

ВЪЗДЕЙСТВИЕТО НА ГОЛЕМИТЕ ДАННИ (BIG DATA) ВЪРХУ ОФИЦИАЛНАТА СТАТИСТИКА: ВЪЗМОЖНОСТ ИЛИ ПРОВОКАЦИЯ¹

Богдан Богданов, Галя Статева***



„Целта се постига по-лесно не като се бърза, а като не се спира.“

Плутарх

Въведение

Известно е, че фрагменти от статистиката като наука и практика възникват още в древността поради необходимост от по-добро държавно управление. Като целенасочена дейност тези фрагменти се превръщат в стройна система в края на XVIII век. Постепенно статистиката се превръща в официално призната дейност в целия свят. Вече в продължение на повече от два века статистиката като наука и практика е неотделима част от средствата, необходими за управлението на държавата. С лавинообразното разрастване на информационните технологии и начините за създаване на данни официалната статистика е изправена пред необходимостта от реформиране. Един важен аспект на реформата е свързан с възникналото през последните години ново явление в социално-икономическата наука и практика, получило наименованието Big Data. Буквалният превод на

* Д-р, заместник-председател на НСИ; e-mail: bbogdanov@nsi.bg.

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

¹ Участие на авторите при написването на статията: д-р Б. Богданов - въведение, част III и заключение; Г. Статева - част I и част II.

понятието на български език е „големи данни“, но то вероятно не отразява толкова мащабно и експресивно това, което се съдържа в него. Въпреки това, по-нататък в изложението ще използваме понятията Big Data и „големи данни“ като взаимнозаменяеми, тъй като и двете понятия са придобили масова популярност сред читателите.

Според американски и европейски експерти Big Data са достатъчно значими и вездесъщи в съвременното информационно общество. Тези данни са с огромен обем, разнообразие и скорост, като разкриват нова информация, факти, връзки и отправни точки, които никога не са съществували в миналото. Когато тази нова информация се обхване и се управлява по подходящ начин, може да се превърне в голяма сила, която ще повиши ефективността на държавното управление и знанията за обществото. Очаква се, че чрез тази информация представата за света с нарастващо население ще доведе до управление, което ще редуцира напрежението в предоставянето на услуги и инфраструктура; тотално ще подобри здравеопазването; интензивно ще отстрани заплахите в сигурността на обществото и националните граници. Целта е да се намерят адекватни подходи и средства за събиране, обработване, съхранение и анализиране на големите данни, което гарантира ефективното им използване (възможности и предизвикателства) и включването им в областите за управление на държавата и обществото.

Теоретичните постановки, мнения и заключения на експертите могат да се обобщят и представят по следния начин:

1. Да се разбере **изкуството на възможностите**. Така ще се придобие опит, който липсва сега.
2. Да се идентифицират от два до четири **ключови бизнеса** или мисии, изискващи използването на Big Data.
3. Да се направи **инвентаризация** на „белите петна“ в информационните масиви.
4. **Оценка на експертните способности и ИТ средства**, необходими за постигане на целите.
5. **Достъп до информация с източници Big Data** с цел да се предизвика използване на опита на външни бизнес потребители.

Политиката по отношение на Big Data трябва да бъде следната:

1. В областите, където е необходимо използването на Big Data, **да се премахват бариерите** за тяхното разбиране, изучаване и сертифициране.
2. Да се постави **акцент върху възможностите** за техния анализ.

3. Да се установи трайна и **устойчива връзка между бизнеса и академичния подход** за използване.

4. Да се насърчават **стратегии за постоянно изучаване** и прилагане на нови техники и средства за усвояване на данните.

5. Да се разработят **правила и да се установи сътрудничество** с бизнеса и притежателите на данни, за да се осигури конфиденциалност и защита на данните, като се създаде необходимата култура за тяхното използване.

I. Дефиницията за Big Data и видимите хоризонти за развитие

На тази основа се представят теорията, разбирането и **дефиницията за Big Data**. Естеството на Big Data показва, че те не трябва да се разглеждат като технологичен, а по-скоро като съдържателен феномен от обширно количество необработена информация, пресичаща общественото пространство и генерираща се чрез бизнеса и държавните организации. Като пример може да се посочи, че през 2011 г. е създадена 1.8 зетабит глобална информация, като ежегодно тази информация удвоява обема си. По този повод възникват въпросите: Как да се натрупва, управлява и използва огромното количество данни, произвеждани в реално време? Как да се осигури тяхната защита? Как да се подобри процесът на организиране на данните и тяхното разпространение за разширяване на разумната връзка с тях? Как да се изгради потребителско доверие към данните чрез опознаване на тяхното естество и произход? Какви са най-добрите начини за визуализация, средства и форми за представяне на данните, за да се осигури бърз достъп за анализ? Как да се направи мост между таланта и човешкия капитал, за да се постигнат позитивни цели?

Дефиницията за Big Data може да се представи по следния начин:

Big Data е понятие, което описва голям обем на високоскоростен набор от постоянно променящи се данни, изискващи модерни средства и технологии за обработване, съхранение, разпространение, управление и анализ на информацията.

