

ГОЛЕМИТЕ ДАННИ, ПРОМЕНЕТЕ В СТАТИСТИКАТА И НОВИТЕ ПРЕДИЗВИКАТЕЛСТВА ПРЕД ПОЛИТИКОНОМИЧЕСКИТЕ СИСТЕМИ

Предложен е политикономически подход в изследването на промените, които настъпват в статистиката в резултат от трансформациите в три групи фактори – технологии, политика и икономика, а оттук и в методите за събиране, складиране и анализ на данни. Разгледани са основните етапи в развитието на статистиката в зависимост от измененията в технологиите за обработка на данни и от променящите се нужди на политикономическите системи. Анализирани са „големите данни“ и предизвикателствата, които поставят пред традиционната статистика. Те, от една страна, позволяват да се разкриват невидими преди това тенденции и да се оптимизират все повече човешки дейности, но от друга, повдигат много нови въпроси, които чакат своя отговор. Тези въпроси произтичат от факта, че променяйки статистиката, големите данни съдържат тенденцията да преобразуват политиката, икономиката, обществата и всяка дейност, изискваща вземането на решения.¹

JEL: A12; C10; C8; O25; P16

Ключови думи: индустриална революция; дигитална революция; големи данни; статистика; алгоритми

Индустриални революции и статистика

Управлението на държавата и икономиката е невъзможно без нужното количество информация за вземането на рационални решения. Тази информация се събира от държавните институции по множество различни канали, а етапите на развитие на държавата, свързани с ускоряването и усложняването на социалната система, са неотделими от етапите на създаване на система за събиране и обработване на информация за управлението. Официалната статистика е именно такъв основен инструмент в управлението на държавата, чиито етапи са свързани и с етапите на развитие на икономическата наука, а също и на пазарната икономика и различни други форми на събиране на емпирична и конкретно на статистическа информация от различни субекти.

* УНСС, катедра „Политическа икономия“, prodanovhristo@gmail.com

¹ Chief Assist. Prof. Hristo Prodanov, PhD. BIG DATA, CHANGES IN STATISTICS AND THE NEW CHALLENGES TO POLITICO-ECONOMIC SYSTEMS. *Summary:* A politico-economic approach is presented for the exploration of the changes that occur in statistics as a result of the transformations in three sets of factors – technologies, politics and economics, and hence in the methods of collecting, storing and analysing data. The main stages in the development of statistics are examined, taking into account the changes in the data processing technologies and the ever changing needs of politico-economic systems. Big Data and the challenges it poses to traditional statistics are analysed. On the one hand, it is able to reveal previously unseen trends and to optimize more human activities, but on the other hand it gives rise to many new questions that are waiting to be answered. These questions stem from the fact that by changing statistics big data tends to change politics, economics, societies and any activity that demands for decisions to be made. *Keywords:* industrial revolution; digital revolution; Big Data; statistics; algorithms.

Основната теза в изследването е, че една от характеристиките на всяка индустриална революция е качественият скок в увеличаването на информацията, както и във възможностите за събирането, складирането и използването ѝ, което води до промени не само в политиката и икономиката, но и във взаимодействието помежду им като политикономическа система, съответно и до промени в статистиката. От своя страна статистиката има системни следствия, тъй като оказва съществено влияние в три основни социални подсистеми – политиката и в частност държавата, икономиката и поведението на икономическите субекти, икономическата и социалните науки изобщо, чието развитие е силно обвързано със събирането и обработването на емпирична информация. С всяка индустриална революция се увеличава количеството генерирана информация, усъвършенстват се комуникационните технологии, извършват се последователно преходи от събиране и натрупване на информация на хартиен носител, различни видове дискове, за да се стигне в наши дни до натрупването ѝ в гигантски размери в различни бази данни, облачни услуги, до събирането и използването ѝ от високотехнологичните компании и държавите.

Самото възникване на държавата в най-ранните ѝ антични версии изисква съответна информация, за да може тя да функционира. Затова статистически данни могат да се открият още в първите държавни формирания и големи империи, които събират сведения за населението, търговията с определени стоки, богатството. Смята се, че статистиката става обект на внимание още през IX век, когато в своя „Ръкопис за дешифриране на криптографски съобщения“ арабският учен Ал-Кинди (801-873) описва как статистиката и честотният анализ могат да бъдат използвани при дешифрирането на криптирана информация. Интерес представлява и историята на Флоренция на Джовани Вилани (1276-1348), който дава обширна информация за населението, наредбите, търговията, образованието и религиозните обекти (макар че терминът „статистика“ все още не съществува по онова време).

Въпреки тези и някои други примитивни опити за статистическа обработка на данни възходът в развитието на статистиката обикновено се свързва с разгръщането на това, което е известно като модерна държава, притежаваща легитимния монопол за употребата на принуда върху определена територия. Исторически за неин предвестник се смята Вестфалският мирен договор, с който завършва Тридесетгодишната религиозна война в Европа. В договора е записано, че всеки владетел притежава пълен монопол върху суверенитета на определена територия. За да може да реализира този суверенитет и да упражнява икономическа власт, да провежда икономическа политика и да взема рационални решения обаче, са необходими съответни данни за процесите в рамките на държавата.

Така се поставя началото на първия етап в развитието на статистиката в политикономическите системи на меркантилисткия тип доиндустриален капитализъм. Разгръщането на абсолютната монархия, която се нуждае от постоянен поток от данни за населението, ресурсите и икономиката, е важен

фактор, стимулиращ този процес. За дълъг период обаче става въпрос за събиране основно на демографски и икономически данни, а основният субект, който прави това, е държавата. Само тя разполага с необходимите ресурси да извършва подобна дейност и главно тя има нужда от такива статистически данни за човешките и материалните ресурси, които може да използва по един или друг начин. Тези данни се събират бавно и не са достатъчни за установяването на опасни тенденции и за предвиждането и предотвратяването на конфликти (например на Английската и Френската буржоазни революции).

