

УДК: 338.1

JEL Classification: C15

О. В. ЖУРАВЛЬОВ,
доцент кафедри менеджменту
зовнішньоекономічної діяльності,
Національна академія статистики, обліку та аудиту

Імплементация концепції “великих даних” у державну статистику

Стаття присвячена дослідженню розвитку концепції “великих даних” в світі та її впливу на трансформацію статистичного моделювання економічних процесів. Наголошено на важливості імплементції “великих даних” у систему державної статистики України, що дасть можливість корегування економічної політики у відповідності зі змінами інституційного середовища.

Ключові слова: “великі дані”, державна статистика, інституційне середовище, відкриті дані, Інтернет, інформація.

Постановка проблеми. Імплементация концепції “великих даних” (або мегаданих) у державній статистиці може бути основою ретельної діагностики трансформацій, які відбуваються в економіці України під впливом глобалізації. Саме такі дані надають можливість сформулювати інституційне середовище, сприятливе для економічних змін і подальшого розвитку. Вони визначають напрямок коригувань інституційного середовища при розробленні соціальної, економічної, демографічної політики та інших важливих компонентів господарського розвитку. Крім того, використання мегаданих підкреслює значущість нових технологій, які дозволяють перевести статистику на кардинально новий рівень у відповідності з поточним технологічним укладом.

Разом із достатньою кількістю прикладів успішного використання “великих даних” існує проблема розроблення механізму імплементції їх у систему державної статистики, а також охоплення “великими даними” такої важливої сфери як інституційне середовище. Тому впровадження нових індикаторів за допомогою “великих даних” є важливою науковою проблематикою.

Аналіз останніх досліджень та публікацій. На думку О. О. Васечко, “використання великих даних у статистичних цілях є одночасно і дуже перспективним, і революційним з погляду на організацію виробничого процесу та методологію). Дуже привабливим моментом є те, що, на відміну від традиційних статистичних даних, мегадані можуть бути швидким барометром суспільних настроїв та благополуччя. Але водночас вони мають суттєвий недолік – складність управління інформаційним потоком, оскільки дані є неструктурованими або погано структурованими (не мають заздалегідь визначеної реляційної моделі опису) та надходять в онлайн режимі (потік даних)” [1, с. 5].

О. Л. Єршова та Т. В. Томашевська підкреслюють, що “більшою мірою ефект Big Data залежатиме від ефективності управління і використання ресурсів. Країни з великими підприємствами, глобальними зв’язками, з розвинутою інфраструктурою інформаційно-комунікаційних технологій зможуть отримати значно більшу вигоду, ніж ті, що в цьому відстають. Таким чином, наслідки від використання великих даних будуть більш відчутними в Північній Європі, водночас у більшості країн нової Європи і Південної Європи результат буде значно меншим. В Україні застосування великих даних лише на слуху у вузьких професійних колах ІТ-індустрії” [2, с. 34].

Ми погоджуємось із результатами наукового пошуку згаданих авторів та продовжуємо досліджувати шляхи імплементції концепції “великих даних” у державну статистику. Крім того, враховуючи все більший інтерес економістів до інститутів як детермінант економічного розвитку та важливість “великих даних” у розумінні формальних і неформальних інститутів, нами запропоновано таку категорію як “інституційна статистика” – галузь офіційної статистики, яка досліджує “правила гри”.

Метою дослідження є розроблення шляхів імплементції “великих даних” у державну статистику.

Викладення основного матеріалу дослідження. Проблема імплементації “великих даних” до системи державної статистики обговорюється на глобальному рівні. Найбільш значущі конференції, проведені останнім часом, відображають визнання необхідності вироблення спільних підходів до розгляду проблем “великих даних”:

– I Міжнародна конференція з використання “великих даних” державною статистикою, м. Пекін (КНР), 28–30 жовтня 2014 р.;

– II Глобальна міжнародна конференція з використання “великих даних” державною статистикою, м. Абу-Дабі (ОАЕ), 20–22 жовтня 2015 р.;

– III Глобальна міжнародна конференція з використання “великих даних” у державній статистиці, м. Дублін (Ірландія), з 30 серпня по 1 вересня 2016 р.