Основните четири характеристики на Big Data са: скорост (Velocity), обем (Volume), разнообразие (Variety), истинност (Veracity) (Crawford, 2011, с. 11). Може да се отбележи, че само 15% от информацията е структурирана. Останалите 85% от информацията е неструктурирана като тази в социалните мрежи, колцентровете, сензорните устройства, мобилните телефони, сателитите и

т.н. Това е информация, която в своето ускорено движение непрекъснато променя формата и конфигурацията си, задоволявайки потребностите на обществото. Нейното систематизиране и „овладяване“ ще предотврати хаоса в информационното пространство. Успоредно с това ще намали драстично опасността от недостоверна и вредна информация, като усили позитивната енергия, съдържаща се в данните. Феноменът Big Data съществува обективно и от изследователите зависи какво ще бъде неговото влияние в обществото и света.

Дескриптивно представяне на Big Data

Характеристика	Описание	Атрибути	Управление
Обем	Данни, които интензивно се натрупват, за да бъдат анализирани и предопределят вземането на определени решения.	Експоненциално нарастване на информацията в световен мащаб.	Нарастване на източниците на данни с висока резолюция на сензорите.
Скорост	Данните се придвижват с висока скорост, променяйки се в динамика.	Достъпност: информацията за потребителите трябва да бъде на необходимото място в необходимото време. Приложимост: релевантност на феномена, за който са данните. Полезно време: реално време на данните и анализа.	Нарастване на източниците на данни. Подобряване на връзките. Всестранно нарастване на силата на произвежданите данни чрез умножаване на средствата.
Разнообразие	Нарастване на източниците на информация отвътре и отвън. Това оказва натиск върху възможностите на информационните технологии.	15% от информацията е структурирана. 85% от информацията не е структурирана. Самоструктуриране: комбинирането на структурирани и неструктурирани данни става върхово постижение. Комплексност: когато данните се придвижват и остават.	Мобилни средства Сензори Социални мрежи Сателити Видео
Истинност	Качество и произход на получаваните данни.	Качеството на данните зависи от дефинирането, комплексността, описанието, латентността, модела за апроксимация.	Базите данни изискват проследимост и уточняване.

Според експертите на OECD² съществува и друга характеристика, която допълва същността на Big Data - ценност (Value) (Exploring data..., 2013, с. 12). Тази пета характеристика за големите данни е свързана с нарастване на социално-икономическата стойност, получена от използването им. Налице е нов фактор, който предполага производство на продукт с определена цена. Това, от своя страна, показва, че обективно съществуват нови сегменти в информационните пазари. Подобно обстоятелство дава огромни възможности за развитие на частния бизнес. По този начин се създава нова икономическа култура в глобалния свят, която е шанс за страните независимо от техния икономически статус и степен на развитие, т.е. вече и по-слабо развитите икономически страни имат по-голям шанс за ускоряване на своето развитие чрез Big Data. Това е нова информационна среда за управление на финансовите потоци, инвестициите, работната ръка и обмяната на материални ресурси, стоки и услуги. На практика Big Data в определена степен провокират разработването на подходящи електронни средства и ИТ инфраструктура, чрез които става възможно получаването на такъв тип информация. Например сензорните устройства могат да се използват с еднакъв успех в кухните на домакинствата, болничните стаи, производствата, опасни за човешкото здраве, трафик контрола по пътищата и магистралите. Друг пример са образованието и пазарът на труда, където Big Data ще бъдат обект на изучаване и възможност за развитие на нови техники и средства за анализ. В този смисъл потокът от големи данни очертава широк спектър от стратегии за постигане на различни цели. По този начин анализът на събраната информация ще рефлектира в позитивна светлина върху живота на обществото, подпомагайки разработването на възможно най-добрите политики за действие.

Съществен момент при използването на Big Data е изграждането на необходимата архитектура. Основната структура на архитектурата е: интернет/интранет, сателити, социални сайтове, видео/картини, мобилни средства, документи. Всеки елемент на тази структура предполага включване на подструктура, която може да има разклонения в различни направления. По този начин се създава спектър, където се съчетават бизнес интереси, социално-икономически варианти на държавно управление и обществени потребности. Парадигмата на Big Data е съчетание от интереси на всички нива в хоризонтален и вертикален аспект, включващи потребностите на обществото в глобалната икономика.

² Organisation for Economic Co-operation and Development.

В процеса на създаването на архитектурата на тези данни важен момент е успоредното разработване на нови средства за екстрахиране на информацията в съчетание с добре познатите средства за постигане на планираните цели. Съхраняването на данните е съществена част от целия процес, тъй като се създава възможност за: тяхното многократно използване; съчетаването им с други източници на информация, вкл. статистически; сравнения във времето и пространството; проследяване на динамиката на изследваните явления и процеси. Във връзка с това изключително важен е процесът по изграждане на специални информационни центрове, където може да се споделя и стандартизира общата ИТ инфраструктура за използване на Big Data.