Между 1642 и 1688 г. по време на Английската буржоазна революция през 1663 г. английският демограф Джон Граунт публикува книгата „Природни и политически наблюдения върху регистрите на смъртността“, където, анализирайки смъртността, прави първото статистическо преброяване на населението на Лондон. Първоначално статистиката се базира върху ограничено количество данни, които се използват за нуждите на държавното управление и за ръководство на икономиката. Оттук идва и етимологията на термина, употребен за първи път през 1749 г. в Германия от Готфрид Ахенвул (1719-1772). През 1791 г. Джон Синклер (1754-1835) въвежда понятието „статистика“ на английски език в статистическите отчети, които прави в Шотландия.

Вторият етап в натрупването и събирането на информация, утвърждаващ разгръщането на политикономическата система на съвременната модерна индустриална държава, започва с Първата индустриална революция. В нейното начало обаче възможностите за събиране и натрупване на информация все още са ограничени. Затова в разгръщащата се политическа икономия за дълъг период теорията доминира емпириката, а вземането на управленски решения се основава върху ограничено количество данни. Дори Адам Смит през 1776 г. поради липса на достатъчно данни е принуден да работи с абстракции като „невидимата ръка на пазара“ и „икономическия човек“, с които обяснява функционирането на икономиката. Това се дължи на факта, че все още методите за събиране, организиране и обработка на емпирични данни са сравнително ограничени, а основен източник за събиране на такива данни е държавата с нейните статистически служби, чиито цели са свързани най-вече с нуждите на държавното управление. Статистиката има относително периферно значение в научните изследвания на процесите в политиката и икономиката чак до XIX век. Тя става важен елемент в изследванията на автори като Уилям Пети (1623-1687) и Томас Малтус (1766-1834), но въпреки това все още оказва слабо влияние върху научните изследвания поради ограниченията в методите за събиране и обобщаване на данни при съществуващите тогава технологии, както и поради необходимостта от огромни ресурси за целта, които са на разположение на малък брой субекти, сред които на първо място е държавата.

С началото на Първата индустриална революция и с разгръщането на индустриалния капитализъм, свързан с изобретяването на парната машина от Джеймс Уат през 1776 г. и на телеграфа от Самюел Морз през 1836 г., с усложняването на процесите в обществата, с нарастването на броя на нуждаещите се

от статистическа информация субекти, с постепенната диференциация на научното познание и развитието на различни дисциплини с все по-строг и стеснен предмет на изследване, статистиката започва да се обособява като самостоятелна наука със собствени инструментариум и методология, което може да се разглежда като втори етап от развитието на съвременната статистическа наука. Той стартира с работата на английските изследователи Франсис Галтън (1822-1911) и Карл Пиърсън (1857-1936), които превръщат статистиката в строга математическа наука и започват да я използват за научни, индустриални и политически изследвания и анализи. Те развиват и въвеждат понятия и методи като „стандартно отклонение“, „регресионен анализ“, „корелация“, т. нар. метод на моментите, създават първото статистическо списание „Биометрика“, а след това и дисциплината „Математическа статистика“ в Университети Колидж, Лондон, привличаща студенти от целия свят.

В началото на Първата индустриална революция с утвърждаването на водещите национални държави в развитите страни се формират и техните институции. През 1800 г. е създадено Бюрото по статистика в Париж, а през 1834 г. във Великобритания – първата официална статистическа институция „Кралско статистическо общество“. Изграждат се националните банки и борсовите индекси, които непрекъснато генерират данни и информация. Държавата поетапно усъвършенства своята роля в намесата в една или друга сфера.

Третата вълна в развитието на статистиката като основа на следващ етап в разгръщането на една много по-силно регулираща икономическите процеси държава започва с Втората индустриална революция. По време на Дългата депресия между 70-те и 90-те години на XIX век, когато започва прехода към Втора индустриална революция, се създават национални статистически институции в голяма част от съществуващите и нововъзникващите държави. В България веднага след Освобождението с формирането на държавните институции в началото на 1880 г. се изграждат статистически звена към Министерството на правосъдието и Министерството на финансите.

Втората индустриална революция води не само до продуцирането на все повече информация, на хора, занимаващи се с нейната обработка, но и до подобряването на технологичните възможности за събиране, обобщаване, обработка на данни и информация. Съществена роля за това изиграва развитието на информационно-комуникационни (радио, телевизия, телефон, грамофон, табулираща машина) и транспортни технологии (автомобили и самолети), което предизвиква увеличаване на източниците на данни и усложняване на методите за тяхното събиране, организиране и анализ. Допълнителен стимул за развитието на статистиката дават фордисткият начин на производство, кейнсианската (на Запад) и социалистическата (на Изток) регулация на икономиката, засилващата функциите си централизирана национална държава, контролираща увеличаващо се количество процеси на своята територия и нуждаеща се за целта от все повече данни. Появяват се и се използват активно нови индекси и показатели за състоянието и развитието на икономиката. Най-значимият

от тях, създаден в резултат от голямата икономическа криза през 1929-1933 г., е брутният вътрешен продукт. Умножават се субектите, които събират данни. Наред с държавните статистически служби това започват да правят и все повече частни субекти, каквато е например основаната от Джон Мудис през 1909 г. рейтингова агенция „Мудис“. Възникват концепциите за достатъчност, спомагателни статистики, линейният дискриминатор на Фишер, започват да се тестват хипотези, изграждат се техники за обработка на малки количества данни, върху чиято основа се правят съответните обобщения, което дава тласък на развитието на емпиричната социология.

Увеличава се броят на наблюдаваните икономически показатели – от индекса на цените до безработицата. Оттогава до наши дни постепенно се разширява предметът на статистиката и нейните резултати започват да служат за проверка на хипотези във все повече науки и за вземане на все повече управленски решения. Днес трудно може да се срещне научна и управленска област, която не използва достиженията на статистиката. Започват да се мерят например такива процеси като инфлацията и дефлацията. Развива се идеята за националния доход и се предприемат измервания на различията в богатството и доходите. Поставено е началото на модерното преброяване на населението, при което вече се изисква информация не за домакинството, а за отделния човек. Именно извършването на качествени промени в събирането количество информация и показателите за функционирането на икономиката и обществата прави възможно много по-рационалното активно регулиране на процесите в държавата, водещо до индикативно и директивно планиране, до социалната държава на Запад и версията на догонваща планова държава в Източна Европа.