Стратегічну цінність “великих даних” для державної статистики визначають за такими критеріями:

– великий обсяг даних може сприяти отриманню більш детальних відомостей за конкретними напрямками статистики;

– висока швидкість зміни інформації сприяє збільшенню частоти статистичних оцінок;

– різноманітність “великих даних” може відкрити можливості для виробництва статистики в нових галузях (наприклад, дистанційне зондування та географічні інформаційні системи), отримання інформації про не спостерігається офіційною статистикою сферах;

– різноманітність походження “великих даних” може сприяти багатоаспектності вимірювання, а відтак підвищенню надійності статистичних оцінок;

– “великі дані” є важливою складовою формування інституційної статистики.

О. Васечко підкреслює, що “сучасними джерелами мегаданих є: адміністративні дані (електронні медичні карти, карти соціального страхування); дані фінансового сектору (банківські, страхові); торговельні та операційні дані (операції з банківськими картками між двома суб'єктами); реєстраційні дані датчиків (супутникові зображення, дорожні, кліматичні); дані відстеження (з мобільних телефонів, глобальної навігаційної системи – GPS); дані поведінки (запити в онлайн-пошукових системах з метою отримання інформації щодо продуктів, послуг); оціночні дані (відгуки та коментарі щодо послуг на сайтах компаній та у соціальних мережах)” [1, с. 7].

В умовах проникнення Інтернету в життя людей формування “великих даних” є необхідною вимогою часу. На рис. 1 представлено динаміку зростання користувачів Інтернету в Україні і Нідерландах. Нами наведено приклад використання даних про користувачів і про очікувану тривалість життя. Саме поєднання таких нетрадиційних чинників формує уяву про інституційне середовище – наскільки доступним є Інтернет, чи використовують користувачі інформацію про способи зміцнення здоров'я та профілактики захворювань тощо.

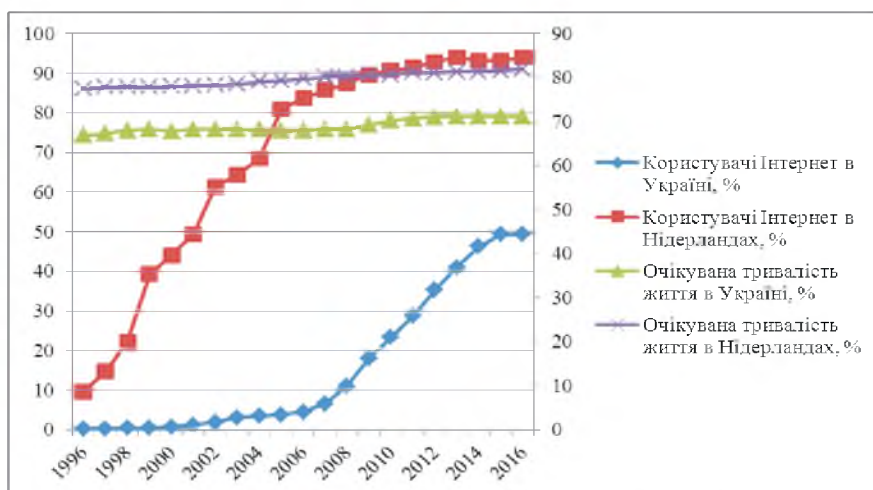


Рис. 1. Динаміка користувачів Інтернету і тривалості життя в Україні і Нідерландах

Джерело: [5]

У вересні 2013 року 49,8% дорослого населення України користувалися Інтернетом. Таким чином, зростання числа користувачів триває навіть більшими темпами, ніж прогнозувалося. Темп приросту протягом лютого 2012 – жовтня 2013 років склав 16%, що трохи поступається рекордному стрибку в 34% у період з березня 2011 по лютий 2012 року [3].

Разом із тим варто зауважити, що в Нідерландах частка Інтернет-користувачів становить майже 95% населення [3].

Саме це ставить задачу виявлення залежності між очікуваною тривалістю життя та доступу до Інтернету, адже глобальна мережа – це екзабайти інформації, які подвоюються кожні 2–3 роки. Саме доступ до цих знань може змінювати життя людини як з позиції підтримання стану здоров'я, так і з позиції міграції, доступу до вакансій, консультацій лікарів онлайн тощо.

