изучава? С други думи, как да се интегрира Статистиката в рамките на съществуващите образователни системи? Какво е положението в тази област у нас и по света? Как трябва да се развива Статистиката като наука? А кой и как трябва да се занимава с нейните приложения? За да не стигаме до популярния парадокс, с който най-често се осмива неправилното (понякога преднамерено) използване на статистиката: „Има три вида лъжа - обикновена лъжа, нагла лъжа и ... статистика“. А какви ли не епитети се срещат още: придворна статистика, подвеждаща статистика, поръчкова статистика, обслужваща статистика, погрешна статистика, „изсмукана от пръстите“ статистика и т.н.

Разбира се, да се отговори на тези въпроси в рамките на едно кратко есе не е лесно (колкото по-кратко, толкова по-трудно). И все пак, какво е това Статистика? За да отговорим на този въпрос, първо трябва да отбележим, че Статистиката има няколко лица, т.е. този термин има различни значения. За математиците Статистиката е математическа дисциплина, която в известен смисъл решава обратната задача на Теорията на вероятностите (ТВ). За да подчертаят това, те обикновено я наричат Математическа статистика (МС). Често тяхното обединение се нарича Стохастика, т.е.  $ТВ + МС = Стохастика$ . Това ще се опитаме да прецизираме по-нататък. За обикновения човек Статистика означава най-често някакви числови данни, от които се правят някакви изводи. И тук е „тънкият момент“, т.е. има две различни дейности - събиране на информация за даден обект или явление (статично или динамично) и изводи от събраната информация. Това също ще уточняваме по-нататък. Тук обаче трябва да отбележим, че терминът „статистика“ е добил гражданственост, особено на Запад, като една търговска марка, като събирателно понятие за всички тези математически и недотам

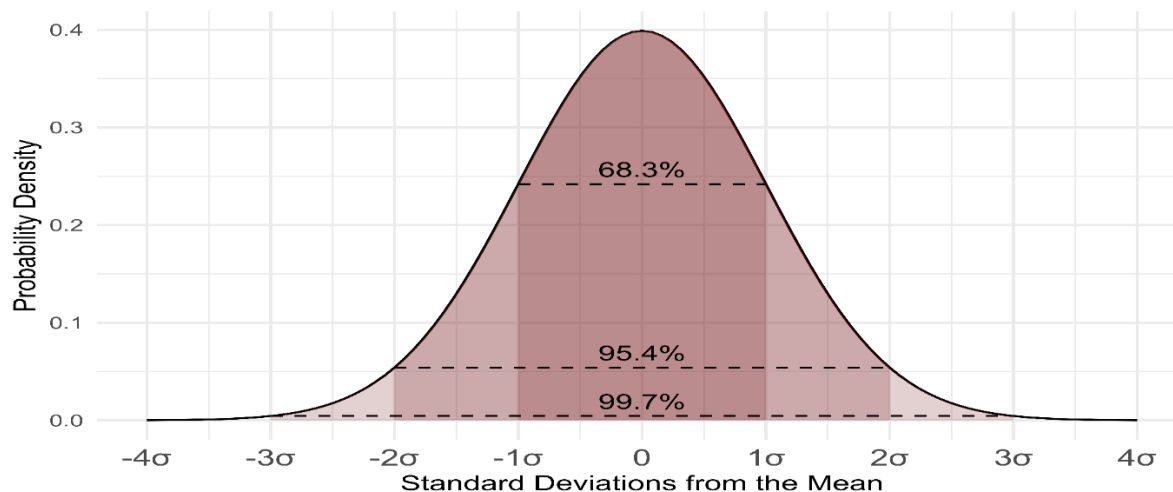
математически дейности. Така че като честваме Международната година на статистиката, ние трябва да разбираме Статистиката в цялата нейна многостранност.

Във всеки приличен университет днес има задължително катедра (или департамент) по Математика (където понякога се занимават и с ТВ и МС), а отделно има катедра (или департамент) по Статистика (където задължително се занимават с ТВ и МС). Освен това в по-големите университети, които имат медицински центрове, са създадени големи департаменти по Биостатистика. Към икономическите факултети (или центрове) има департаменти по Икономическа статистика и т.н. Така че в общественото пространство фигурират съчетания като медицинска статистика, инженерна статистика, икономическа статистика, държавна статистика, селскостопанска статистика, промишлена статистика, военна статистика, финансова статистика, борсова статистика, банкова статистика, търговска статистика (външна или вътрешна), макроикономическа статистика, спортна статистика и т.н. А всеки ден от медиите ни заливат с какви ли не потресаващи статистики (убийства, катастрофи, пожари, наводнения, бедствия, епидемии, болести, смъртност, корупция, кражби, безработица и т.н.). Има, разбира се, и „приятни“ статистики (раждания, бракосъчетания, ръст на БВП, спестявания, покупателна способност, ръст на заплати и пенсии, пътувания в чужбина и т.н.). Но като че ли вторите се губят в потока на първите (добрата вест не е новина за съвременната журналистика). Както отбелязахме, живеем в информационно общество. И както казват някои млади хора, няма защо да се учи: каквото ни трябва, намираме го в интернет. Решихме да проверим и написахме думата „Статистика“ в една от търсачките. Ето какво намерихме:

### **„Статистика**

*от Уикипедия, свободната енциклопедия*

Диаграма на **нормално разпределение**, често използвано в статистиката



**Статистиката е математическа дисциплина**, която изучава добиването на информация чрез анализ и интерпретация на емпирични данни, използвайки **теорията на вероятностите**. Статистическата дейност включва също планирането и организирането на събирането на данни чрез проучвания и експерименти. Статистиката възниква във връзка с нуждите на емпиричната наука и се отличава от повечето клонове на математиката по своята приложна насоченост. Статистически методи се прилагат в широк кръг области като природните и обществените науки, държавното управление и бизнеса. Една от основните подобласти на статистиката е **описателната статистика**, която се занимава с обобщаването на систематизирани данни. То е от особена важност при емпиричните изследвания и описването на резултати от експерименти. Чрез методите на статистиката данните могат да бъдат анализирани, като се отчита случайността и несигурността на наблюденията, и въз основа на това да се правят изводи за съдържащи се в тях закономерности“.