Четвъртият етап в развитието на статистиката започва през 70-те години на XX век с разгръщането на информационно-комуникационните технологии на Третата индустриална революция, компютрите и интернет. Те се превръщат във важен инструмент за събиране, складиране и обработка на всякакви данни. Нараства значението на емпиричната основа на познанието за сметка на теорията. Неща, които преди не е било възможно да бъдат разкрити емпирично, вече стават видими. Огромната изчислителна мощ на компютърните технологии позволява в много по-голяма степен отпреди да се събират и анализират данни за всякакви процеси и тенденции, да се правят съответните корелации и прогнози, свързани с усложняването на обществените системи. Използването на компютри съкращава времето за сложните статистически изчисления, улеснява проверката на хипотези, повишава достоверността на резултатите и води до създаването на нови методи, невъзможни дотогава поради ограниченията в ръчното анализиране на данни. Същевременно се наблюдава и обратната тенденция, свързана с ускоряването на всички процеси, с експоненциалния характер на развитието на технологиите, с усложняването на социалните и икономическите системи. Всичко това изисква растящо количество непрекъснато подавани данни, за да могат да се взимат рационални

решения от различните държавни институции и икономически субекти. Ето защо в САЩ например има 13 основни правителствени статистически агенции, за които през 2017 г. са похарчени 2.257 млрд. USD. Предлаганата от тях информация е важна за компаниите, които се ориентират към събирането на голямо количество данни, но те не са достатъчни за достигането до оптимални бизнес решения. Данните на компаниите се съчетават с множество данни, събирани от правителството, за да се вземат по-ефективни решения. При това се отчита, че в епохата на скоростно развитие на информационните технологии публичното събиране и разпространение на данни изисква непрекъснато развитие и подобрене, за да бъде в съответствие с бързите структурни промени в икономиката (вж. Hughes-Cromwick and Coronado, 2019, p. 132).

Същевременно усложняването на всички процеси и прехвърлянето на функции от държавата към пазара води до растеж на количеството данни, генерирани от частни субекти, мозъчни тръстове и международни организации – от индекса на човешко развитие, през индекса на корупция, свободата на медиите, благосъстоянието, икономическата свобода, тероризма, робството, до индекса на иновативността, конкурентността, глобализацията и т.н. Протича процес на приватизация на все по-увеличаващо се количество данни, рейтинги, показатели, генерирани от мозъчни тръстове, консултантски компании, неправителствени организации и др. Появяват се стотици различни национални и световни рейтинги и показатели, чрез които се оценяват състоянието, тенденциите и развитието на всички сфери на обществото. Формирането на нишови пазари е важен фактор в това отношение, тъй като те изискват много по-точно таргетиране на съответните ниши потребители, а това във все по-голяма степен зависи от статистическите данни, събирани от компаниите. Наред с пазара и ценовите сигнали, статистиката се превръща в особено важен фактор в определянето накъде да бъдат насочени производството, размяната, разпределението и потреблението. Но тя става важна и за други, на пръв поглед, странични дейности – спорт, забавления, туризъм, психология, медицина, водейки до оптимизирането на все повече сфери на обществения, икономическия и политическия живот.

Трябва да се отбележи още, че етапите в развитието на статистиката и събирането на данни, необходими за управлението и държавата, са свързани с циклични кризисни процеси. От тях се излиза с нова технологична революция и с усъвършенстване на начините за събиране, преработка и използване на данни за икономиката и обществените процеси. Те служат за основа на нови икономически теории, за да могат държавата и икономическите субекти да реагират по-бързо и успешно, да вземат рационални решения, с които да се избягват предходните кризисни процеси. Статистиката и всички останали форми на събиране на емпирична информация за социалните процеси стават все по-значими с развитието на националните държави и с действията им по контрола и управлението на процесите на тяхна територия. Постепенно държавите не само непрекъснато увеличават събираните данни, но и въвеждат нови показа-

тели за измерване на една или друга тенденция. Голямата икономическа криза през 1929-1933 г. например става фактор за появата на БВП като основен показател за състоянието и развитието на икономиката и за репрезентативните емпирични социологически изследвания, с които се мери отношението на населението към политиката и останалите социални явления. Кризата през 2008 г. пък дава тласък за развитието на Четвъртата индустриална революция и на „големите данни“ като нов етап за емпирична оценка на различни процеси и вземането на решения от държавни институции и от редица други субекти.

„Големите данни“ – фактор за нов етап в развитието на статистиката като информационна основа на функционирането на държавата и икономиката

В началото на XXI век с разгръщането на феномена „големи данни“ настъпва нов етап от развитието на технологиите за събиране, складиране и анализ на данни. Той е свързан с навлизането на света в това, което започва да се осмисля с понятия като „дигитална революция“, „Четвърта индустриална революция“, „дигитална икономика“, „Индустрия 4.0“ и т.н. Определящо за този етап е, че стъпва върху дигитализацията на растящо количество процеси, а дигитализацията от своя страна се основава върху данните. За всички дигитални технологии – от компютрите и смартфоните, през „интернет на нещата“, социалните медии и търсачките до автономните превозни средства и роботите, е характерно, че чрез тях се събират и натрупват данни. Всеки човек, докоснал се по един или друг начин до заобикалящите го от всички посоки дигитални устройства, се превръща в генератор, в „сноп от данни“, които след това могат да се използват за определени цели. Те идват от сензорите по улиците, от „интернет на нещата“, търсачките, социалните мрежи, интернет сайтове, телефонни обаждания, електронната търговия, облачните технологии, мобилните приложения и се смята, че позволяват много по-ефективно регулиране на процесите. Така се раждат т.нар. големи данни (Big Data), отличаващи се с внушителен, непознат в историята обем, огромна скорост на производството им, разнообразие, нестабилност, неорганизираност, трудност за разбиране и осмисляне без специална обработка. Те невинаги са достатъчно достоверни, което се превръща в проблем при тяхното събиране, структуриране, анализиране, а оттук и за употребата им от традиционната статистика, която трябва да инкорпорира нови методи и модели, за да може да използва познавателния им потенциал.