Особливе місце у “великих даних” займає концепція відкритих даних, яка відображає ідею про те, що певні дані повинні бути вільно доступні для машинного використання і подальшого передруку без обмежень авторського права, патентів та інших механізмів контролю (рис. 2).



Рис. 2. Великі дані та офіційна статистика

Джерело: розробка автора

Звільнити дані від обмежень авторського права можна за допомогою вільних ліцензій, таких як ліцензій Creative Commons. Якщо який-небудь набір даних не є суцільним надбанням, або не пов'язаний ліцензією, що дає права на вільне повторне використання, такий набір даних не вважається відкритим, навіть якщо він викладений в машинному вигляді в Інтернеті.

Важливе місце в аналізі даних займає теоретична дефініція “тріади даних”. В англійській мові існує тріада слів – data, analytics & insights (дані, аналітика та розуміння). Оскільки з першими двома термінами при перекладі на українську не виникає питань, то для терміна “insights” слід надати додаткове визначення.

Insights – це “додана вартість” у даних, отримана за допомогою аналітики. Проникність, отримана за допомогою аналітики, є неймовірно потужною і може використовуватися для різноманітних цілей – від вивчення перспектив для бізнесу до фундаментальних досліджень в багатьох галузях знань. Зв'язок між даними інституційної статистики та економічної політики представлено на рис. 3.

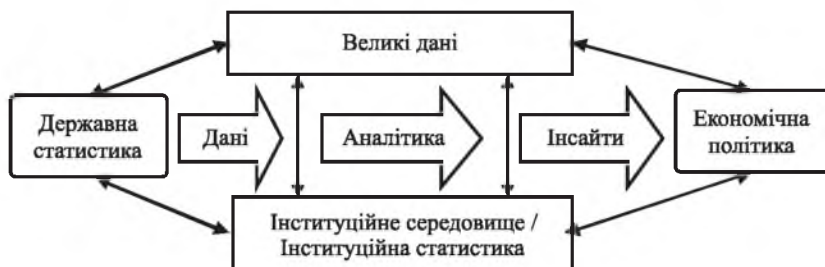


Рис. 3. Блок-схема взаємодії компонентів державної статистики, інституційного середовища і економічної політики

Джерело: розробка автора

Робота з “великими даними” – це, в першу чергу, дієвий інструмент ефективного перерозподілу ресурсів. Той інструмент, що дозволить отримати швидкі та ефективні “перемоги” тут і зараз. Це та аналітика, що сприятиме абсолютно іншій культурі полікотворення – коли політика формується на основі фактів, а не теоретичних міркувань. Доступ до відкритих даних та широке використання аналізу даних покращить процес прийняття рішень в публічному секторі, в тому числі в питаннях законодавства та державного управління на національному та місцевому рівнях.

Застосування “великих даних” у державній статистиці пов’язано з вирішенням проблем забезпечення конфіденційності, фінансування відповідних робіт, а також наявності фахівців для ведення великих даних і аналітики.

Реальна цінність “великих даних” та їх аналітики полягає в їх здатності надавати цікаві ідеї. Можна мати багато точок даних, але працювати з ними таким чином, щоб ці дані могли надавати цінну інформацію.

Прикладом використання концепції відкритих даних в Україні є портал відкритих даних, який затверджено Кабінетом Міністрів України (рис. 4).



Рис. 4. Портал відкритих даних України <http://data.gov.ua/>

Кабінет Міністрів України ухвалив постанову від 30 листопада 2016 р. № 867 “Деякі питання оприлюднення публічної інформації у формі відкритих даних”, якою затверджено Порядок ведення Єдиного державного веб-порталу відкритих даних.

Цілями створення і використання порталу відкритих даних є:

- надання на єдиному інформаційному ресурсі найбільш повних переліків і відомостей про набори відкритих даних, що публікуються органами державної влади та органами місцевого самоврядування;

– створення інформаційного майданчика для взаємодії з широкою громадськістю з питань формування, публікації та використання відкритих державних даних, включаючи тематичне інформування та отримання зворотного зв'язку щодо їх якості та затребуваності;

– формування і реалізація єдиної технологічної політики в сфері відкритих державних даних.