Не зная кой е авторът на това описание, но то е доста неточно и непълно. И докато непълнотата може да бъде оправдана с желание и необходимост за краткост, то грешките могат да подведат непредубедения читател. Още първото изречение съдържа невярна информация, като определя Статистиката като математическа дисциплина. Както беше отбелязано, само една част от Статистиката, а именно Математическата статистика, е безспорно част от Математиката. Защото, когато срещаме например изрази като „статистика на кражбите“ или „борсова статистика“, те нямат нищо общо с Математиката. Те са просто съвкупности от данни, които също е прието да се наричат „статистика“ (при това най-често с някакво прилагателно). Всъщност етимологията на думата статистика е свързана с думите „state“ (държава, състояние) и „status“ (състояние,



положение). В този смисъл зараждането на статистиката може да бъде отнесено към Древния Китай, където още преди 5 000 години са започнали да събират сведения за състоянието на държавата. Ясно изразен образец на държавна статистика можем да намерим в Римската империя. Между другото, това е един от основните признаци на държавността. Днес всяка държава има специална институция с най-различни названия (министерство, агенция, институт, комитет, бюро, комисия), която се занимава с т.нар. държавна (или официална) статистика.

У нас държавната статистика започва да се развива от 1880 г. в Княжество България с учредяването на Статистическото отделение към Министерството на правосъдието. Оттогава, независимо от организационните промени на статистическата институция, тя събира, обработва и съхранява важна статистическа информация за демографските и икономическите процеси в страната. Сега статистическата дейност на държавата се осъществява главно от Националния статистически институт (НСИ), който е основан със Закона за статистиката, приет през 1991 г. от Народното събрание. По-подробно с историята, визията, целите, устройството, организацията и въобще с всички дейности на НСИ любознателният читател може да се запознае на интернет страницата на НСИ: <http://www.nsi.bg/index.php>. И ако не сте посещавали тази страница, ще останете приятно изненадани да разберете колко интересна информация има и каква огромна дейност се развива в областта на държавната статистика. Ето някои извадки: „Главната цел на развитието на НСС (Националната статистическа система) през периода 2008 - 2012 г. е да се усъвършенства разработването, производството и разпространението на статистическа информация за всички групи потребители чрез подобряване на институционалния капацитет на системата и ускорено внедряване на съвременни информационни и комуникационни технологии в условията на пълноправно членство на Република България в ЕС.

Официалната статистическа информация - продукт на НСС, е предназначена за анализи и научни изследвания, за планиране, прогнозиране и вземане на управленски решения на микро- и макроравнище. Статистическите информационни продукти и услуги са насочени към: органите на държавно управление и местната власт; бизнеса и неправителствените организации; научната и академичната общност; медиите; ЕС, ООН, МВФ и други международни организации; отделни граждани. Развитието на НСС през периода 2008 - 2012 г. предполага осъществяването на следните **основни цели**:

- Развитие на методологията и обогатяване на съдържанието на съществуващите статистически изследвания и внедряване на нови изследвания и показатели в

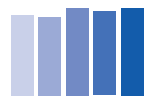


съответствие с нуждите на потребителите и приоритетите на Статистическата програма на ЕС 2008 - 2012 година.

- Усъвършенстване на производството, разпространението и съхранението на статистическа информация.
- Подобряване на средата, информационната инфраструктура и ресурсната осигуреност на НСС“.

Събирането на достоверна статистическа информация е важна и отговорна задача, но само по себе си това би било безполезно, ако получените статистически данни не се анализират, т.е. от тях да се получат съответните достоверни изводи и прогнози. А това вече не е възможно без използване на методите на Математическата статистика.

Както беше отбелязано, МС решава обратната задача на ТВ. Образно казано, те са като двете страни на една монета. Да се опитаме да изясним тази връзка, защото тя е много важна. Най-общо казано, предмет на ТВ и МС е математическият анализ на понятието „случайност“. ТВ и МС, или с други думи, Стохастиката, е тази част от Математиката, която създава и изследва математически модели на случайни явления, величини и процеси. Тук е моментът да припомним, че едно от големите постижения на ХХ век беше аксиоматизацията на ТВ, предложена от А. Н. Колмогоров през 1933 г. (VI проблем на Хилберт). Последвалото бурно развитие на ТВ превърна тази част от Математиката в естествен фундамент на цялото естествознание, защото случайността лежи в основата на самата природа и колкото повече се приближаваме до микросвета, толкова това става по-ясно. Съгласно Колмогоровската аксиоматика основните математически обекти в ТВ (случайни величини и случайни процеси) се разглеждат в т.нар. вероятностни пространства, където вероятността е определена като една неотрицателна, нормирана и адитивна мярка. По този начин в услуга на ТВ е впрегнат практически целият съществуващ математически арсенал, което естествено води до бурно развитие на самата ТВ. От друга страна, построяването на конкретно вероятностно пространство означава създаването на определен математически модел за даден стохастичен феномен. Това стимулира изключително многобройните и разнообразни приложения на ТВ. И тук естествено стигаме до МС. Образно казано, ТВ е изкуство за пресмятане на едни вероятности посредством други, т.е. при зададени начални разпределения на вероятностите се описват вероятностните разпределения на последващи по-сложни събития и процеси. А МС решава обратна в известен смисъл задача, т.е. по наблюденията над събитията, случайните величини или случайните процеси да оцени тези първоначални разпределения, чрез които се описва



математическият модел. Чрез МС се проверяват и хипотезите за адекватност, т.е. доколко математическият модел е близък до реалната действителност. При това основните методи на доказателство и изследване се основават на ТВ, като се използва, разбира се, и целият останал математически апарат.

Например, важна роля играят редица оптимизационни и апроксимационни методи. Може би е добре да илюстрираме тези идеи с един несложен реален пример.

**Задача 1.** При статистически качествен контрол се прави случайна извадка с обем  $n$  от партида с еднородни изделия. Известно е, че вероятността всяко отделно взето изделие да бъде дефектно, е  $p$ . Каква е вероятността в извадката да има точно  $k$  дефектни изделия? Това е една типична задача от ТВ. В този случай не е трудно да се покаже, че случайната величина  $X$ , равна на броя на дефектните изделия в извадката, има биномно разпределение  $Bi(n, p)$ , т.е. вероятностите  $P(k) = P\{X = k\}$  са т.нар. биномни вероятности,  $k = 0, 1, 2, \dots, n$ . Ето как би изглеждала съответната ситуация от гледна точка на МС.

**Задача 2.** При статистически качествен контрол се прави случайна извадка с обем  $n$  от партида с еднородни изделия. При прегледа се оказва, че сред тях има точно  $k$  дефектни изделия.

а) Как да определим неизвестната вероятност  $p$  всяко едно отделно взето изделие да бъде дефектно?

б) Ако  $p^*$  е оценка за тази вероятност, как да определим точността на тази оценка?

в) Как да проверим хипотезата, че бракът е под 1%?

г) Как да различим две хипотези, при които бракът е съответно 2% или 5%?

Не е трудно да се предложи  $p^* = k/n$  (често наричана честота) като една естествена оценка за  $p$ . Отговорът на останалите въпроси обаче не е толкова очевиден и изисква прилагането на определени статистически методи.