В зависимост от начина на генерирането им тези данни могат да се обособят в три типа: създадени в комуникацията между индивидите, свързани с комуникациите между човек и машина и генерирани от технически устройства. „Големите данни“ все повече се превръщат във фактор за промени в маркетинга, който адаптира своите послания към интересите и психологическия профил на отделни хора и групи потребители. Тяхното значение става важно за банковата система, здравеопазването, сферата на производството и цялостното

функциониране на обществото. Анализът на данните се превръща в конкурентно предимство не само в бизнеса, но и в политиката и останалите социални сфери и дейности. Поради съществуващите геотехнологически сблъсъци, колизии и войни все повече държави ограждат дигиталното си пространство от външни вмешателства, за да опазят данните за своите граждани и институции в условия на растяща дигитализация на държавното управление. Разгръщането на тези процеси създава поредица от нови противоречия и предизвикателства пред традиционната статистика, а оттук и по отношение на политическата икономия на „големите данни“.

1. В резултат от подобряването на технологичните възможности за събиране, обобщаване и анализ на данни наред с растящата нужда от данни за икономиката, политиката и всички управленски сфери *„големите данни“ бързо започват да се възприемат като важен източник за официалната статистика, тъй като позволяват да бъде получено голямо количество информация с относително малко усилия и разходи.* Това, което ги отличава от предходните форми на събиране на статистическа и изобщо на емпирична информация, е: *първо*, ако голяма част от предходни форми са свързани с репрезентативни извадки, тук е налице цялото възможно количество данни в определения сегмент; *второ*, тези данни се продуцират и обработват в реално време с помощта на съответни алгоритми, което дава възможност веднага да се реализира и да се вземе съответно управленско решение; *трето*, данните не се събират и анализират от хора, а в нарастваща степен от самите дигитални технологии. Чрез тях могат да се създават и използват нови показатели и индикатори за процесите в икономиката и в обществото. Тези данни се обработват с помощта на машинното обучение и съответни алгоритми, а не чрез традиционните статистически методи. Тъй като изводите се опират на много по-голямо количество информация, „големите данни“ обикновено са и по-точни.

Поради много по-широкия си обхват те често служат и за значително по-точни прогнози на определени тенденции в ситуация на нарастващи кризисност и неопределеност в обществата, когато поради експоненциалното ускоряване на всички процеси прогнозирането става все по-трудно, а тенденцията е все повече прогнози да не се реализират. „Големите данни“ се превръщат в особено важен източник за преодоляване на негативите от тази кризисност, тъй като позволяват процесите да се управляват в реално време, като паралелно чрез тях могат да се прогнозира икономически, социални, политически, климатични, здравни предизвикателства и рискове; да бъде подпомагано наблюдението на финансовите потоци; да се предупреждава за предстоящи заплахи; да се прогнозира икономическият растеж; да се оптимизира данъчното и митническото наблюдение; да се оценяват различни типове неравенства; да се идентифицират поведенчески модели и модели на потребление, равнище на доходите, инфлация, туризъм, търговски потоци.

Получаването на навременна информация е особено важен фактор за вземането на адекватни решения във всяка управленска сфера и поради това

анализът на „големи данни“ създава възможности за предотвратяване на кризи, както и за оценка на финансово-икономически и политически рискове. Те навлизат масово в икономическите и финансовите изследвания и стават все по-значим елемент във вземането на решения както в частния, така и в публичния сектор. В политически план се очертава тенденция анализът на „големи данни“ да се използва за идентифициране и предотвратяване на политически и управленски рискове, във връзка с което много правителства се опитват да изградят нови подходи в организацията на своята дейност както на национално, така и на международно равнище.

2. Дигитализацията на системите за управление и създаването на електронни правителства дава възможност растящо количество данни да се събират от публичната администрация, да се използват и от статистиката, а публичните институции да могат да повишат качеството на своята дейност. „Големите данни“ се разглеждат като фактор за ограничаване на бюрокрацията, която поради усложняването и ускоряването на всички процеси в икономиката и обществото през последните десетилетия непрекъснато нараства. За да може държавата да реагира бързо и успешно на постоянно възникващите проблемни ситуации, започват да се изградят електронни правителства и управленски системи, които се превръщат във важен инструмент за ограничаване на бюрокрацията и в източник за генериране на данни за нуждите на публичните институции.

Рационализацията на държавата и на правителствата не може да стане без качествен скок в бързината и количеството на събираната и преработвана информация за всички процеси в обществото. Наличието на голямо количество данни за поведението на индивидите, продуцирани чрез дигитализацията на икономиките, управлението, цялостния живот на хората, допринася за подобряване на количествения анализ в социалните науки, а на институциите и на вземащите решения се предлагат обработени, структурирани и свързани системи от данни. Това става възможно именно чрез „големите данни“, чрез които се следят например нивата на замърсяване и шума по улиците, предвиждат се рискове от кризи и социални конфликти, решават се екологични, политически, икономически и социални проблеми, вземат се решения в областта на образованието, здравеопазването, инфраструктурата, правосъдието и др., като процесът е алгоритмично автоматизиран. Друг типичен пример е данъчната система и обвързването на касовите апарати с държавни институции, позволяващи във всеки момент да се контролират количеството и стойността на продаваните стоки, получаваните доходи от тях, промените в цените и т.н. Благодарение на „големите данни“ институциите получават допълнителни възможности да проследяват тенденции, миграция, плътност на населението, развитие на финансови или друг тип пазари, да вземат решения и да регулират определени процеси.

Статистическите данни, събирани чрез въпросници, все повече се допълват и заместват от данни, получени от съответни административни структури.

Например някои страни вече не извършват обширни анкетни проучвания сред населението, а правят преброявания чрез съчетаване и анализиране на данни от различни административни източници. По този начин националните статистически институти се обвързват още по-тясно с информационните системи на правителствата, намалявайки или изоставяйки традиционните въпросници за събиране на данни. Това прави резултатите от съответните проучвания не само по-точни, но и в редица отношения уникални, защото само националните статистически институти имат възможност за събиране и обработване на всички нужни данни от националните информационни системи на правителствата (вж. Struijs, Braaksma and JH Daas, 2014, p. 2-3).