Редица страни използват Big Data при решаване на задачи от икономиката и управлението на обществото. Така например интересен е моментът при възлагането на съществена роля на сателитните снимки за селското стопанство и екосистемите. Информацията от тези снимки дава значими данни за посевите и растенията в страни като Австралия, Китай, Мексико, Колумбия и други. В Нидерландия се провеждат обещаващи опити за изследване чрез социалните мрежи на човешкото поведение, потребителските интереси, емоции и чувства. В Италия и Китай чрез тези данни се правят оценки за броя на свободните работни места и по този начин се подпомага официалната статистика. Швейцария и Швеция имат добър опит и партньорство при оценката на индекса на потребителските цени. От тази гледна точка следва, че организирането на процесите при разработване на потоците от данни в национален и глобален аспект поражда редица предимства за създаване на добри практики и партньорство без загуба на време, човешки и финансови ресурси. Съществен момент при използването на Big Data е изграждането на обществено доверие към тяхното прилагане. Очаква се Big Data да „попълнят“ белите информационни полета, до които по една или друга причина (най-често финансова) официалната статистика не може да достигне чрез провеждане на традиционните статистически изследвания. Това се отнася особено до малките териториални отграничения (области, ареали, общини и т.н.). Големите данни предизвикаха промени и в конюнктурата на пазара на труда. Непрекъснато нараства броят на т.нар. научни изследователи за работа с данни (*data scientists*) и администраторите на данни, което изисква специфични знания и умения. Особено чувствителни в това отношение към момента са пет сектора: публичната администрация, създаването на реклами, здравните грижи, предоставянето на услуги, логистиката и транспортът (Exploring data..., 2013, с. 13). В недалечно бъдеще се очаква разширяване на този обхват.

II. Big Data: възможност или заплаха за официалната статистика

Сливането на технологии и социално-икономически тенденции, в т.ч. интензивната миграция на социална и икономическа активност в интернет и намаляването на разходите за получаване на данни, транспорт, съхранение и анализ, води генерално към огромен обем от данни - обобщено представени като Big Data, където има възможност да се разработят нова индустрия, процеси и продукти на ниска цена. Очевидно е, че при сравненията между разходите на средства за Big Data и данните от официалната статистика при анализа на конкретни и важни социално-икономически области ще се докаже еднозначно, че разходите на средства ще бъдат значително по-високи, ако се използва класическият и традиционен начин за получаване на данни чрез провеждане на статистически изследвания.

На редица работни срещи и конференции се разглеждат възможностите и предизвикателствата при използване на новите източници на информация, наричани Big Data. Основният въпрос, който възниква, се отнася за кохезията на тези данни с официалната статистика. По-конкретните въпроси са:

- Каква може да бъде ролята на официалната статистика в глобализирания се свят при непрекъснато нарастващите нови източници на информация?
- Дали натискът за промяна на официалната статистика е продуктивен?
- Новият поток от източници на данни, достъпни чрез различни средства, може ли веднага или в краткосрочен план да подмени официалната статистика?

Към настоящия момент релефно се открояват три основни групи източници на Big Data. Първият успешен опит за класифициране на типовете източници на Big Data е направен през 2013 г. от работната група към Статистическата комисия на ООН:

1. Социални мрежи (информация, генерирана от населението): тази група съдържа информация, която по същество е запис на човешкия опит; в миналото е описвана в книги и произведения на изкуството, фотографии, аудио- и видеозаписи. В наши дни информацията е почти изцяло дигитализирана и се съхранява навсякъде: от персоналните компютри до социалните мрежи. Данните са неструктурирани и са трудно обработваеми и управляеми.

1100. Социални мрежи: Facebook, Twitter, LinkedIn и др.

1200. Електронни блогове

1300. Лични документи

1400. Снимки: Instagram, Flickr, Picasa и др.

- 1500. Видеоклипове: Youtube и др.
- 1600. Интернет търсачки
- 1700. Данни от мобилни телефони: текстови съобщения
- 1800. Електронни карти
- 1900. Електронна поща

2. Традиционни бизнес системи и уебсайтове (данни, генерирани от информационни системи): тази група съдържа информация, която е в резултат от стопанска дейност като например регистриране на потребители, производство на продукти, заявки, финансови трансакции и други. Данните са структурирани в определен формат, предимно табличен, дефинирани са връзките между тях и метаданни за тяхното съдържание. Съхраняват се в релационни бази данни или системи (някои източници, принадлежащи към този клас, могат да попаднат и в категорията „Административни данни“).

21. Данни, генерирани от държавни агенции

2110. Медицински записи

22. Данни, генерирани от бизнеса

2210. Търговски трансакции

2220. Банкови записи/борсови операции

2230. Електронна търговия

2240. Кредитни карти

3. Интернет на нещата (данни, генерирани от машини/сензорни устройства): тази група съдържа информация, генерирана от машини и сензорни устройства, които измерват и записват събитията и ситуациите във физическия свят. Данните са добре структурирани и са подходящи за компютърна обработка, но размерът и скоростта им са извън възможностите на традиционните методи за обработка на данни.

31. Данни от сензори

311. Постоянни сензори

3111. Домашни автоматизирани системи

3112. Сензори за времето/замърсяването

3113. Пътни сензори/уебкамери

3114. Температурни сензори

3115. Охранителни/наблюдателни видеокамери

312. Мобилни сензори (проследяващи устройства)

3121. Локация на мобилни телефони

3122. Навигационни системи за автомобили

3123. Сателитни снимки/изображения

32. Данни от компютърни системи

3210. Логове

3220. Уеблогове

Това разделение е важно, тъй като по този начин се определя стратегията за разработването на тези източници и се дефинират познавателни задачи за бизнес процесите в различни социално-икономически сфери.