Съвременното развитие на ТВ и МС се обуславя от два мощни фактора: вътрешното развитие на теорията и непрекъснатия приток на нови и все по-сложни проблеми от заобикалящата ни действителност. Така са създадени редица нови направления както в самата математика, така и в колаборация с други науки като например Стохастичен анализ, Стохастични диференциални уравнения, Стохастична геометрия, Стохастични матрици, Финансова математика, Теория на риска и надеждността, Статистическа механика, Статистическа физика, Биостатистика, Икономическа статистика, Статистически контрол на качеството, Статистическа демография и т.н.

Заобикалящият ни свят и общество са стохастични по своята същност. Поведението на индивида или обществото не могат да бъдат описани с детерминистични функции, те имат ярко изразен стохастичен характер<sup>3</sup>. И тук стигаме до един от големите парадокси на човечеството: нашето възпитание по своя характер е строго детерминистично и това води до конфликти със заобикалящия ни стохастичен свят. На нас от малки ни втълпяват, че всичко е свързано в причинно-следствени връзки, пренебрегвайки изобщо категорията **случайност**.

По-възрастните сигурно си спомнят, а по-младите може би са чували за периода, когато под влияние на идеите на академик Лисенко в СССР и т.нар. социалистически лагер бяха разгромени редица научни направления, обявени като „буржоазни“ (и на първо място генетиката). Тогава беше издигнат и лозунгът: „Науката е враг на случайността!“ Под угроза за забрана са не само ТВ и МС, но и всички клонове на науката, където те се прилагат. Математиците Хинчин и Колмогоров обаче остроумно и своевременно издигат контратезата: „Да, науката е враг на случайността, но врагът трябва да се изучава. А това прави Теорията на вероятностите“. Това се харесало на диктатора Сталин, създател на тезата за „враговете на народа“, които били изпращани в концлагери. Казват, че всмуквайки от знаменитата си лула, Сталин заявил: „Правилно! Врагът трябва да се изучава!“ И това спасило ТВ и МС от пълен разгром, заедно с редица други области на науката.

Сега може да ни се струва нелепо, но тогава много учени са заплатили с живота си само за това, че са били истински учени. Този пример е ярко доказателство за вредата от фундаменталния детерминизъм, който се проявява в редица религиозни норми и ненаучни идеологии, донесли толкова беди на човечеството. И тук естествено стигаме до въпроса за стохастичното образование и възпитание, защото стохастичната култура е важна част от общата култура на индивида и нацията. Днес практически в почти всички университетски специалности се изучават в различна степен ТВ и МС. В развитите страни елементи на стохастиката са станали вече неизменна част от средното образование по математика, като в някои страни то започва от началните класове или дори от детските градини. В това отношение ние значително изоставаме от водещите световни тенденции.

---

<sup>3</sup> Вж. Б. Димитров и Н. Янев. Вероятности и статистика. Издателство на СУ - първо изд. 1989, второ изд. 1996; Софттекс - трето изд. 2007, София.

В процеса на еволюцията у човечеството постепенно се развива стохастична интуиция вследствие на дългите наблюдения и опит, свързани с лова, земеделието, риболова, предсказването на времето и т.н. Човекът е свикнал да играе хазарт с природата и неслучайно хазартните игри са се появили от най-дълбока древност. Всъщност за начало на ТВ се счита 1654 г., когато започва кореспонденцията между Ферма и Паскал по повод на някои въпроси, поставени от кавалера де Мере и свързани с хазартните игри. По-нататъшното развитие на стохастиката е свързано с имената на Галилей, Хюйгенс, Моавр, Лаплас, Гаус, Поасон, Коши, Лобачевски, Чебишев, Марков, Ляпунов, Борел, Бернщайн, Леви, Хинчин, Винер, Фишер, Нейман, Пирсън, Гнеденко, Фелър, сред които отбелязахме вече фундаменталния принос на Колмогоров (1903 - 1987 г.), наричан често „баща на съвременната стохастика“.

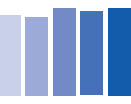
Съвременната МС е силно свързана с развитието на информатиката. Сега всеки потребител на статистическа информация разполага с мощни пакети от статистически процедури, с които лесно и бързо могат да бъдат решени редица стандартни статистически задачи. Това доведе до развитието на нов дял от МС, наричан често компютърна статистика.

Като се връщам към спомените си за Световния статистически конгрес през далечната вече 1994 г., организиран от ISI (International Statistical Institute) в Чепълхил, Северна Каролина, не мога да не отбележа визитата в кампуса на компанията SAS, най-големият производител на статистически софтуер. По това време, както ни беше казано, там са работели над 3 000 души, които заедно със семействата си представляват едно малко модерно градче, разположено сред вековни гори и наброяващо над 10 000 жители. В демонстрационната зала на централния офис имаше над 150 монитора, на всеки от които се представяха последните постижения на фирмата в различни области на статистиката. Беше истинско удоволствие да се видят високопрофесионалните дискусии между участниците в конгреса (предимно статистици теоретици) и демонстраторите (специалисти по компютърна статистика). От някои участници в конгреса бяха изказани пожелания за разширяване на възможностите на съществуващия софтуер или предложения за нови статистически процедури. Това доведе моментално до предложения за сключване на договори за развитие на тези идеи. Много участници в конгреса получиха покани (сред тях и авторът) да изпратят своите доклади и други научни статии с цел евентуална разработка на нови статистически програми (авторът трябва да признае, че не се е възползвал от тази възможност). Това беше и ярка

демонстрация на сътрудничество между наука и бизнес. Всъщност SAS беше един от главните спонсори на конгреса.

Възможностите на съвременните компютри за бързо симулиране на случайни величини и процеси доведе до развитието на много нови дялове от статистиката и статистическото моделиране. По този начин практически всяка задача може да намери едно сравнително добро приближено решение. Особено силно се развиха т.нар. методи Монте Карло за пресмятане на интеграли, диференциални и интегрални уравнения. С подробности за организацията на честванията на Международната година на статистиката по цял свят любознателният читател може да се запознае на интернет сайта <http://www.statistics2013.org/>. На него са регистрирани над 2 000 научни организации, свързани със статистиката, много университети, научноизследователски институти, обществено-професионални статистически организации и други. България е добре представена в различните раздели, като трябва да отбележим преди всичко участието на ИМИ-БАН, НСИ, ФМИ-СУ, БСД (Българско статистическо дружество) и други.

