

– гуманітарно-релятивна концепція, в якій у центрі уваги була мотивація та задоволення співробітників.

Перерозподілом обов'язків між органами лінійно-функціональної структури намагаються впорядкувати та скоротити довжину горизонтальних зв'язків у процесі управління, мінімізувати ті негативні наслідки багаторівневої лінійної системи, що характерні класичним схемам побудови організаційної структури. Стало очевидним, що сучасні вимоги зовнішнього середовища не сумісні з традиційними концепціями організації. Новітні організаційні концепції направлені на розв'язання таких основних завдань: як підприємству реагувати на потреби суспільства та індивідуума; як установам поступати з задачами, які суперечать традиційним організаційним формам; яким чином розв'язати задачу прискорення темпів, зниження затрат, покращення якості, більшій орієнтації на клієнта. Серед останніх розробок можна виділити: *організаційна модель, що орієнтована на процес*; *організаційна модель, що орієнтована на співробітників*; *організаційна модель, що орієнтована на завдання*.

Розглядаючи як традиційні, так і новітні підходи до побудови організаційних структур установи можна вказати ті необхідні передумови, що забезпечують на своєму рівні управління можливість своєчасного прийняття вірних рішень: контроль за виконанням завдань; екстраполяція минулого в майбутнє; передбачення змін, які в сучасних умовах можуть виникати раптово.

На сучасному етапі політичного та економічного розвитку України в управлінські структури прийшло молоде покоління управлінців, яке сформувалось в умовах відкритості суспільства, демократизації і гласності та працює поруч з людьми старшого віку, що мають величезний професійний досвід.

С. І. Джус,

аспірант кафедри статистики,

Національна академія статистики, обліку та аудиту, м. Київ

ПОТРЕБА ВИКОРИСТАННЯ DATA SCIENCE & BIG DATA ANALYSIS (НАУКА ПРО ДАНІ ТА АНАЛІЗ ВЕЛИКИХ ДАНИХ) У СУЧАСНОМУ СТАТИСТИЧНОМУ ТА ФІНАНСОВОМУ СВІТІ

Актуальність теми продиктована сьогоденними викликами в обробці великих масивів даних (англ. – Big Data) та створенні нового статистичного фаху, названого «вчений з даних» (англ. – Data Scientist), який станом на початок 2017 року займає чільне місце в обробці й аналізі всієї економічної та соціальної інформації у світі.

У найсучасніших дослідженнях та публікаціях у сфері великих даних можна виділити декілька авторів. Найвідомішими є іноземні статистики, програмісти, фінансові та бухгалтерські фахівці: Стів Ландефелд (старший радник Відділу статистики ООН), Карлота Перес, Файе Чуа, Джефрі Дін і Санджей Гемават, Сіу-Мінг Там і Фредерік Кларк, Бертольд Брааксма, Кіз Зеленберг, Мікаела Агафі, Фабріца Граса, Уім Клоек, Фернандо Реіс і Соріна Вайю, Амір Гандомі, Муртаза Хейдер [1] та ін. Із вітчизняних – С. Шельпук, Д. Карамшук та С. Омельчук.

Також варто виділити офіційні публікації закордонних статистичних інституцій, на результати яких, у сфері Big Data, Україна повинна опиратися, а іноді і перевершувати їх. До них відносяться: Відділ статистики ООН, Європейська економічна комісія (ЄЕК), Бюро Конференції європейських статистиків, Статистичне бюро Європейської комісії (Євростат), Організація економічного співробітництва та розвитку (ОЕСР) та Економічна і соціальна комісія для Азії та Тихого океану (ЕСКАТО), Комісія з великих даних при Фонді «ТехАмерика», Національне статистичне бюро Китаю [2–4] Міждержавний статистичний комітет Співдружності Незалежних Держав та Австралійське бюро статистики [5; 6], Федеральна служба державної статистики РФ [7], Американська асоціація дослідників громадської думки [8], Міжнародний союз електров'язку (Женева, Швейцарія) [9], Національний інститут статистики Італії [10], Статистичне бюро Нідерландів [11] та ін.

Програмні комплекси, що використовуються при обробці та аналізі великих даних, в свою чергу, займають фундаментальне місце, оскільки за допомогою саме них і відбувається обробка даних. Статистичні програми, додатки, надбудови та мови програмування, такі як STAT Plus, STATA, STATISTICA, SPSS, R-programming, Apache Hadoop, Tableau і Quikview, мови програмування Python, SQL/NoSQL, MapReduce та ін., використовуються провідними науковцями та практиками в сфері статистики та математичного аналізу. MS Excel та Access теж представляють собою певну структуру обробки та візуалізації даних (розрахунок формул, імпорт/експорт, зіставлення даних, створення діаграм та графіків, фільтрування і т. д.). Проте, основним результатом обробки даних, у сучасному розумінні, є їх візуалізація та трансформація у зручний формат для підведення підсумків, розробку висновків та прийняття відповідних рішень.