3. *Събирането и анализът на данни намира приложение както при използването им в реално време, така и за прогнозиране на определени процеси.* „Големите данни“ могат да се анализират в реално време, както е например с отчитането на пътния трафик, „интернет на нещата“, автомобилите без шофьори, роботите, нефтопроводите, при които е нужна незабавна реакция, за да се отстрани някакъв проблем или да се изследва разгръщането на определени процеси в реално време. Данните за локализацията на мобилните телефони, с каквито разполагат повечето хора днес, могат да бъдат използвани за почти всекидневно проследяване на движението на населението, както и в статистиката в областта на туризма. Данни за инфлацията могат да бъдат получени от ценовата информация, която се съдържа при продажбите в мрежата.

В Холандия например всяка минута пътни сензори събират 24 хиляди единици данни от пътна мрежа от над 6 хиляди километра и ги публикуват на всеки 75 секунди, което дава възможност на шофьорите да избягват задръстванията и повишава сигурността по пътищата. От своя страна, анализирайки тези данни, официалната статистика може да предоставя сведения за пътния трафик – брой на колите, влизащи и излизащи от страната, видовете повредени автомобили, количеството коли от различни страни и градове, и т.н., което не може да бъде постигнато с традиционните статистически инструменти. В Австрия статистическата служба използва спътникови снимки на Земята, чрез които се оценява как се експлоатира нашата планета. В Естония се събират и обработват анонимно от компании, осигуряващи мобилните комуникации, данни за техните потребители и по този начин се получава информация за трансграничните пътувания (вж. Giczi and Szoke, 2018, p. 170).

Посланията в социалните мрежи пък могат да се анализират и обобщават като индикатор за потребителското или политическото поведение на клиентите, което служи за целите на политическата и икономическата реклама и маркетинг. Паралелно „големите данни“ се проучват и за да се регулират някои процеси в перспектива, когато става дума за събиране и анализ на информация с цел постигането на определени резултати в бъдещето или за доказване на дадени научни хипотези. Типичен пример е превърналата се в бестселър книга на Тома Пикети „Капиталът на XXI век“, в която благодарение

на компютърните изчисления се изследват тенденции със столетия назад. Известен е и т. нар. индекс на „Джини“, проследяващ чрез новите технологии за обработка и анализ на данни развитието на неравенствата от времето на Римската империя до днес.

4. *„Големите данни“ отговарят на нови въпроси и са фактор за създаването на нови индикатори.* Статистическите данни традиционно се използват за търсене на съответни каузални и функционални зависимости, изследващи развитието на дългосрочни тенденции и разгръщането на определени процеси на макроравнище. Същевременно с помощта на „големите данни“ се разкриват модели на взаимовръзка и корелации, чиято роля е не да обясняват защо нещо се е случило, а да сигнализират, че нещо се случва и може да е нужно да се реагира и така процесът да се насочи в желана посока. Те служат за регулиране на процесите най-вече на микроравнище – например за пътен трафик, идентифициране на някакви непосредствени заплахи, таргетиране на точно определени потребители. Така могат да бъдат разработени нови индикатори, свързани с получаването на корелации в реално време и със създаването на ранни предупредителни системи срещу специфични или системни рискове в различни сфери в дадена държава – външна и вътрешна сигурност, реална и дигитална икономика, финансова, здравна или друга система. Известно е например, че търсенията в „Гугъл“ и постингите във „Фейсбук“ вече се използват за предсказване на ликвидността на фондовите пазари и за конструирането на индекси на нагласите, от които могат да се направят прогнози. Смята се, че чрез анализ на данните от мобилните телефони, както и от различните приложения, телефонни обаждания, сателитни снимки може да се прогнозира социално-икономически тенденции, движение на населението, кредитни рискове и т.н.

Получената чрез „големите данни“ информация може да помогне да се видят взаимовръзките между различните политики и техните резултати, давайки възможност да се реагира на рискове от политически напрежения, нарастващи социални неравенства или недоволства на населението. В разработка на Международния валутен фонд се правят предложения за използването на „големите данни“ за проследяване, анализ и решения в няколко основни посоки: (1) използване на данни от системата за банков трансфер SWIFT за проследяване на глобалните финансови потоци; (2) създаване на система за ранно предупреждение, основана на проучване на промените в масовите нагласи на различни групи от хора, от чиито действия зависят едни или други процеси; (3) прогнозиране на тенденциите в БВП чрез използване на данни от „Гугъл“; (4) автоматизиране и разширяване на обхвата на прилагане на определени индекси; (5) събиране на данни за движението на паричните потоци, управлявани от държавата, с цел повишаване на контрола и качеството на провежданите политики; (6) използване на анализа на данните за подобряване на администрирането на данъците и митата (вж. Hammer, Cornelia, Kostroch, Quirós and STA Internal Group, 2017, p. 11-14).

5. Данните днес са повече от когато и да било в историята на човешката цивилизация, но понякога това прави *по-трудни тяхната организация, обобщаване и анализ, за да може данните, които имат някакво значение, да бъдат отделени от онези, които не са толкова съществени*. Причината е, че за разлика от традиционните статистически данни, които са специално събирани за определена цел, повечето от „големите данни“ са страничен продукт в бизнеса, управлението, социалните мрежи, „интернет на нещата“, търсачките на информация в интернет от рода на „Гугъл“, използването на мобилните телефони за най-различни цели. Голяма част от тях не са структурирани, а някои, засягащи дейността на бизнес системите, обикновено представляват структурирани информации за трансакции. Друга част са генерирани от устройства, свързани с „интернет на нещата“ – данни от сензори, уебкамери, установена чрез GPS устройствата на мобилните телефони локация и т.н. Поради това редица източници на „големи данни“, на първо място, тези в социалните медии, не са добре структурирани и към тях трудно се прилагат традиционните методи на репрезентативната извадка. Например „Фейсбук“ или „Гугъл“ могат да правят анализ само на данните, с които разполагат, но не и на онези групи от населението, които не използват съответните търсачки или социални мрежи. Сензорите по пътищата дават данни за трафика по дадена пътна отсечка, но не и за трафика изобщо, ако липсват сензори по всички пътища на страната. „Амазон“ може да създава профили и да таргетира онези потребители, които пазаруват от него, но не и останалите.