Изумително е огромното количество хора и организации в цял свят, заети със статистическата информация в цялата гама от аспекти. Така например според някои оценки само в САЩ с ТВ и МС се занимават между 60 и 70% от професионалните математици. И въобще в САЩ бизнесът започва със статистика, развива се със статистика и прави финални заключения пак с помощта на статистиката. Всичко това естествено рефлектира върху сериозното количество издавана специализирана литература по Стохастика и съответните периодични списания с фундаментален или приложен характер. Интересно, че в сериозните списания по експериментални науки вече не се приемат статии, в които няма съответни статистически модели и изводи. За съжаление, стохастичната колегия у нас е твърде малка дори за мащабите на нашата страна, особено в областта на приложната статистика. Очертава се сериозна ниша, в която, да се надяваме, ще се появят млади и способни кадри. Един доказан начин за това е въвеждането на по-ранно и задълбочено запознаване с Вероятности и Статистика в средните училища. Да се говори и пише за Статистиката (особено пред професионалисти) не е нито лесно, нито просто, особено ако се търси по-нестандартен ракурс. От друга страна, всеки от нас има вече някакви изградени представи, а предложената тема е толкова широка, че спокойно може да обхване съдържанието на една дебела книга, където всички тези неща могат да бъдат разгледани по-подробно, в исторически аспект и с многобройни примери.



Целта на това кратко есе е по-скоро да представи някои въпроси и тезиси за размисъл и дискусии. Липсата на цитирана литература се обяснява просто с факта, че авторът не е ползвал такава (с изключение на това, което е останало в съзнанието от някъде-някога прочетеното). Останалото са собствени мисли, нахвърляни „експромпт“, не без известно удоволствие, защото да се говори и пише за Статистиката е наистина удоволствие. Разбира се, Международната година на статистиката е добър повод за сравнения, анализи и изводи, които ще бъдат последвани от дискусии и други мероприятия.

## МАТЕМАТИКАТА С ПОГЛЕД КЪМ СТАТИСТИКАТА

*Богдан Богданов\**

**РЕЗЮМЕ** Информацията „Математиката с поглед към статистиката“ съдържа статията на д-р проф. Николай Янев, публикувана в сп. „Статистика“ през 2012 г., бр. 3 - 4 със заглавието „Международна година на статистиката - 2013“. Тази информация се представя на страниците на сп. „Статистика“ във връзка с 141-годишнината от създаването на държавната статистика в България. Нейната актуалност е особено важна за хората, приели статистиката за своя професия. В същото време информацията притежава и значителна познавателна стойност за читателите на списанието.

Статията съдържа интересни разсъждения на проф. Янев за същността на статистиката като наука и нейната връзка с математиката. Представена е значимата роля на статистиката като теория и практика за обществото и държавата. Подчертава се високата научна стойност на статистиката за развитието на информационното общество.

Статията е представена с предисловие, написано от д-р Б. Богданов.

---

\* Д-р, заместник-председател на Националния статистически институт; e-mail: [VBogdanov@nsi.bg](mailto:VBogdanov@nsi.bg).

## МАТЕМАТИЧЕСКИЙ ВЗГЛЯД НА СТАТИСТИКУ

*Богдан Богданов\**

**РЕЗЮМЕ** Информация „Взгляд математики на статистику“ содержит статью д-ра мат. наук, проф. Николая Янева, опубликованную в журнале „Статистика“ в 2012 году, в 3 и 4 номерах, под заголовком „Международный год статистики - 2013“. Эта информация представлена на страницах журнала „Статистика“ в связи с 141-й годовщиной создания государственной статистики в Болгарии. Ее актуальность особенно важна для людей, которые выбрали для себя профессию статистика. В то же время, информация имеет значительную познавательную ценность для читателей журнала.

Статья содержит интересные рассуждения профессора Янева о сущности статистики как науки и ее связи с математикой. Представлена значительная роль теоретических и практических аспектов статистики для обеспечения потребностей общества и государства. Подчеркивается высокая научная ценность статистики в развитии информационного общества.

Предисловие к статье написано доктором Б. Богдановым.

---

\* Д-р, заместитель председателя Национального статистического института; e-mail: BBogdanov@nsi.bg.



## MATHEMATICS WITH A LOOK AT STATISTICS

*Bogdan Bogdanov\**

**SUMMARY** The information “Mathematics with a look at statistics” contains the article by D.S. Prof. Nikolay Yanev, published in the “Statistics” magazine in 2012, issue. 3 - 4 with the title “International Year of Statistics - 2013”. This information is presented on the pages of the “Statistics” magazine in connection with the 141st anniversary of the establishment of state statistics in Bulgaria. Its relevance is especially important for people who have adopted statistics as their profession. At the same time, the information has a significant cognitive value for the readers of the magazine.

The article contains interesting reflections of Prof. Yanev on the nature of statistics as a science and its connection with mathematics. The significant role of statistics as a theory and practice for society and the state is presented. The high scientific value of statistics for the development of the information society is emphasized.

The article is presented with a preface written by Dr. B. Bogdanov.

---

\* Dr., Deputy President of the National Statistical Institute; e-mail: BBogdanov@nsi.bg.

## ПЪРВО ПРЕБРОЯВАНЕ НА ИНДУСТРИАЛНИТЕ И ТЪРГОВСКИТЕ ЗАВЕДЕНИЯ В ЦАРСТВО БЪЛГАРИЯ ПРЕЗ 1926 ГОДИНА

*Соня Златанова\**



В доклад № 4001 от 17 март 1926 г. до Върховния статистически съвет са изложени мотивите за допълване на демографското преброяване с преброяване на стопанските предприятия.

Върховният статистически съвет одобрява проекта на Дирекцията на статистиката за обединяване на демографското, жилищното и стопанското преброяване, като приема единодушно следното решение: „Едновременно с общото демографско преброяване по досегашния кадър да се предприеме и преброяване на селското стопанство и занаятчийско-индустриалното производство със специални карти, чрез които да се съберат най-общи сведения за установяване количествено и качествено, от една страна земеделческите и от друга - всички останали неземеделчески стопанства и предприятия в нашата страна“.

На 31 декември 1926 г. се провежда първото общо преброяване на индустриалните и търговските предприятия в Царство България едновременно с преброяването на населението и жилищния фонд. Извършва се посредством карта „З“ за неземеделските, търговските и индустриалните заведения.

---

\* Младши експерт в отдел „Публикации, библиотека и дигитални продукти“, Национален статистически институт; e-mail: SZlatanova@nsi.bg.

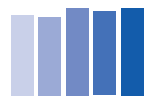
Едновременното провеждане на демографското и стопанското преброяване през 1926 г. има своите предимства и недостатъци. Оптималните моменти за двете преброявания не съвпадат. За преброяването на населението такъв момент е през зимата, когато населението е оседло, а за индустриалното - през лятото, когато дейността е засилена. Поради сливането на трите преброявания - демографско, индустриално и селскостопанско - сезонните индустрии или са пропуснати, или са регистрирани с намален персонал. Повечето от несезонните индустрии са с намален капацитет на дейност към момента на преброяването. Като по-сложни стопанските преброявания би трябвало да продължават по-дълго време и да се извършват от много добре обучени агент-преброители. При сливането на трите преброявания агентите са претрупани с работа, поради това те нямат възможност да се задълбочат в отделните обекти. Това обстоятелство се отразява на качеството на събраните сведения. Попълването на индустриалната карта е в пряка връзка с отговора за професията в личната демографска карта „А“ и се извършва в жилището на собственика на индустриалното или търговското заведение, а не в самото заведение. Поради тази причина се появява опасност много предприятия или да бъдат пропуснати, или да бъдат преброени два пъти.