Сучасна статистика – це вже не тільки збирання й обробка вхідних даних, аналіз та поширення вихідних даних державних статистичних спостережень (далі – ДСС), а й всі ці процеси разом із *візуалізацією* результатів на приватному (бізнесовому) рівні. Тобто органи державної статистики, що виконують функції офіційної статистики на рівні держави, вже зменшили свою частку в усіх вироблених статистичних даних у країні до

50%. Усі чотири процеси, що становлять суть статистичних робіт, виходять на новий рівень у зв'язку зі швидкою глобалізацією та захопленням нових ринків глобальними корпораціями, які вносять значний вплив у розвиток обробки та аналізу даних (напр., Google Inc., Facebook Inc., IBM Inc. та ін.). При цьому, використання «великих даних» розглядається зараз у спрощенні процесу перепису населення багатьох держав [12] та інших фінансових та нефінансових секторах економіки.

Величезна кількість даних, яка постійно збирається, зберігається і передається завдяки технологіям, змінює пріоритети сучасної економіки та науковців і ставить важливі питання до його лідерів. Ринок аналітики «великих даних», як не дивно, швидко зростає і, згідно з прогнозами 2013 року, має сягнути 23 млрд доларів США до 2016 року [13].

У доповіді Асоціації Присяжних Сертифікованих Бухгалтерів [14, с.6] вказано, що рух з відкритим вихідним кодом і індустрія програмного забезпечення розробили рішення, такі як нові моделі програмування і нові модулі інструментів даних. Управління «великими даними» фактично вимагає відповідних працівників. Рух «відкритого вихідного коду» заснований на ідеї, що вихідні коди програми повинні бути спільними. У поєднанні зі збільшенням обчислювальної потужності, ці рішення дозволяють синтезувати величезну кількість інформації з небаченою раніше швидкістю і точністю, але вони є тільки одним з елементів реагування.

Будь-який аналіз впливу «великих даних» необхідно розпочати з визначення того, що власне означає цей термін, який часто використовується, але зазвичай не всі його розуміють. Він стосується насамперед величезної кількості даних, які постійно збираються за допомогою пристроїв і технологій, таких як платіжні картки та картки лояльності клієнтів, Інтернету та соціальних медіа і все частіше, через датчики Wi-Fi та електронні мітки. Велика частина цієї інформації є неструктурованою – тобто це дані, які не відповідають певній, заздалегідь визначеній, послідовності. У свою чергу, *дані (Data)*, згідно з визначенням «Енциклопедії кібернетики» за редакцією академіка В. Глушкова [15, с. 230], – це факти та ідеї, подані у формалізованому вигляді, завдяки чому їх можна передавати чи обробляти за допомогою певного процесу (й відповідних технічних засобів). Тобто «великі дані» – це наступний етап розвитку даних. Забезпечення якості даних є ключовим фактором. Процес рівно настільки якісний, наскільки якісні дані, які в ньому використовуються.

Оскільки автоматична обробка даних є однією з основних прикладних задач кібернетики, то прописування кодів команд із набором даних повинно перетворюватися у візуальне відображення зрозумілого змісту та форми для пересічного користувача статистичної інформації. У цьому полягає складність обробки даних, оскільки статистик стає не тільки front-end-розробником (розробка програмних компонентів для розрахунків та

візуалізації аналізу), а й back-end-розробником (розробка/адміністрування бази даних), оскільки повинен забезпечувати адміністрування бази даних через різні компоненти системи обробки даних (СОД), таких як SQL/NoSQL, Python, MS Access та ін.

«Гартнер» (надалі Gartner), дослідницька і консультативна фірма в США з питань інформаційних технологій, вперше розробила модель для «великих даних» ще в 2001 році. Її модель «3V» [16] охоплювала обсяг, швидкість і різноманітність даних (англ. – volume, velocity, variety). Із часом до них практики додали ще три «V» – veracity, validity, volatility – правдивість, дійсність та волатильність, що задовольняють умови сучасних статистичних досліджень.