Особено важен момент се съдържа във възможността за търсене на отговор на въпроси, по които до момента няма ясна и точна информация и данни в общественото пространство. Например: Как скритата икономика влияе върху икономическата, държавната и ценностната система в обществото? Какви са негативните и позитивните страни на глобализацията в стойностно и морално изражение? Какви са оптималните/критичните граници на неравенството на населението по доход, които предопределят просперитет или срыв? Как новопоявяващи се феномени като миграционни потоци, замърсяване на околната среда, тероризъм, локални войни и много други променят света? Това са въпроси, на които официалната статистика трудно може да даде отговори. При изследванията в тези области се намират данни и частични информационни сегменти, очертаващи процеси, притежаващи мощна сила, която не се контролира и управлява с определени политики. Във връзка с това следва да се отбележи, че анализът на Big Data може предварително да фокусира вниманието към зараждащите се явления и процеси, които изискват редуцирането на негативи и своевременно предотвратяване на нежелани последици за обществото. И обратното: към засилване на позитивните страни на явленията и процесите, създаващи условия за просперитет. При официалната статистика традиционните изследвания по-трудно и бавно откриват важни за обществото моменти от неговото развитие в динамика. Много често това обстоятелство се свежда до констатиране на факти след натрупани негативи, което по същество ограничава възможността за превантивни решения. В крайна сметка анализът на събитията е осъществен след голям разход на

време и средства и е относително закъснял или подранил чрез загатнат (трудно разбираем) силует на предстоящ общественно-икономически проблем.

Съществен момент при големите данни е наличието на предизвикателството и потенциала, които носят със себе си по отношение на търсенето на решения за обработка и анализ, подобрявайки съществуващата практика. Споменатите вече основни характеристики на Big Data са резултат от постоянното нарастване на информационните потоци в най-различни сфери и развитието на техническите средства за комуникация. В този смисъл експертите сравняват Big Data със създаването на ядрената енергия в началото на XX век. Това е причината да възникнат логичните въпроси: Могат ли тези данни да бъдат контролирани в съответствие с етическите измерения в обществото? Може ли да се носи отговорност, когато те се използват, за да няма поражения в общественото пространство? (Exploring data..., 2013, с. 21)

Сега се пишат теорията и практиката на Big Data. Редица работни групи в международните икономически организации, университети, изследователски организации и институции (Евростат, националните статистически институти, OECD, Световната банка, всички международни организации, правителствата на различни страни, международни компании и др.) дефинират основните принципи на работа, които обхващат различни аспекти: същност, съдържание, качество, технология, правна рамка, интегрираност, конкурентоспособност, финансиране, добри практики, сътрудничество, отговорности, етичност, демократичност, преоценки на риска, сигурност на управление, международен достъп и използване, цялостното бъдеще на тези данни. Стремежът е тези принципи да се материализират в информационни системи и мрежи, отворени, организирани, защитени и даващи възможност за използване и обмяна. Тези дейности се предопределят от потенциалните възможности, които се съдържат в Big Data и които нарастват непрекъснато във времето и пространството. В конкретен вид могат да се обобщят по следния начин:

- Създаване на нова статистика.
- По-голяма детайлизация на съществуващата статистика.
- Статистика, която е навременна - получава се в кратки интервали от време.
- Статистика, която позволява да се правят прогнози и да се разработват показатели за кратки периоди от време.
- Статистика, позволяваща да се подобри качеството на данните.
- Редуциране на равнището на неотговорилите респонденти.

- Редуциране на разходите за статистически изследвания и повишаване на тяхната продуктивност и достоверност.

Разрастването на информационните технологии и потоци от данни следва да се разглежда по-скоро като необходимост от цялостна реформа, а не само като необходимост от разширяване на обхвата на официалната статистика. Това означава разработване на методика за контролирана синергия между Big Data и официалната статистика. Това е производствен процес, който има обща рамка, включваща необходимия инструментариум, технологични средства и принципи за анализ. Във връзка с това трябва да се обърне внимание, че сега съществуващият „Общ модел на статистическия производствен процес“ („Generic Statistical Business Process Model“ (GSBPM) е описание на технологията за производство на данни в съответствие с дейността на официалната статистика към момента. Комбинирането на изследвания с използването на административни източници и Big Data ще промени тотално статистическата култура и практика на националните статистически институции. Очевидно е, че принципите, същността и спецификите на процеса ще бъдат актуализирани в зависимост от поставените цели общо и от конкретната област на изучаване. В недалечно бъдеще може да се очаква, че ежегодно планираните изследвания на официалната статистика ще бъдат постепенно комбинирани, сегментирани и заместени с източници на Big Data (включително данни от административните източници). Планирани изследвания ще се правят по определена тематика с допълнение на Big Data, което очевидно ще измести начините и подходите за получаване на информация. Успоредно с това трябва да бъдем готови и за обстоятелства, при които иновативните подходи за работа с големите масиви от данни няма да позволяват коректни сравнения с резултатите от минали периоди на изследване. С други думи, сравненията на показателите в динамичните редове ще бъдат валидни с известни условности и ограничителни условия. Това е неизбежно, като се вземат предвид глобалните промени, следващи естествения ход на времето. Успоредно с това възникват редица въпроси за валидирането на данните, тъй като ще нарасне опасността от изместване на оценките като резултат от отсъствието на еднозначно дефинирана генерална съвкупност в съответствие с теорията на извадковите изследвания (Cooper, 2013, с. 5).