От друга страна, сливането на преброяванията дава възможност за взаимна проверка между различните карти, особено на въпроса за професията в демографската карта със сведения за характера на стопанската дейност. Разходите при преброяванията при една и съща организация са значително по-малки, отколкото биха били при отделни преброявания. Последното съображение е надделяло при вземането на решението за обединяване на преброяванията.

### **Обект на преброяването с карта „З“**

Стопанските субекти, преброени с карта „З“, са:

- Индустрия, занаятчийство, домашна индустрия (работа вкъщи за своя сметка или за сметка на даден индустриалец или търговец);
- Рударство, каменни, глинени и други кариери и солници;
- Строителство, вкл. строителни дружества, архитектурни, инженерни, земемерни бюра, бояджийство на сгради;
- Инсталиране на електричество и водопроводи;
- Търговия;
- Банкерство, спестовни каси и други подобни;



- Осигурителни дружества, агенции за осигуряване, болнични каси, посмъртни каси, вкл. касите за злополука, безработица и др. при професионалните сдружения;

- Транспортно дело, експедиция, пътнически превоз и паспортни бюра, митнически комисионери;

- Хотели, кръчми, гостилници, пансиони;

- Болници, бани, лаборатории и други подобни;

- Театри, опери, кинематографии;

- Вестникарство, издателства;

- Пчеларство, бубарство, скотовъдство, млекоарство и други производства от този род, несвързани със земеделието.

Въз основа на практиката на германската статистическа служба, по-точно в резултат на проведеното през 1925 г. преброяване в Германия, като частни случаи към обектите на карта „З“ са отнесени следните стопански субекти:

- Неземеделско градинарство и цветарство в специални помещения, развъждане на гъби в парници и др.;

- Скотовъдство и птицевъдство;

- Пчеларство, бубарство и дърварство;

- Въглищарство;

- Казаните за производство на ракия са броени с карта „З“ само когато собствениците им са купували материал, за да варят с търговска цел, или са предоставяли казана на други стопани срещу възнаграждение;

- Притежатели на вършачки, роначки, триори и др., когато са експлоатирали тези машини с търговска цел и редовно са извършвали услуги на други лица срещу възнаграждение.

Освен изброените стопански субекти, преброени с карта „З“, са регистрирани и частни случаи, разпределени в следните категории:

- Търговски складове, които са имали персонал или двигателна сила;

- Театри и други развлекателни заведения без значение дали са частни, общински или държавни;

- Образователни заведения - всички учебни заведения, експлоатирани от частни лица с цел формиране на доход, както и всички професионални и технически училища и курсове;
- Железопътни, телеграфо-пощенски и въздухоплавателни работилници и радиостанции;
- Работилници при затворите и войсковите части.

### **Единица на наблюдението**

До Първата световна война са водени много спорове в международен мащаб относно единицата на наблюдение при промишлените преброявания. Едва след войната се наблюдава доближаване на схващанията в различни страни. Международният статистически институт с резолюция от 1925 г. препоръчва като единица на производствените преброявания „заведението“. Най-пълно въпросът е разрешен от германската статистическа служба по време на преброяването в Германия през 1925 година. На практика са предвидени три вида разработки на събраните сведения с три различни единици на броене:

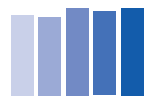
1. Разработка по стопански единици - предприятията, които се разглеждат като съвкупност от стопанската дейност на едно физическо или юридическо лице, доколкото то за своя сметка и на собствен риск посредством едно или повече заведения произвежда блага за задоволяване на чужда потребност.

Чрез тази разработка на данните се цели да се отговори на въпроса за концентрацията на стопанския живот в страната според правната форма, в която се извършва стопанската дейност.

Всяко предприятие може да се състои от много подразделения, разположени на различни места. Всяко такова подразделение се разглежда като клон на предприятието. Като клонове се разглеждат и съставните части на едно предприятие, които коренно се различават по вида производство, дори когато не са разделени по места. Всички предприятия без клонове и всички клонове и централи на предприятията с клонове представляват отделни „заведения“.

2. Втората разработка на германското преброяване е по териториален принцип на предприятия, като целта е да се получи ясна представа за географското разпределение на неземеделската стопанска дейност.

3. Единица на наблюдение в третата разработка е техническата единица, която е основата на стопанския живот на страната. Това са ателието, отделението,



работилницата и други. Тази разработка дава най-подробна представа за техническата структура на стопанския живот.

Главната дирекция на статистиката, опирайки се на опита на Германия, възнамерява да осъществи две разработки - по предприятия и по заведения. Оказва се, че това намерение не е осъществимо главно поради избрания начин за попълване на картите. Да се обединят всички заведения на една и съща фирма, за да образуват едно предприятие, не е възможно на този етап. Поради тази причина резултатите от преброяването се предоставят само в една разработка - по заведения.

В структурата на българското стопанство се оказват няколко тясно свързани помежду си стопански субекта, но коренно различни като видове стопанска дейност, упражнявани на едно място. За тях Дирекцията на статистиката включва специални параграфи във възприетата номенклатура и разглежда тези комбинации като едно заведение. Такива случаи се отнасят за бакалници и кафенета, кръчмарници и бръснарници, кръчми и касапници и други. Тези комбинации са срещани най-вече по селата в Царството.

### **Домашна индустрия**

Особена категория образуват заведенията от *домашната индустрия*. Обективните признаци за принадлежност към тази категория са:

- Извършването на дейността по поръчка или предназначена за търговия в жилищното помещение, така че не може да се разграничи работилницата от жилището;
- Извършване на дейността в двора или в отделна стая на жилището, в селата или по-малките градове, за пазара или за търговски посредник;
- Извършване на дейността за търговец или посредник в специално помещение.

Така дефинирано, понятието за домашна индустрия е много по-широко от приетото в западноевропейските статистики, където домашната индустрия обхваща само дейностите за предприемач или посредник. Възприетата от българската статистика разработка на домашната индустрия, давайки отделни сведения за заведенията, работили за собствена сметка или финансирани за външна поръчка, позволява лесно да се правят международни сравнения.

### Момент на преброяването

Преброяването, извършено с карта „З“, няма критичен момент. Броят на заетите лица е показан към края на 1926 година. Ако в деня на преброяването някое предприятие е преустановило дейността си поради стачка, локаут или други причини, събраните сведения са за състоянието преди причината за прекратяване на дейността. Ако предприятието не е работило поради сезонни причини, данните, попълнени в картата, се отнасят до състоянието му по време на преброяването.