Възниква проблем и при интерпретацията на данните, например поради това, че даден мобилен телефон може да е регистриран на нечие име, но да се използва от някой друг, или пък човек да си създаде фалшив профил в социалните мрежи, което от своя страна прави невъзможен анализа на характеристиките на неговия притежател. Освен това източниците на „големи данни“ са различни дигитални устройства и данните от тях се подчиняват на различна логика. Обикновено те са селективни, тъй като се събират от определени търсачки, социални мрежи, сензори с конкретен профил потребители, а трябва да бъдат репрезентативни, за да могат да служат за нуждите на официалната статистика. Например при изследването на понятието за население в официалната статистика се предполага определена репрезентативност на данните. Ако в този анализ се използват примерно данните на потребителите на мобилни телефони, очевидно е, че те сами по себе си не могат да дадат репрезентативни данни за цялото население.

Трудността да бъдат прилагани традиционните статистически методи, базирани на репрезентативни извадки, се превръща в предизвикателство пред границите на използваемост, качеството и точността на данните. Това налага изработването на нови показатели и методи, чрез които да се търсят съответните корелации и да се управляват процесите не само в перспектива, но и в реално време. Така се ражда феноменът „тъмни данни“ – неорганизиранни, неструктурирани, некачествени и непригодни за изследователски цели данни,

които в много случаи достигат над 90% от общото количество и това изключително затруднява тяхното разделяне от онези, които имат смисъл и могат да послужат за получаването на някаква информация.

Ето защо преди „големите данни“ да бъдат използвани в статистиката, те трябва да бъдат селектирани и структурирани правилно в зависимост от целите на конкретното проучване. Изследването на данните предполага необходимостта от разкриване на тяхната структура, модели и качество. Трябва да бъдат идентифицирани грешките, аномалиите и липсващите данни. Нужни са и съответни методологии и модели на интерпретация на данните, отчитащи различията в характеристиките им и съответно прилагането на различни типове анализи (вж. Giczi and Szoke, 2018, p. 164-168).

6. *Възникват противоречия във връзка с данните, притежавани от частния и от публичния сектор.* Възходът на икономиката на „големите данни“ има съществени следствия за организациите, за които досега производството и анализът на данни и информация са водещи. Това е важно, особено за социалните науки, както и за държавните и наднационалните статистически институции като „Евростат“ и различните международни организации, продуциращи официалната статистика, която е в основата на дейността на политици, държавни и наднационални структури и частния бизнес. Ако досега тя се е ръководела от държавата и от различни специализирани в тази дейност частни структури, а немалка част от данните са получавани чрез интервюта от врата на врата, по телефона или чрез експертни оценки, в днешно време данните изтичат от всички посоки, събират се от всякакви дигитални устройства и се анализират чрез съответните алгоритми, което позволява на относително ниска цена да се оптимизират все повече процеси в един все по-нелинеен и непрогнозируем свят. Те са толкова сложни, че традиционните методи за събиране, съхраняване и изследване се оказват неприложими, налагайки необходимостта от нови умения, технологии и инфраструктура. Всички тези данни се събират и анализират от огромен брой субекти и започват да играят все по-голяма роля в официалната статистика. В същото време държавата губи изключителните си права върху тях за сметка на големите дигитални монополи. Оказва се, че частните компании често притежават много повече възможности за генериране и обработка на информация, отколкото държавните институции, които разчитат основно на административни източници и поради това нерядко се обръщат към частния сектор за получаването ѝ.

Същевременно данните в частния и в публичния сектор в много случаи влизат в несъответствие помежду си, давайки противоречиви интерпретации за едни и същи тенденции. Една от причините е, че „големите данни“ са доста по-различни от тези на традиционната статистика, тъй като значителна част от нея отразява предимно дългосрочни процеси. За разлика от тях „големите данни“ са относително нов феномен, изразяват това, което става в момента, и служат за нуждите най-вече на частните компании. Те са динамични, постоянно променящи се и еволюиращи. Системите, мрежите и методите за техния анализ

са в процес на постоянно развитие и усъвършенстване, а това създава не само нови възможности, но и предизвикателства и ограничения в статистическата им употреба. Освен това данните за официалната статистика се събират с определени цели и са предварително фиксирани, докато „големите данни“ са разпилени в най-различни дигитални устройства и системи – „интернет на нещата“, социални мрежи, търсачки, административни и корпоративни системи. Традиционните статистическите служби работят в повечето случаи с репрезентативни извадки, докато „големите данни“ не могат да се разглеждат като такива. Официалните статистически служби използват утвърдени методологии и стандарти, давайки съответните гаранции за качеството на информацията, която извличат от тях, докато „големите данни“ се характеризират с непълнота, прекъснатост, нестабилност, което често ги прави непригодни за получаване на достатъчно достоверна информация.

Оказва се, че в случая с „големите данни“ традиционните статистически процедури за обработка на данни често не работят. Така, въпреки че развитието на технологиите прави много по-лесен процеса на събиране, съхраняване и обработка на данни, традиционната статистика все още се опитва да изработи правилните подходи и методи за техния анализ. Затова трябва да се има предвид, че данните, които генерират националните статистически институти и тези, произведени от частни компании, много често имат различен характер, използват различна методология и преследват различни цели. Докато в първия случай те служат най-вече за предвиждане и регулиране на дългосрочни тенденции, засега „големите данни“ се ограничават до проследяване на краткосрочни развития и регулиране на процеси в реално време, например в т. нар. умни градове, където те позволяват оптимизиране на трафика, пътникопотока и избягването на инциденти. По този начин „големите данни“ водят до появата на нови статистически показатели, подпомагат бързия анализ, скоростното вземане на решения и прогнозирането на тенденции, превръщат се в нов източник на информация за официалната статистика. Те създават нови възможности пред статистиката, но и нови предизвикателства, които изискват нови подходи, законодателна база, умения, професии, нови форми на сътрудничество между тези, които притежават данните, и учените и статистиците, нови методи на събиране и обработка на данни, нови начини за вземане на решения.

7. Противоречие има във връзка с нуждата от нови работни места, свързани с новите умения и експертиза, които ще са нужни в обществата на „големите данни“. Тяхното включване като нови източници на данни, допълващи или заместващи традиционните, води до нови методологични, организационни и бюджетни предизвикателства. Усъвършенстването на информационните системи, анализът и експлоатацията на „големите данни“ налагат необходимостта от съвсем нови умения, различни от тези, нужни в традиционната статистика. Все повече ще бъдат търсени аналитици, които имат съответните знания как да събират и да анализират експоненциално нарастващото количество данни. Ще се създават нови модели, техники и подходи

към „големите данни“, чрез които да бъдат оптимизирани всякакви процеси. Старите подходи, свързани с традиционната статистика, все повече ще отстъпват пред алгоритмите и възможността чрез тях да се правят бързи изчисления и прогнози за все по-сложни процеси, изискващи все по-комплексна експертиза.