Порядок ведення Єдиного державного веб-порталу відкритих даних визначає основні поняття та механізм ведення Єдиного державного веб-порталу відкритих даних з метою забезпечення доступу до публічної інформації у формі відкритих даних та взаємодії з користувачами щодо відкритих даних. Він також забезпечує нормативно-правове врегулювання питань функціональних можливостей, обов'язки держателя та адміністратора порталу, процедуру реєстрації розпорядника інформації, розміщення прикладної програми тощо.

“Великі дані” характеризуються такими ознаками як великий обсяг інформації, висока швидкість зміни інформації, різноманітність і різнорідність даних.

У сучасному світі відкриті дані є не лише інструментом підвищення прозорості та ефективності роботи органів державної влади, а й потужним джерелом для стимулювання розвитку економіки країни. На жаль, більшість українських урядових топ-менеджерів на сьогодні ще не повністю уявляють потенціал та можливості відкритих даних саме в економічному аспекті. Експерти Програми EGAP підготували черговий матеріал щодо впливу відкритих даних на економіку ЄС. Сподіваємось, що цей матеріал стане у нагоді для посилення політичної волі та громадської активності в Україні для подальшого розвитку даних.

У 2015 р. в ЄС було проведено одне з найбільш потужних досліджень оцінки впливу та потенціалу відкритих даних щодо розвитку економіки країн та Єврозони в цілому. Як результат було розраховано основні показники зростання ринку відкритих даних. Зокрема, визначено розмір прямого (direct market size) та непрямого ринків відкритих даних (indirect market size), які разом формують загальний ринок відкритих даних (total market). Прогнозований розмір прямого ринку відкритих даних країн Єврозони на 2016 р. складе 55,3 млрд євро. У період з 2016 по 2020 рік очікується зростання цього ринку на 36,9% до 75,7 млрд євро. Загальна ринкова вартість відкритих даних оцінюється в 193–209 млрд євро для 2016 р. та 265–286 млрд євро до 2020 р., з поправками на інфляцію [2].

Вплив використання відкритих даних у розрізі секторів економіки виглядає таким чином. За оцінками експертів, найбільшу вигоду від розвитку відкритих даних (22 млрд євро у 2020 р.) отримує саме державний сектор. Це підтверджує, що державний сектор є основним користувачем своїх власних даних. Для таких секторів як сільське господарство, мистецтво і розваги очікувані переваги є значно меншими і становлять 379 млн євро для кожного. Відкриті дані у цих секторах мають значний потенціал, проте знадобиться час, щоб його розкрити.

Що стосується ринку праці, то визначається кількість робочих місць, які будуть безпосередньо пов'язані з повторним використанням відкритих даних у приватному секторі країн ЄС. Через це не враховувались співробітники, що працюють у статистичних та інших урядових організаціях. Згідно з прогнозами, максимальна кількість робочих місць, безпосередньо пов'язаних з відкритими даними, у 2016 р. становитиме 75000. До 2020 р. цей показник має зрости до 100000, тобто на 32% (7,3% на рік) [4].

Органи державного управління прагнуть точніше оцінити економію витрат, яку вони можуть отримати від відкритих даних. Економію можна розрахувати на підставі прогнозованого ВВП країн Єврозони до 2020 р. Сукупна економія для країн ЄС у 2020 прогнозується на рівні 1,7 млрд євро. Ефективність є важливим фактором як для державного, так і для приватного сектору у всіх галузях. Підвищення ефективності спрямовано на поліпшення розподілу ресурсів у такий спосіб, щоб мінімізувати видатки та підвищити результати, використовуючи при цьому одну і ту ж саму кількість ресурсів. Для розрахунку ефективності використовувались три ключові індикатори: збереження життя, збереження часу та користь для навколишнього середовища від використання відкритих даних [5].

“Великі дані” не лише здійснюють революцію в маркетингу і продажах, вони мають значний потенціал використання і в публічному секторі: від містопланування до координації під час надзвичайних ситуацій, від забезпечення логістики ліків у закладах охорони здоров'я до розслідувальної журналістики та прогнозів результатів виборів. Якщо ми маємо до них доступ, можемо користуватись та ділитись ними – це вже відкриті дані.

“Великі дані” можуть використовуватись не лише в системі державної статистики України, а й вітчизняними підприємствами.