Наред с това трябва да се отбележат и някои негативни тенденции, които се открояват и сега при използването на големите данни. На първо място е необходимостта от филтриране на информацията, за да не се допуска дезинформация в потока от данни. На второ място е възможността за натрупване на данни, чрез които се осъществява манипулация и насаждане на негативни настроения в определени слоеве от населението (например расова и религиозна

нетърпимост). Не на последно място се откроява и възможността за създаване на складове с данни, които съдържат информация, деструктурираща обществото (например страх от тероризъм и заплахи от ядрена световна война). В крайна сметка се налага изводът, че внимателният подход към бъдещите информационни системи изисква:

- Материализиране на новите идеи, но без да се бърза с отхвърлянето на старите методи.
- Погледът към бъдещето не означава забравяне на миналото, а надграждане на постигнатото.
- Непрекъснато натрупване на нови познания и умения.
- Да не се подценява стойността на създадените добри теории и практики при прехода към иновации.

Не трябва да се забравя, че внедряването на Big Data за статистически цели изисква използването на точен и ясен механизъм за описание на метаданните. Изследователите сравняват големите данни със самородна буца злато, но са наясно, че тяхната навигация за целите на държавното управление е дълъг и труден процес, изискващ много умения, знания и визия за развитие.

Някои национални статистически офиси вече осъществяват европейски и международни проекти за големите данни. Техният опит от успешното използване на големите данни може да бъде изучаван и споделян с други страни с цел извличане на ценни познания и прилагане на добри практики по отношение на източниците на големи данни. Освен това националните статистически организации са окуражавани от Евростат да включат официално въпросите за големите данни в техните годишни програми и стратегически документи чрез осъществяване на изследователски и пилотни проекти в избрани области и чрез разпределяне на подходящите ресурси за тези цели. В процеса на приложение на Big Data в официалната статистическа практика Евростат играе ключова роля, предвид факта, че той е източник на финансови, идейни и технически ресурси, включително организатор на редица специализирани обучения, семинари и научни конференции по тази тематика. В допълнение, за да посрещне предизвикателствата и да създаде подходящи условия за сътрудничество между частните единици и националните статистически институти за достъп до нови източници на данни, Евростат е разработил пътна карта и план за действие. Всички заинтересовани страни от политиката, науката, статистиката и бизнеса, както и обществото като цяло ще бъдат привлечени. Целта на пътната карта е да направи възможно постепенното

интегриране на източниците на големи данни в производството на европейска и национална статистика в дългосрочен, средносрочен и краткосрочен аспект, като по този начин ще допринесе значително за постигане на целите на ЕСС Визия 2020.

Бъдещето на официалната статистика в ерата на големите данни е все още въпрос на обсъждане и експериментиране. Но това, което със сигурност е ясно, че цялата международна и европейска статистическа общност е необходимо да се адаптира към новата реалност и да отговори адекватно на възможностите и предизвикателствата, които тази реалност предлага. За да се случи това, е необходимо широко и ползотворно сътрудничество с различни „играчи“ във и извън статистическата общност чрез формиране на разнообразни мрежи, които могат да изградят нови пътища за генериране на статистически данни. Това е гаранция за успешно обществено развитие в условията на глобалната икономика.

III. Глобализация и Big Data

Връзката между глобализацията и Big Data е взаимозависима. Разрастването на информационните технологии и създаването на огромни потоци от информация предопределят растежа на глобалната икономика. И обратното, глобалната икономика не може да се развива без информация, която направлява в необходимата посока пари, ресурси и работна сила. Забележителното е, че тази информация се създава където и когато данните на официалната статистика не са достатъчни и трябва да се реагира бързо. Процесът на глобализация в икономически аспект може да се дефинира като неизбежна интеграция на пазари, държави и технологии в степен и мащаби, недостигани досега в човешката история. Основната концепция на глобалната икономика е свободният пазар, което предполага оттегляне на държавата от икономическия живот на отделната страна. Модерната и скоростна световна икономика изисква и промяна в качеството на държавното управление - умна и ефективна администрация, която може да регулира свободния пазар, без да го задушавя, но и без да изпуска контрола върху него. На практика това означава установяването на стабилна финансова система и създаването на необходими условия за привличане на инвеститори и насърчаване на частния бизнес. Тази ситуация очертава три основни аспекта по отношение на изискванията към бъдещото развитие на статистиката:

Първо: Когато се използва статистическият инструментариум за индикиране на развитието на пазарната икономика и приближаването на страната ни към Европа и към икономически развития свят, следва да се наблюдават редица процеси като: обхвата на частния сектор, равнището на инфлация и стабилност на цените, балансираността на държавния бюджет, значението на

премахването и намаляването на митата върху вносните стоки, влиянието на отстраняването на ограниченията върху частните инвестиции, а също на квотите и домашните монополи, увеличаването на износа, приватизацията на домашната индустрия и обществените услуги, отказа от регулирането на капиталовите пазари, превръщането на валутата в конвертируема, премахването на проблемите за навлизане на преки чужди инвестиции, редуцирането на корупцията и организираната престъпност, ограничаването на правителствените субсидии и паричните помощи, отварянето на банковата и телекомуникационните системи за частния капитал и конкуренцията.

Второ: Когато се обработва статистическата информация, следва да се вземат предвид многообразието, количеството и уникалността на наблюдаваните процеси. Това означава решаването най-малко на три първостепенни задачи: получаването на изходи с висока аналитична стойност; възможност за международна сравнимост на данните; съхраняване на първичната информация във вид и при условия, позволяващи коректно формиране на дълги динамични редове за проследяване на генезиса и развитието на наблюдаваните процеси.