Картите се попълват в жилището на собственика, но данните се отнасят по местонахождение на предприятието, а данните от картите на амбулантните търговци и занаятчиите се отнасят към населеното място, където те са заварени в момента на преброяването.

### Заети лица

Заетите лица са разпределени по пол и положение в предприятието в следните категории:

- Стопани, заети в предприятието;
- Членове на производителните кооперации;
- Управители, директори и други висши административни служители;
- Търговски и канцеларски персонал (прокуристи, началници на отдели, чиновници и др.);
- Технически ръководен персонал (инженери, кондуктори, машинисти, майстори специалисти и др.);
- Работници;
- Членове на семейството на собственика на стопанския субект. Това е категорията, в която са допускани най-често грешки от агент-преброителите. Тук трябва да фигурират само онези членове на семейството, които не могат да бъдат отнесени в други категории.

Главната дирекция на статистиката се е стремела да избегне двойното броене на лицата, затова заетите в две индустриални предприятия или в селското стопанство са регистрирани само там, където е главното им занятие. В дребните предприятия, където работи собственикът, но той има друго главно занятие, предприятието е отнесено в групата на предприятията с нула заети лица.

## **Впрегатен добитък и превозни средства**

На въпроса за впрегатния добитък в карта „З“ са получени много неточни отговори поради това, че добитъкът, използван в селското стопанство, е объркан с този, използван в индустриалното или търговското предприятие. Поради тази причина отговорите са послужили само за разпределение на предприятията на групи „със“ или „без“ впрегатен добитък. В подобна степен това важи и за въпроса относно употребяваните транспортни средства.

## **Двигателни машини**

В този раздел са регистрирани:

- Броят и номиналните ефективни конски сили на двигателя по вид;
- Броят и консумираните конски сили на електрогенераторите;
- Резервната и неизползваната мощност на двигателите.

Тези сведения дават възможност да се разпредели, макар и приблизително, цялата двигателна сила, която е:

- послужила за движението на работните машини;
- послужила за производството на електрическа енергия;
- останала неизползвана и в резерв.

За да се избегне двойното броене на двигателните машини, работили в две предприятия, карта „З“ се попълва за това предприятие, в което двигателят е зает повече време. Другото предприятие е отнесено към групата на предприятията с неопределена двигателна сила. При разработката не се прави разлика между собствени и наети двигатели.

## **Номенклатура**

Главната дирекция на статистиката използва две номенклатури. Едната е разработена от френския статистик Жак Бертийон, а другата е изготвена от българските статистици и е използвана в две анкети за преброяване на насърчаваната от държавата индустрия, проведени в началото на ХХ век. Желанието да се постигне лесна сравнимост на данните с международната статистика, налага използването на номенклатурата на Бертийон в общи линии като схема за разработката на карта „З“. Отпечатана е „Подробна номенклатура на индустриалните (включително занаятчийските и от домашната индустрия) търговски и други заведения“. Стопанските отрасли са групирани в 6 отдела, 18 класа (табл. 1 и 2 от приложението) и 307 вида. За



нуждите на преброяването е отпечатан и систематически указател за ползване на подробната номенклатура, който съдържа подробна разбивка на отделите и класовете. Отпечатан е и Азбучник на подробната номенклатура, който улеснява работата на агент-преброителите.

### **Обработка на събраните сведения**

Събраните по време на преброяването сведения се обработват с голямо закъснение, едва през 1927 г., тъй като тогава са отделени карти „З“ от лична карта „А“ и от карта „Ж“. Работата по обработката на данните се извършва от променлив във времето брой чиновници, който никога не надвишава 9 души. В обработката участват средно 7 експерти от Главната дирекция на статистиката.

Подготвителните дейности обхващат следните процеси:

- Проверка на картите и класирането им според възприетата номенклатура;
- Изпращане на въпросници за корекция на нередовните карти;
- Вписване на получените отговори по изпратените въпроси;
- Маркиране с условни знаци на съдържанието на отделните въпроси по бордюра на картата.

За извършване на описаните предварителни процедури са изразходвани 16 000 работни часа. При проверката на картите се оказва, че в около 30% от тях са допуснати грешки и поради това са направени много запитвания с въпросници, за да се установят коректните данни.

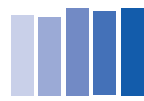
Табличната сводка се извършва на ръка, без машини, като разработката е извършена по отделни градове и околии.

### **Публикуване на резултатите от преброяването**

Предварителните резултати от проведеното преброяване на индустриалните и търговските заведения в Царство България на 31 декември 1926 г. са публикувани в списание „Месечни статистически известия“ на Главната дирекция на статистиката през 1928, 1929, 1930 и 1931 година. В Статистическия годишник от 1928 и 1929/1930 г. също са поместени резултати от преброяването.

Окончателните резултати са издадени в три отделни книги:

- Резултати от преброяване на индустриалните и търговските заведения в Царство България на 31 декември 1926 година. Книга I. Индустриални (включително



занаятчийските и от домашната индустрия, търговските и др. заведения по подробната номенклатура. Общо за Царството и за градовете. С., Държ. печатница, 1932, 260 стр.

- Резултати от преброяване на индустриалните и търговските заведения в Царство България на 31 декември 1926 година. Книга II. Индустриални (включително занаятчийските и от домашната индустрия), търговските и др. заведения по подробната номенклатура. Резултати по градове. С., Държ. печатница, 1932, 400 стр.

- Резултати от преброяване на индустриалните и търговските заведения в Царство България на 31 декември 1926 година. Книга III. Индустриални (включително занаятчийските и от домашната индустрия), търговските и др. заведения по подробната номенклатура. Резултати по околии (без градовете). С., Държ. печатница, 1932, 367 стр.

И трите тома са двуезични с паралелен текст на френски език. Публикациите с резултатите от проведеното през 1926 г. преброяване на стопанските предприятия са част от библиотечния фонд на НСИ и могат да се ползват в читалнята на библиотеката.

**ЦИТИРАНА ЛИТЕРАТУРА:**

**Резултати от преброяване на индустриалните и търговските заведения в Царство България на 31 декември 1926 година (1932).** Книга I. Индустриални (включително занаятчийските и от домашната индустрия, търговските и др. заведения по подробната номенклатура. Общо за Царството и за градовете. С., Държ. печатница, 260 стр.

**Резултати от преброяване на индустриалните и търговските заведения в Царство България на 31 декември 1926 година (1932).** Книга III. Индустриални (включително занаятчийските и от домашната индустрия), търговските и др. заведения по подробната номенклатура. Резултати по околии (без градовете). С., Държ. печатница, 367 стр.