Прикладом ефективного використання “великих даних” у бізнесі є компанія Dunkin Donuts. Разом із касовими чеками компанія отримує дані про найбільш популярні пропозиції у покупців, які потім обробляються. Такий підхід дозволив компанії збільшити прибуток і оборотність товарів на складі. Поштовх для імплементації “великих даних” надало намагання керівництва компанії побачити, чим задоволені чи незадоволені клієнти. Виявивши, що найбільше очікування клієнтів – це швидкість обслуговування, компанія вирішила замінити стандартну систему лояльності (купони чи карти постійних клієнтів) на новий додаток для смартфона.

Через можливість аналізувати таку інформацію компанія може вирішувати, чи достатньо працівників на місцях, чи вистачає продукції в часи пік, робити акційні пропозиції більш індивідуальними тощо. Таким чином, використовуючи “великі дані”, компанія може прогнозувати поведінку споживачів і розробляти стратегію розвитку на основі їх аналізу.

Можливості використання “великих даних” в різних секторах економіки представлено на рис. 5.



Рис. 5. Можливості використання “великих даних” у різних секторах економіки

Джерело: розроблено і побудовано автором

В енергетиці використання “великих даних” забезпечує підвищення точності розподілу потужностей генераторів. У гірничодобувній і нафтової промисловості – ефективність розроблення родовища, відстеження графіку канітального ремонту та стану обладнання, прогнозування попиту на продукцію і ціни.

Для правоохоронних органів “великими даними” є інформація системи ALPR (автоматичне розпізнавання номерних знаків), карти злочинності тощо.

В системі охорони здоров'я “великими даними” є інформація про:

- схильність до захворювань за аналізом крові пацієнтів і донорів;
- зниження витрат держави і підвищення ефективності лікування;
- боротьбу із захворюваннями і моніторинг зростання епідемії;
- потенційно можливі епідемії.

На думку фахівців компанії Делойт [6], “великі дані” – це “паливо” для сучасного бізнесу, і людство створює дані з вражаючою швидкістю. Кожні 2 дні у світі виробляється такий обсяг нових даних, який було вироблено з початку цивілізації до 2003 року.

Цілком зрозуміло, що організації прагнуть розкрити потенціал “великих даних”. Результатом аналізу «великих даних» є прискорення темпів інновацій і руйнування традиційних бізнес-моделей.

Головні проблеми зберігання великих обсягів даних:

- безпека обчислення в розподілених програмних системах;
- безпека нереляційних баз даних;
- безпечне зберігання даних;
- перевірка достовірності;
- моніторинг безпеки в режимі реального часу;
- шифрування управління доступом і забезпечення безпеки з'єднань;
- частковий контроль доступу;
- походження даних.

В Україні концепція “великих даних” може впроваджуватися через:

- використання сенсорів GPS для дослідження потоків транспорту;
- використання геолокації мобільних телефонів;
- використання відкритих даних соціальних мереж;

Однією з проблем для вчених-статистиків є методологічне забезпечення “великих даних” (рис. 6).

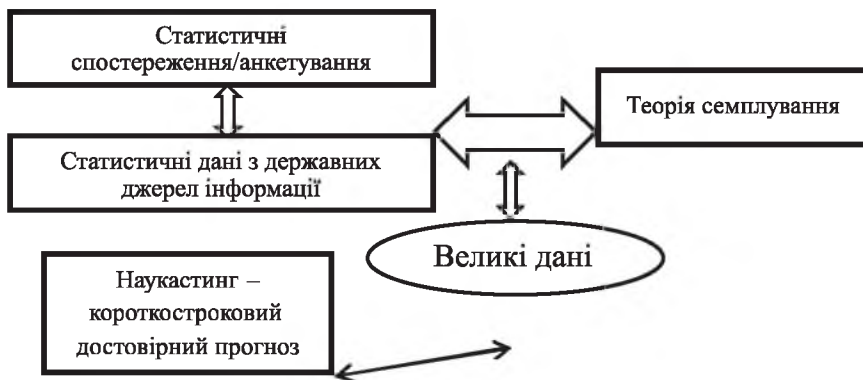


Рис. 6. Методологічне забезпечення “великих даних” у державній статистиці

Джерело: [7]

Проте збирання інформації для формування “великих даних” та включення їх у процес статистичного виробництва не є легким завданням. Наприклад, дані про трафік складаються з вимірів інтенсивності трафіку. Кожен цикл підраховує кількість автомобілів на хвилину, які проходять у певному місці, а також вимірює швидкість і довжину. Такі дані становлять інтерес для руху і транспортної статистики, а також для інших економічних явищ, пов’язаних із транспортом.