Възникват нови професии, свързани с данните – „аналитик на големи данни“ (Big Data analytics) и „специалист по големите данни“ (Big Data scientist), „специалисти по визуализация на данни“ (data visualization specialists), както и нови понятия и дисциплини, каквито са „науката за големите данни“ (Big Data science), „извличане на големи данни“ (Big Data mining), „мениджмънт на големи данни“ (Big Data management), „прогнозна аналитика“ (predictive analytics). Става дума за професии, за които са нужни съвсем нови умения – работещите трябва да са способни да извличат и да анализират данни, да прилагат статистически методи и модели, да притежават критично мислене, да са комуникативни, да са в крак с развитието на технологиите, да са запознати с програмирането, математиката, статистиката, алгоритмите, икономиката, финансите.

8. В този контекст възниква *въпросът дали не се движим към едно „постстатистическо общество“* (вж. Davies, 2017). Той е породен от факта, че статистиката, събирана и обработвана от експерти, все повече дава път на данни, които се произвеждат автоматично и бързо в резултат от дигитализацията на всички сфери. В степента, в която отделните хора, градове, сгради, домакинства, обекти се свързват дигитално, те се превръщат в източник на данни. Тези данни се акумулират лесно и се обработват бързо след машинно обучение чрез алгоритми, независимо дали става дума за „Гугъл“ и „Фейсбук“, или за комплекса от електронни данни, които акумулират незабавно банки, правителства, корпорации. Затова в дискусиата за „големите данни“ се срещат прогнози и твърдения, че те са способни не само да подпомогнат или да променят статистиката, но и дори да доведат до нейното отмиране. Някои изследователи откриват в тях и в алгоритмите нова перспектива, която позволява във всеки момент и чрез най-различни устройства, хардуерни и софтуерни продукти, генериращи и анализиращи непрестанен поток от данни, различни процеси да бъдат регулирани в реално време и в перспектива, без за това да са нужни познатите статистически модели и техники, без каквито и да било епистемологични предпоставки. Те виждат в развитието на „големите данни“ и алгоритмите възможността за автоматизация на целия процес, свързан със събирането, сортирането и анализа на информация. Предполага се, че това е тенденция, която ще се задълбочава с нарастването на изчислителната мощ на компютрите и с разгръщането на 5G мрежите. Тези изследователи са склонни да твърдят, че традиционната статистика си отива под натиска на „големите данни“.

Съществува и обратната гледна точка на онези, според които „големите данни“ допринасят за качествен скок във възможностите на традиционната статистика. Аргументът е, че независимо от увеличаването на източниците и субектите на генериране на данни националните статистически институти имат

привилегирован достъп до голяма част от тях благодарение на електронизацията на правителствата и взаимодействието с международни организации и правителства, както и с различни частни субекти. Те са мястото, където се обобщават всички данни с оглед преди всичко на необходимостта от непрекъснато подобряване на дейността на държавните институции в условията на една все по-усложнена и динамично развиваща се реалност, изискваща бързи реакции за предотвратяване на всякакви рискове. Експоненциалното развитие в условията на дигитализация обаче прави всякакви прогнози трудни за реализиране и поради това са възможни най-различни отношения между статистиката и „големите данни“.

В днешно време от съществено значение е да се осъзнае необходимостта от преминаване на нов етап от развитието на статистиката с помощта на „големите данни“, който ще има две основни характеристики: Първо, все повече информация ще се получава на базата не на репрезентативни извадки от изследваните потребители и процеси, а на цялото количество данни за съответните явления. Второ, статистиката на „големите данни“ ще ги предлага не през определени периоди, а в реално време, създавайки по този начин предпоставки за незабавни реакции и възможност за текущи промени в управлението на едни или други процеси.

Статистиката на „големите данни“ е изходна база за рационализация и качествено подобряване на функционирането на държавите, икономиките и цялостните социални системи в условията на експоненциално ускоряване на промените и на генерираните от технологиите на Четвъртата индустриална революция данни. С нейна помощ може да се постигне нормално функциониране на дигиталната държава и дигиталната икономика, без това да доведе до тежки конфликти и разпадни процеси.

Реализацията на тази задача предполага сътрудничество на националните статистически институти със структурите, които боравят с „големи данни“, най-вече с големите дигитални компании. Те разполагат с възможностите да осигурят достъп не само до конкретни данни, но и до такива, които да бъдат използвани като надежден източник за статистически изводи и могат да бъдат достоверно обработени и валидни. В това отношение статистическата обработка може да бъде полезна и за самите компании. Съществено значение има и взаимодействието с академичните институции, които трябва да решават методологическите и техническите проблеми на обработката на неструктурирани данни. Този проблем става все по-значим за международните организации като Европейския съюз, което е предпоставка в рамките му да започнат да се разработват колективни модели за използване и анализ на „големи данни“ в различни области.

При наличието на огромен и растящ брой източници на всякакви данни, често конкуриращи се и разминаващи се помежду си, националните статистически институти биха могли да се разгръщат като институции, които предлагат достоверни данни за научен анализ и управленски решения. Във връзка с

това през 2014 г. на 45-ата сесия на Статистическата комисия към ООН е създадена Глобална работна група по използване на „големите данни“ за целите на официалната статистика и е признато, че те представляват източник на информация, който не може да бъде игнориран. Година по-рано ръководителите на националните статистически служби в ЕС подписват т. нар. Схвенингенски меморандум за „големите данни“ и официалната статистика, ангажирайки се да ги интегрират официалната статистика. През 2013 г. Евростат стартира проекта ESS (European Statistical System) Vision 2020,² който има за задача да установи кои могат да са основните източници на „големи данни“ и как те да бъдат интегрирани с официалните статистически системи. Подчертано е, че използването на „големите данни“ ще става в съчетание с вече съществуващите източници на информация. Идеята е да се извлича информация от интернет за свободни работни места, за характеристики на предприятията, да се събират данни от автоматични идентификационни системи и мобилни телефони, от измервателни устройства за потребление на електричество и т.н. Така статистическите данни ще се генерират по-бързо и с по-добро качество и ще се правят по-точни прогнози. За тази цел „големите данни“ трябва да станат интегрална част от официалното събиране на данни.