## Приложение

Таблица 1

## Заведения според положението им в предприятието и заети лица

Стопански отрасли (отдели, класове и видове по номенклатура)	Всичко заведения		От тях					
	брой	заети	Самостоятелни		Централи		Клонове	
			заведения	заети лица	заведения	заети лица	заведения	заети лица
Отдел А. - Неземеделско градинарство и животновъдство, дърварство, въглищарство и лов.....	4346	8005	3994	7196	69	285	283	524
Отдел Б. - Добиване минерални вещества (вкл. солници)	499	7268	428	2460	8	171	63	4637
Отдел В. - Индустрия	115663	235939	109068	200835	2773	17778	3822	17326
Клас I. - Текстилна индустрия	9768	23184	8795	18879	207	2636	766	1669
Клас II. - Кожарска индустрия	3340	6369	3208	5981	67	256	65	132
Клас III. - Дърводелска индустрия	9182	15003	8741	13102	141	531	300	1370
Клас IV. - Метална индустрия	16873	33732	16458	30150	218	2026	197	1556
Клас V. - Керамика	3205	5033	3035	4696	65	256	105	81
Клас VI. - Химическа индустрия	763	4332	579	2952	64	725	120	655
Клас VII. - Хартийна индустрия и полиграфия	853	4686	788	4321	22	115	43	250
Клас VIII. - Прецизна механика, музикални инструменти, игралки, скъпоценни изделия и др.	856	1542	822	1390	15	79	19	73
Клас IX. - За храни и пиетета	29121	41060	26271	34857	1329	4487	1521	1716
Клас X. - Тютюнева индустрия	188	16766	100	7187	22	3909	65	5670
Клас XI. - За облекло, обувца и тоалет	26319	48729	25584	46561	351	1132	384	1036
Клас XII. - Мебелна индустрия	2917	9868	2829	9019	53	566	35	283
Клас XIII. - Строителна индустрия	4991	10244	4787	7932	103	306	101	2006
Клас XIV. - Добиване и доставяне: вода, електричество и лед	144	1057	92	450	14	165	38	442
Клас XV. - За преносни уреди	7143	14334	6979	13358	102	589	62	387
Отдел Г. - Съобщения и пренасяне	13141	12519	12843	11564	107	689	191	266
Отдел Д. - Търговия	63282	99598	58533	85446	2426	7642	2323	6510
Клас XVI. - Чисто търговски заведения	34431	51837	31591	45419	1409	3643	1431	2775
Клас XVII. - Други търговски заведения	6109	16476	5318	10357	239	2814	552	3305
Клас XVIII. - Хотели, кафенета, ресторанти и кръчми	22742	31285	21624	29670	778	1185	340	430
Отдел Е. - Развлечение и обучение	396	1937	359	1659	16	152	21	126
Отдел Ж. - Здравеопазване	457	2756	442	2552	8	194	7	10
Всичко за Царството	197784	368022	185667	311712	5407	26011	6710	29399

Таблица 2

**Брой на заведенията и заети лица в  
окръг София по околии без градовете**

Околии на Софийски окръг	Брой на заведенията	Заети мъже	Заети жени	Заети общо
1. Околия Брезник	503	290	16	306
2. Околия Ихтиман	1192	1408	250	1658
3. Околия Новоселци	935	1040	65	1105
4. Околия Орхание	1601	1090	160	1250
5. Околия Пирдоп	831	645	192	837
6. Околия Самоков	1023	916	97	1013
7. Околия София (градска)	13679	42868	8418	51286
8. Околия София (селска)	4455	9293	671	9964
9. Околия Трън	1016	584	93	677
10. Околия Цариброд	901	728	85	813

## ПЪРВО ПРЕБРОЯВАНЕ НА ИНДУСТРИАЛНИТЕ И ТЪРГОВСКИТЕ ЗАВЕДЕНИЯ В ЦАРСТВО БЪЛГАРИЯ ПРЕЗ 1926 ГОДИНА

*Соња Златанова\**

**РЕЗЮМЕ** В доклад № 4001 от 17 март 1926 г. до Върховния статистически съвет са изложени мотивите за допълване на демографското преброяване с преброяване на стопанските предприятия.

Върховният статистически съвет одобрява проекта на Дирекцията на статистиката за обединяване на демографското, жилищното и стопанското преброяване, като приема единодушно следното решение: „Едновременно с общото демографско преброяване по досегашния кадър да се предприеме и преброяване на селското стопанство и занаятчийско-индустриалното производство със специални карти, чрез които да се съберат най-общи сведения за установяване количествено и качествено, от една страна земеделческите и от друга - всички останали неземеделчески стопанства и предприятия в нашата страна“.

На 31 декември 1926 г. се провежда първото общо преброяване на индустриалните и търговските предприятия в Царство България едновременно с преброяването на населението и жилищния фонд. Главната дирекция на статистиката използва две номенклатури. Едната е разработена от френския статистик Жак Бертийон, а другата е изготвена от българските статистици и е използвана в две анкети за преброяване на насърчаваната от държавата индустрия, проведени в началото на XX век.

Окончателните резултати са издадени в три тома, които са двуезични, с паралелен текст на френски език. Публикациите от проведеното през 1926 г. преброяване на стопанските предприятия са част от библиотечния фонд на НСИ и могат да се ползват в читалнята на библиотеката.

---

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

## ПЕРВАЯ ПЕРЕПИСЬ ИНДУСТРИАЛЬНЫХ И ТОРГОВЫХ ПРЕДПРИЯТИЙ В ЦАРСТВЕ БОЛГАРИЯ В 1926 ГОДУ

*София Златанова\**

**РЕЗЮМЕ** В докладе № 4001 от 17 марта 1926 г. Высшему статистическому совету были изложены причины, по которым наряду с переписью населения необходимо провести перепись сельского хозяйства и промышленных предприятий и организаций.

Высший статистический совет одобрил проект Статистического управления по объединению переписи населения, жилищного фонда и экономической переписи, единогласно приняв следующее решение: „Одновременно с проведением всеобщей демографической переписи, провести перепись сельского хозяйства, а также ремесленно-промышленного производства с использованием специальных карточек, посредством которых собрать общие сведения для количественной и качественной оценки, с одной стороны земледельческих, а с другой - всех остальных неземледельческих хозяйств и предприятий в нашей стране“.

31 декабря 1926 года стартовала первая всеобщая перепись индустриальных и торговых предприятий в царстве Болгария, которая проводилась одновременно с переписью населения и жилищного фонда. Главное управление статистики использовала две номенклатуры. Одна из них была разработана французским статистиком Жаком Бертильоном, а другая была подготовлена болгарскими статистиками и уже использовалась в двух анкетах исследований для переписи индустрии, поощряемой государством, проведенных в начале двадцатого века.