Трето: Динамиката и естеството на процесите поставят съответни изисквания по отношение на синтеза и комплексността на статистическия анализ. Статистиката като наука и практика има съществено значение за количеството, качеството и бързината на достъп до информацията. Парадигмата на света, в който живеем, се усложнява, видоизменя и разнообразява непрекъснато. Днес повече от всякога традиционните граници между различни аспекти на общественото развитие - политика, култура, технологии, национална сигурност, финанси, образование и опазване на околната среда, се размиват и изчезват. В хода на статистическите изследвания, анализ и коментар на данните става невъзможно обяснението на едно явление, събитие или процес в дадена област, без да се търси причинно-следствена връзка, доказателство и синтез с другите. Очевидно лаговете на изменение и влияние както на опосредствани връзки, така и на едностранни, независими перспективи на промени се съкращават непрекъснато. Умението да се вплете най-същественото и значимото, когато се рисува картината на света, има първостепенно значение, тъй като се създава реална представа за мястото, ролята и насоките за движение на всяка отделна страна в новия свят на глобализация.

Сложността и скоротечността на изследваните процеси промениха статистическата система. Каква трябва да бъде статистиката в новото информационно общество? Един възможен и предизвикателен отговор: **внедряване и използване и на Big Data в развитието на официалната статистика.**

Развитието на индустрията постави въпроса за необходимостта от информация за управление на икономиката и обществото. Това утвърди и статистиката като институция, където се създават методологията, организацията и крайният информационен продукт. Интересно е да се отбележи, че в миналото тези дейности са се осъществявали от различни колективи от специалисти и между тях са съществували отчетливо забележими граници. Новите информационни технологии размиха тези граници, съкратиха времето за получаването на информационния продукт и неговото предоставяне на потребителите. Успоредно с това интернет създаде условия за събиране на хора с познания в едно виртуално общество за обмен на информация. С други думи, интернет се превърна в инструмент за формиране на информационното общество, а то, от своя страна, се превърна в двигател на глобализацията като процес. Тази ситуация предопредели съвършено новата роля на статистика професионалист през новото хилядолетие. Новият информационен пазар ще изисква от него да синтезира и обединява отделни фрагменти информация, да извлича най-същественото от съществуващите информационни масиви и бази данни, да бъде методолог с нестандартно мислене, което означава да оценява значимостта и достоверността на определени изводи, да балансира, претегля и ранжира фактите, да бъде арбитър при използване на информация от различни източници. Именно тези източници формират феномена Big Data.

В своята книга „Трусове във властта“ А. Тофлър отбелязва: „Електронната инфраструктура на напредналите икономики ще притежава шест характерни белега, някои от които вече започват да се проявяват. Шестте ключа към бъдещето са: интерактивност, преносимост, преводимост, съвместимост, вездесъщност и глобалност. Влезли в органично съчетание, тези шест принципа водят до тотална трансформация не само на способите, по които влизаме във връзка един с друг, но и на мисленето ни, на начина, по който виждаме себе си в света и вследствие на това оценяваме взаимоотношенията си с различните власти“ (Тофлър, 1996, с. 401). Спецификата и скоростта на протичащите процеси създават прецеденти, при които националната статистическа система невинаги е готова с адекватно информационно осигуряване. Отсъствието на готовност много често е свързано с липсата на финансови средства, тъй като става дума за уникалността на нововъзникващи и важни за обществото процеси, както и за необходимостта да се подготви информация за малки подсъвкупности от населението, които невинаги са леснодостъпни. Може да се каже, че с нарастване на разнообразието на протичащите процеси намалява хомогенността на обектите за наблюдение, т.е. признаковото пространство нараства и се променя сравнително по-бързо отколкото обектите на статистическо наблюдение. Това означава, че акцентът на статистическата дейност постепенно

трябва да бъде преместен от мащабни и регулярно осъществявани изследвания към въвеждане на нови подходи за екстрахиране на информация, максимално използване на данни от ограничени по обхват наблюдения и достигане на по-голяма дълбочина и детайлизация на характеристики чрез използването на статистико-математически модели. В този смисъл може да се отбележи, че **на дневен ред е предизвикателството Big Data**. В следващата таблица е показан SWOT анализ за големите данни, който обобщава в кратка форма основните изводи по отношение на възможностите за тяхното използване.

SWOT анализ на Big Data

	ПОЛОЖИТЕЛНИ ФАКТОРИ	ОТРИЦАТЕЛНИ ФАКТОРИ
ВЪТРЕШНИ ФАКТОРИ	<p style="text-align: center;"><u><i>СИЛНИ СТРАНИ</i></u></p> <ul style="list-style-type: none"> ✓ Многообразие на данни в най-различни области на живота. ✓ Източник на информация в реално време, която позволява чрез анализ да бъде начало на позитивни и креативни решения в различни сфери на общественото пространство. ✓ Позволява селектирането на неподправена и достоверна информация, създаваща благоприятна среда за устойчиви и дългосрочни решения за развитие. ✓ Данни, които могат да окажат съществено влияние върху представите ни за света и да променят в голяма степен стереотипа за човешкото мислене и поведение. ✓ Информация, която е предпоставка за следваща визия в човешкото развитие изобщо и в частност на информационните технологии. ✓ Подпомага развитието на иновативни подходи, включително сензорни устройства и технически средства, които ускоряват 	<p style="text-align: center;"><u><i>СЛАБИ СТРАНИ</i></u></p> <ul style="list-style-type: none"> ✓ Превръщане на Big Data в средство за наблюдение и проследяване. ✓ Разнообразието на информационните потоци изисква задълбочени познания за използването и развитието на съвременните информационни технологии. ✓ Заплаха за провеждането на демократични реформи в обществото като резултат от разпространяването на манипулирана информация. ✓ Възникване на опасност от създаване на фалшива или едностранчива информация (дезинформация). ✓ Необходимост от създаване на условия за ползване на информацията от различни източници. ✓ Необходимост от създаване на правна рамка за коректното използване на информацията от различни източници при спазване на правилата за конфиденциалност.