9. *„Големите данни“ са фактор за получаване на огромно количество статистическа информация, което, от една страна, води до оптимизация на управленските системи, но от друга, се смята, че именно това създава рискове от технократура, контролираща всеки аспект на човешкия живот.* Анализ на тези рискове прави Шошана Зубоф в своята превърнала се в бестселър книга „Епохата на надзираващия капитализъм“, където се развива идеята, че се върви към поредната трансформация на капиталистическата система, определена от автора като „капитализъм на надзора“. Той се осъществява от големите дигитални компании, които посредством „големите данни“, генерирани чрез търсачки, социални медии, хардуерни и софтуерни продукти, наблюдават поведението на хората, насочват го и научават почти всичко за тях – често дори повече, отколкото потребителите знаят за себе си, като тази информация се превръща в нова форма на експлоатация. Така всъщност наблюдението върху човека и неговото поведение се превръща в бизнес, генериращ милиарди долари печалби за дигиталните гиганти, които търгуват с тази информация, контролират и насочват човешкото поведение. При това, докато държавата малко или много е ограничавана в тази насока посредством механизмите за баланс и контрол, частните корпорации трудно се поддават на публичен контрол. Те се превръщат в един своеобразен „Big Brother“, знаещ благодарение на „големите данни“ повече, отколкото Оруел някога си е представял в знаменития си роман „1984“. Става дума за тотални мониторинг и контрол върху поведението на всеки потребител, за програмиране с помощта на алгоритми на неговите чувства, мисли, потребности.

² През 2015 г. към този проект като партньор се включва и НСИ на България.

Днес, с разгръщането на дигиталната или Четвъртата индустриална революция, на „големите данни“ и на алгоритмите, „невидимата ръка на пазара“ изглежда все по-видима чрез данните. Те се превръщат във фактор, който насочва производството, размяната, разпределението и потреблението. Така се разгръщат по същество процеси на програмиране на капиталистическата система, особено що се отнася до големите дигитални компании, при които не пазарът, а данните за потребителското поведение придобиват водещо значение. Все повече процеси, които преди са били незабележими, стават явни, и все повече сигналите на пазара отстъпват пред анализа на информация, което води до преход от пазарна към цивилизация на „големите данни“. Ускореното въвеждане на машинното обучение и на изкуствения интелект при обработването и използването на „големите данни“ в реално време за вземане на рационални икономически решения става предпоставка за радикални промени в цялостното функциониране на политическите и икономическите системи, което може да получи алтернативни измерения в зависимост от това кой има достъп до и борави с такъв тип данни. Анализът на тези алтернативни варианти на използване на изкуствения интелект при обработката на статистиката на „големите данни“, съответно от малка група хора или от цялото общество при равенство в правата и възможностите на всеки индивид, тепърва предстои (Todorova, 2019).

*

Използването на „големи данни“ дава възможност за качествен скок във възможностите на традиционната статистика, на политиката, на икономиката и на всяка сфера, която изисква вземането на управленски решения. Увеличават се източниците и субектите на генериране на данни, а националните статистически институти събират растящо количество информация чрез електронизацията на правителствата и взаимодействията с международни организации и правителства, както и с различни частни субекти. Те се превръщат в място, където се обобщават все повече данни с оглед преди всичко на необходимостта от непрекъснато подобряване на дейността на държавните институции в условията на една по-усложнена и динамична реалност, изискваща бързи реакции за предотвратяване на всякакви рискове, опасности, кризисни процеси, с които се сблъскват икономиките и обществата, представляващи все по-неравновесни и нелинейни системи.

Така се ражда и базовият диалектически парадокс на „големите данни“ и алгоритмите – те са способни да допринесат за рационализирането на ключови процеси в обществата и в икономиката, да насочват производството, размяната, разпределението и потреблението, така че да оптимизират целия производствен цикъл; могат да дадат отговор на фундаменталните за икономиката въпроси какво, как, колко и за кого се произвежда; притежават възможността да подпомагат политиката, прогнозирането, изграждането на репутация, но в същото време именно в тяхната способност да оптимизират всичко се

крият и най-големите рискове. Както лъкът и стрелата в първобитното общество са можели да служат за набавяне на прехрана също толкова, колкото за извършването на убийства, така и анализът на „големи данни“ в началото на XXI век разполага с потенциала да доведе не само до подобряване и оптимизиране на основни сфери от политическия, икономическия и социалния живот, но и до нови форми на контрол върху човека, до нови технологични войни и конфликти, каквито вече са познати от историята. Затова важни са не „големите данни“ сами по себе си, а това как ще се използват и как ще бъдат ограничени рисковете от злоупотреби с тях както от страна на отделните частни субекти, така и от различните държави.

Използвана литература:

Daas, P. J. H. and M. J. Puts, B. Buelens, P.I. A.M. van den Hurk (2015). Big Data as a Source for Official Statistics. *Journal of Official Statistics*, Vol. 31, N 2.

Davies, W. (2017) How statistics lost their power – and why we should fear what comes next. *The Guardian*, January 19.

Giczi, J. and K. Szoke (2018). Official Statistics and Big Data. *East European Journal of Society and Politics*, Vol. 4, N 1.

Hammer, C. L. and D. C. Kostroch, G. Quirós, STA Internal Group (2017). *Big Data: Potential, Challenges, and Statistical Implications*. IMF

Hughes-Cromwick, E. and J. Coronado (2019). The Value of US Government Data to US Business Decisions. *Journal of Economic Perspectives*, Vol. 33, N 1.

Struijs, P. and B. Braaksma, P. J. H. Daas (2014). Official statistics and Big Data. *Big Data & Society*, April-June.

Todorova, M. (2015). Counterfactual Construction of the Future: Building a New Methodology for Forecasting. *World Future Review*, Vol. 7(1), p. 30 –38

21.06.2019 г.