Окончательные результаты переписи были опубликованы в трех двуязычных томах с параллельным текстом на французском языке. Публикации о всеобщей переписи 1926 года входят в библиотечный фонд НСИ и с ними можно ознакомиться в читальном зале библиотеки.

---

\* Младший эксперт отдела „Публикации, библиотека и цифровые продукты“ в НСИ; e-mail: SZlatanova@nsi.bg

## FIRST CENSUS OF THE INDUSTRIAL AND COMMERCIAL ESTABLISHMENTS IN THE KINGDOM OF BULGARIA IN 1926

*Sonya Zlatanova\**

**SUMMARY** In Report 17 4001 of 17 March 1926 to the Supreme Statistical Council are set out the reasons for supplementing the demographic census with a census of business enterprises.

The Supreme Statistical Council approved the project of the Directorate of Statistics for unification of the demographic, housing and economic census, unanimously adopting the following decision: ‘Simultaneously with the general demographic census on the current staff to undertake a census of agriculture and handicraft production with special maps, through which to collect the most general information for establishing quantitative and qualitative, on the one hand the agricultural and on the other - all other non-agricultural holdings and enterprises in our country’.

On December 31, 1926, the first general census of industrial and commercial enterprises in the Kingdom of Bulgaria was held simultaneously with the census of the population and the housing stock. The Directorate-General of Statistics uses two nomenclatures. One was developed by the French statistician Jacques Bertillon, and the other was prepared by Bulgarian statisticians and used in two surveys to count the state promoted industry, conducted in the early twentieth century.

The final results are published in three volumes, which are bilingual, with a parallel text in French. The publications from the 1926 census of business enterprises are part of the NSI library fund and can be used in the library reading room.

---

\* Junior expert in ‘Publications, Library And Digital Products’ Department, NSI; e-mail: SZlatanova@nsi.bg.





## В памет на проф. Поля Ангелова



Почина проф. Поля Ангелова<sup>1</sup>. Тъжна вест, която ни донесе много печал. В нейно лице загубихме голям приятел, професионалист и преподавател. Статистиците никога няма да забравят нейната усмивка и мъдри думи. Статистиката като наука и практика загуби учен с голям потенциал, опит и визия.

Загубихме приятел, съмишленик, колега и приятен събеседник. Ще останат завинаги в паметта ми срещите с проф. Ангелова и колегите от катедра „Статистика и приложна математика“ в Стопанска академия „Д. А. Ценов“ - Свищов. Винаги при тези срещи проф. Ангелова разговаряше професионално, точно и образно. Съумяваше да предразположи събеседниците си, да бъде общителна и внимателна в изказа. Тя умееше да се шегува, находчиво да иронизира и да дава критични оценки за събития, личности и случки от света, който ни заобикаляше. Беше и си остава една от нас: жизнена, обичаща живота, посветена на своята работа в науката, приемаща отговорността на лидера в битността си на заместник-ректор на Стопанска академия „Д. А. Ценов“.

Проф. Ангелова живя достойно, остави ярка следа след себе си и даде своя забележим принос към развитието на статистиката като теория и практика. Тя се нареди до тези, които със своя опит и преподавателски умения щедро раздаваха знание. Нейният пример ще остане като образец на човек отдаден на науката и познанието. Участието ѝ в Научно-експертния съвет на Националния статистически институт беше отчетливо професионално, прагматично и винаги резултатно. Със своето активно присъствие в редколегията на сп. „Статистика“ тя беше един от стожерите. Със своите научни публикации проф. Ангелова остави забележими,

---

<sup>1</sup> Дългогодишен ръководител на катедра „Статистика и приложна математика“ в Стопанска академия „Д. А. Ценов“, заместник-ректор на академията.

дълбоки бразди в статистиката като наука. Със своите статии в сп. „Статистика“ тя ще остане като забележителен изследовател, анализатор и учен, който отвори широко прозорците на познанието за идващите след нас. Нейните писани слова в книги и статии ще останат в читалните на библиотеките, ще се четат и ще бъдат като ярката светлина на фар в бурното море на науката.

Почивай в мир, Поля!

Д-р Богдан Богданов - зам.-председател на НСИ  
Редколегия на сп. „Статистика“

## НА ВНИМАНИЕТО НА АВТОРИТЕ

Статиите, предложени на редакцията на сп. „Статистика”, трябва да бъдат оригинални, непубликувани в други издания.

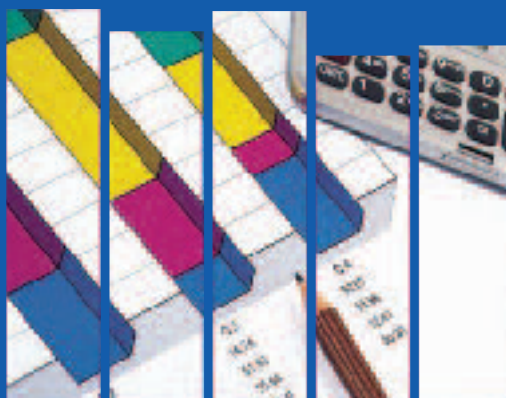
По решение на Редколегията на авторите се предоставя възможност статиите им да бъдат публикувани или на български, или на английски език. За целта е необходимо авторът, който желае статията му да бъде публикувана на английски език, да представи в редакцията ръкописи и съответни файлове и на български, и на английски език при спазване на посочените изисквания и за двата ръкописа. Авторите трябва да имат предвид, че английският текст ще бъде публикуван без редакционна намеса, което означава, че те носят пълна отговорност за коректността на превода.

Статиите, придружени от кратко резюме на български и на английски език и номерата на научната област от JEL класификацията, се представят в редакцията на сп. „Статистика”. Текстът трябва да бъде на Word, а графиките - Excel, със задължително приложени данни. Формулите в текста и тези на самостоятелен ред (без числата в тях) трябва да бъдат задължително курсивни (Italic). Ако графиките са дадени отделно, в текста трябва да се посочи точното им място. Авторите носят отговорност за коректността на данните и цитатите. Ръкописите трябва да са придружени от трите имена, научната степен и звание, точния адрес, телефона, длъжността, мястото на работа и електронния адрес (e-mail) на автора.

Поместването на статия на страниците на списанието не означава, че предложенията и евентуалните критични бележки, които авторът отправя, се споделят от Редакционната колегия или от ръководството на Националния статистически институт.

Редакцията на сп. „Статистика” би приветствала различни инициативи на читателската аудитория под формата на дискусии на страниците на списанието, научно-професионални „реплики” по повод на отпечатана статия, самостоятелни кореферати по повдигнатата тема - изобщо всяка инициативност по теми и проблеми на науката и практиката на статистическите изследвания от сферата на българската и европейската статистика.

**Редакция на списание „Статистика”**



**СТАТИСТИКА**  
**STATISTICS**

1/2021

[www.nsi.bg](http://www.nsi.bg)