	<p>позитивните процеси в развитието на обществото.</p>	<ul style="list-style-type: none"> ✓ Необходимост от създаване на приемственост между експертите и предаване на натрупания опит и рутина при използване на информацията от различни източници. ✓ Създаване на система за обучение на експерти за работа с Big Data.
<p>ВЪНШНИ ФАКТОРИ</p>	<p style="text-align: center;"><u>ВЪЗМОЖНОСТИ</u></p> <ul style="list-style-type: none"> ✓ Подобряване на диалога с потребителите. ✓ Води до осъвременяване на техническата инфраструктура. ✓ Създава условия за разработване на софтуер с възможност за повторно използване при други задачи с малка промяна в него. ✓ Бърз достъп на експертите до външни източници на информация по електронен път. ✓ Създава условия за развитие на креативни подходи за анализ на явленията и процесите с участието на експерти от най-различни области на теорията и практиката. 	<p style="text-align: center;"><u>ЗАПЛАХИ</u></p> <ul style="list-style-type: none"> ✓ Затруднения при намиране на достатъчно квалифицирани експерти. ✓ Зависимост на официалната статистика от външни софтуерни фирми изпълнители. ✓ Отсъствие на ясно дефинирани стратегия и тактика за използване на съществуващия капацитет. ✓ Стихийно усвояване на ИКТ без определен план и очертаване на приоритети. ✓ Липса на взаимно разбиране между статистици и ИТ експерти по отношение на феномена Big Data. ✓ Отсъствие на изградена система за усвояване на добрите практики. ✓ Отсъствие на ясни правила за управление на качеството при използването на големи масиви от данни.

В книгата си „Архиви на бъдещето: история на следващите 50 години“ Ричард Уотсън нахвърля бъдещите щрихи за развитие на света. Той отбелязва, че „...интелигентността ще започне да нараства безкрайно и експоненциално, щом достигнем определено ниво в развитието на генетиката, нанотехнологията и роботиката и до сливането на тези технологии с човешката

биология“ (с. 81). При тези условия следва „сливане на компютрите с комуникациите“, „...инженерството започва да се слива с компютрите“, „в автомобилите инженерството пък се слива с науки като компютърната, а тя, от своя страна, силно се влияе от биологията и неврологичните науки“. Границите между отделните науки се размиват в определени аспекти. В резултат на това „се сливат и продуктите“. Полето за производство на Big Data се разраства в невиджани мащаби. В Кеймбридж учен е създал прототип на компютър, който може да „чете“ мислите на потребителите си, като „заснема и интерпретира изразите на лицето като например при концентрация, болка или объркване“. Компанията „Тойота“ вече работи по проект, при който „емоционалното състояние на шофьорите се обвързва с механизмите за сигурност“. Сред потенциалните клиенти са и застрахователни компании, които искат да „изловят недобросъвестните искиове, банки, които са се прицелили към измами със самоличността, институции, искащи да разкрият измами със социалните осигуровки“ (с. 86) и т.н.

Друга интересна сфера за приложение на Big Data са прогнозите в почти реално време. „В бъдеще прогнозите за трафика ще са точно толкова обичайни колкото прогнозите за времето. Ще има прогнози за замърсяването, за болестите и за войните (с. 87).“

По принцип трябва да се отбележи, че официалната статистика в известна степен е ограничена в сравнение с Big Data, тъй като се нуждае от точно дефинирана крайна съвкупност от обекти, явления и процеси, които в реалния живот невинаги съществуват. В крайна сметка обаче съществува необходимост от дефинирането на такова отграничение във времето и пространството, за да се идентифицира моментът, за който резултатите и анализът на данните са валидни и значими. В този смисъл времето ще бъде основната гаранция за представителността и точността на оценките при равни други условия. По-дългите интервали от време ще бъдат за предпочитане, тъй като се предполага, че включват повече единици за наблюдение и анализ. Под „равни други условия“ се имат предвид два основни фактора: необходимите финансови средства за получаване на информацията и отрязъкът от време, който най-добре характеризира изследваните обекти, явления и процеси. Последният фактор е в пряка зависимост от опита, знанията и интуицията на изследователите. Може да се допълни, че сега съществуващи явления и процеси, които са обект на наблюдение и изучаване, преминавайки през фокуса на Big Data, получават нови възможности за изследване в по-голяма дълбочина, нови нюанси и неподозирани страни от гледна точка на тяхното съдържание. Това са онлайн търговският и рекламният сектор, здравеопазването, обществените услуги, транспортът и други. Целите са известни: повишаване на тяхната ефективност и качество в

условията на ограничени икономически ресурси. Вече са дефинирани и конкретните политики в това отношение, засягащи собствеността на данните, достъпа до данни, заетостта и уменията, инфраструктурата, управлението и т.н.