Для України такий підхід є критично важливим – транспортна стратегія Києва та найбільших міст країни має розроблятися саме на підставі даних про транспортні потоки. При цьому дані з камер відстеження інтенсивності руху мають передаватися до окремої структури в Державній службі статистики – національній бази даних щодо транспорту.

Висновки і подальші дослідження. Проаналізовано концепції “великих даних” і відкритих даних, показано проблеми імплементації “великих даних” у державній статистиці. Представлено шляхи імплементації “великих даних” у державну статистику України через активне використання технологічних можливостей мобільних операторів, навігаційних систем, камер спостереження, соціальних мереж тощо. Показано можливості використання “великих даних” в різних секторах економіки, на також на рівні компаній. Наголошено на проблемах зберігання великих обсягів даних.

Проведене дослідження доводить, що “великі дані” є величезним ресурсом, який слід використовувати в Україні у всіх секторах економіки.

Список використаних джерел

1. Васечко О. О. Сучасні виклики статистичної вищої освіти і науки // Статистика України. 2014. № 4(67). с. 4–10.
2. Єршова О. Л. Томашевська Т. В. Великі дані (Big Data) і модернізація національних систем офіційної статистики // Прикладна статистика: проблеми теорії та практики. 2016. № 2. с. 32–38.
3. World Economic Forum [Electronic resource] URL: <https://www.weforum.org/agenda/2016/05/are-economic-statistics-in-danger-of-becoming-irrelevant>
4. Statistical Database. Eurostat. European Commission [Electronic resource]. URL: <http://ec.europa.eu/eurostat/data/database>.
5. Statistical Year book [Electronic resource]. URL: <http://unstats.un.org>.
6. Deloitte insights [Electronic resource]. URL: <https://www2.deloitte.com/ca/en/pages/deloitte-analytics/articles/data-analytics-and-privacy.html>
7. Lampos Vasileios, Cristianini Nello (2012). Nowcasting Events from the Social Web with Statistical Learning (PDF). ACM TIST. 3 (4, article number 72). Retrieved 12 June 2015.

А. В. ЖУРАВЛІВ,
доцент кафедри менеджмента
внешнеэкономической деятельности,
Национальная академия статистики, учета и аудита

Имплементация концепции «больших данных» в государственную статистику

Статья посвящена исследованию развития концепции «больших данных» в мире и ее влияния на трансформацию статистического моделирования экономических процессов. Подчеркнута важность имплементации «больших данных» в систему государственной статистики, что даст возможность корректирования экономической политики в соответствии с изменениями институциональной среды.

Ключевые слова: «большие данные», государственная статистика, институциональная среда, открытые данные, Интернет, информация.

O. V. ZHURAVLYOV,
Associate Professor of Department
for Management of Foreign Economic Activities,
National Academy of Statistics, Accounting and Audit

Implementing the “Big Data” Concept in Official Statistics

Big data is a huge resource that needs to be used at all levels of economic planning. The article is devoted to the study of the development of the concept of “Big Data” in the world and its impact on the transformation of statistical simulation of economic processes.

Statistics at the current stage should take into account the complex system of international economic relations, which functions in the conditions of globalization and brings new forms of economic development in small open economies. Statistical science should take into account such phenomena as gig-economy, common economy, institutional factors, etc.

The concept of “Big Data” and open data are analyzed, problems of implementation of “Big Data” in the official statistics are shown. The ways of implementation of “Big Data” in the official statistics of Ukraine through active use of technological opportunities of mobile operators, navigation systems, surveillance cameras, social networks, etc. are presented. The possibilities of using “Big Data” in different sectors of the economy, also on the level of companies are shown. The problems of storage of large volumes of data are highlighted.

The study shows that “Big Data” is a huge resource that should be used across the Ukrainian economy.

Keywords: “Big Data”, official statistics, institutional environment, open data, Internet, information.

Посилання на статтю:

Журавльов О. В. Імплементация концепції “великих даних” у державну статистику // Науковий вісник Національної академії статистики, обліку та аудиту: зб. наук. пр. 2017. №3. С.7-15.