Това обстоятелство се дължи на динамичното развитие на модерните технологични чудеса, при които много често хардуерът изпреварва софтуера. Ситуация, изискваща разработването на механизми за контрол, който гарантира включването на цялата информационна енергия в служба на човешките закони и морал. Един от основателите на фейсбук (Крис Хюз) отбелязва: „Хората създават технологиите, хората са длъжни да контролират технологиите“. Очевидно алтернативата да бъдем безпомощни пред главоломно разрастващите се технологии се изключва.

Заклучение

Случващото се в света може да се илюстрира добре със следния цитат от статия на бившия председател на Ситибанк - Уолтър Ристън, публикувана в списание „Форин Афеърс“ през 1997 г.: „Преследването на богатство сега е в основата си преследване на информация и нейното прилагане към средствата за производство. Правилата, навиците, уменията и талантите, необходими да се открие, улови, произведе, запази и експлоатира информация в наши дни, са най-важните активи на човечеството. Конкуренцията за най-добра информация замени конкуренцията за най-добра ферма или възгледобивно предприятие... През следващите няколко десетилетия привличането и управлението на интелектуален капитал ще определя кои институции и нации ще оцелеят и ще просперират“ (Фрийдман, 2001, с. 267).

През XXI век модерното управление означава включване на националната икономика в глобалното развитие на света. Във фокуса на статистическите изследвания застават динамични явления и процеси, тотално променящи световния ред. Те могат да бъдат групирани в две основни направления:

- Явления и процеси, свързани с конфликти, застрашаващи установените държавни закони и обществения морал.
- Явления и процеси, редуциращи негативните аспекти на общественото развитие и утвърждаващи статута както на правителствените институции, така и на частния бизнес и структурите на гражданското общество.

Първото направление включва глобални процеси като престъпност, обедняване, безработица, инфлация, замърсяване на околната среда, борба за енергийни ресурси, демографски срив и други. Второто направление акцентира върху процеси, свързани с движението и концентрацията на

огромни потоци от капитали, трудови ресурси, суровини и стоки, необходими за икономическото развитие на всяка страна. И двете направления се подчиняват на един и същ принцип - бързината, с която както управляващите институции, така и бизнесът трябва да разполагат с необходимата информация. Информация в момента се създава непрекъснато, стихийно и в неструктурирана форма. Превръщането ѝ в необходима, полезна, гравивна и позитивна енергия и форма е задачата на статистиците и експертите по информационни технологии през XXI век. В глобалния свят се налага ново правило, сполучливо формулирано от Клаус Шваб на Световния икономически форум в Давос: „Ние се придвижихме от света, в който големите изяждат малките, в свят, в който бързите изяждат бавните“ (Фрийдман, 2001, с. 262).

В заключение може да се отбележи, че Big Data пространството се променя непрекъснато като обем, мащаб и разнообразие. Това поражда проблеми, които очакват своето решение през XXI век.

ЦИТИРАНА ЛИТЕРАТУРА:

Богданов, Б. (2003). Статистиката в условията на глобалната икономика, сп. Статистика, бр. 4, София.

Какво значение „больших данных“ для официальной статистики? Записка секретариата. Европейская экономическая комиссия. Конференция европейских статистиков. Шестидесят первая пленарная сессия. Женева, 10 - 12 июня 2013 года.

Тофльр, А. (1996). Трусове във властта, Народна култура, София.

Уотсън, Р. (2012). Архиви на бъдещето: история на следващите 50 години, „Локус Пъблишинг ЕООД“.

Фрийдман, Т. (2001). Лексус и маслиновото дърво, изд. „Дамян Янков“, София.

Big data - an opportunity or a threat to official statistics? Conference of European Statisticians, Paris, 9 - 11 April 2014.

Big data - an opportunity or a threat to official statistics? Prepared by Eurostat, Economic Commission for Europe, Conference of European Statisticians, Paris, 9 - 11 April 2014.

Demystifying Big Data: A Practical Guide To Transforming The Business of Government (2012).

Exploring data-driven innovation as a new source of growth. Mapping the Policy Issues Raised by „Big Data“. OECD, 18 April 2013.

Will ‘big data’ transform official statistics? Denisa Florescu, Martin Karlberg, Fernando Reis, Pilar Rey Del Castillo, Michail Skaliotis and Albrecht Wirthmann, All authors are affiliated with Eurostat, L-2920 Luxembourg.

Couper, Mick P. (2013). Is the sky falling? New technology, Changing Media end the Future of Surveys, University of Michigan Survey Research Center, vol. 7, No. 3, pp. 145 - 156.

Crawford, Kate (2011). Six provocations for Big Data, Danah Boyd - Microsoft research, University of New South Wales, Paper to be presented at Oxford Internet Institute’s „A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society” on September 21.

Ricciato Fabio, Peter Widhalm, Massimo Craglia and Francesco Pantisano (2015). „Estimating population density distribution from network-based mobile phone data“, European Commission, Joint Research Centre Institute for Environment and Sustainability.

Struijs Peter, Barteld Braaksma and Piet JH Daas. Official statistics and Big Data, April - June 2014: 1 - 6, Big Data and Society.