Gartner формалізувала своє визначення в 2012 році: «Великі дані – це інформаційні активи великого обсягу, високої швидкості, та/або високої різноманітності, які вимагають нових форм обробки, щоб дозволити здійснювати покращення прийняття рішень, відкриття ідей і оптимізацію процесів» [16]. Використовуючи статистичні методи та інструменти інтелектуального аналізу даних, підвищується ефективність прийняття рішень і передбачення майбутніх моделей поведінки. Підвищений попит на високоякісні аналітичні навички створює важливі можливості для статистиків і фінансових фахівців.

Сучасність диктує те, що статистик повинен бути не тільки підкованим у статистичній методології, формулах та в аналізі результатів отриманих даних, а й у програмуванні та дизайні тієї вихідної інформації, яка отримана шляхом завантаження/імпорту даних та їх обробки. В цьому випадку статистик уже стає «вченим з даних» (Data Scientist). Надалі будемо використовувати цей термін для ідентифікації суб'єкта обробки великих даних. Тому, спираючись на праці К. Перес [17, с. 7], історика технологій та економіки, яка виділяє ряд довгострокових технологічних хвиль (по 50-60 років): Індустріальна революція (з 1771) → Ера парової залізниці (з 1829) → Ера сталі, електрики та машинобудування (з 1875) → Ера нафти, автомобілів та масового виробництва (з 1908) → Ера інформації та телекомунікацій (з 1971) [18, с.9] → «Великі дані»(1988-2030) → «Ера штучного інтелекту та нанотехнологій» (з 2030) →..., питання потреби використання Big Data у статистиці вже є фактом. Останні два пункти є моїм спостереженням та прогнозом у даному ланцюзі подій, але не його закінченням.

Тобто можна стверджувати, що ми, статистики, маємо справу саме з «великими даними», коли ми говоримо про сучасний збір, обробку та аналіз статистичних даних. Навчені визначати критерії інформації та використовувати дані в моделюванні та прогнозуванні, статистики, бухгалтери та фінансові фахівці можуть надати нову і важливу послугу: робити «великі дані» «меншими», здійснювати «дистиляцію» величезних обсягів інформації в дієві ідеї. Оскільки вони відповідальні за цілісність

звітів та рахунків, вони можуть діяти як хранителі масивів фінансових та нефінансових даних і встановлювати якісні та етичні норми для інформації, що використовується в прийнятті стратегічних рішень, і для інформації, яка продається третім сторонам. Ця роль буде ставати все більш важливою, оскільки дедалі більше компаній (бізнесів) шукатимуть способи розробки нових продуктів і послуг на основі даних, які вони виробляють і якими вони володіють, зокрема в контексті зростаючої заклопотаності навколо релевантності і етичного використання даних.

Список використаних джерел

1. Gandomi A. Beyond the hype: Big data concepts, methods, and analytics [Electronic resource] / A. Gandomi, M. Haider // International Journal of Information Management. – 2015. – Vol. 35. – P.137–144. – Mode of access: <http://creativecommons.org/licenses/by-nc-nd/3.0/1-s2.0-S0268401214001066-main.pdf>
2. Большие данные и модернизация статистических систем. [Электронный ресурс] / Экономический и Социальный Совет ООН. – Режим доступа: <http://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-R.pdf>
3. Results of the UNSD/UNECE Survey on organizational context and individual projects of Big Data
4. Prepared by the Statistics Divisions of UN/DESA and UN Economic Commission for Europe [Electronic resource] / Statistical Commission. – Mode of access: <http://unstats.un.org/unsd/statcom/doc15/BG-BigData.pdf>
5. Landefeld S. Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues [Electronic resource] / S. Landefeld. – Mode of access: <http://unstats.un.org/unsd/trade/events/2014/beijing/Steve%20Landefeld%20-%20Uses%20of%20Big%20Data%20for%20official%20statistics.pdf>
6. Там С.-М. Большие Данные, Официальная Статистика и Некоторые Инициативы Австралийского Бюро Статистики [Электронный ресурс] / С.-М. Там, Ф. Кларк. – Режим доступа: http://www.cisstat.com/BigData/CIS-BigData_03.pdf
7. Вега Х.Л. Межгосударственный статистический комитет СНГ. Полезные ссылки [Электронный ресурс] / Х.Л. Вега, А. Аргуэсо, К. Тейхейро. – Режим доступа: http://www.cisstat.com/BigData/bd_links.htm
8. Большие данные и официальная статистика [Электронный ресурс] / Аналитический центр при правительстве Российской Федерации. – Режим доступа: http://www.gks.ru/free_doc/new_site/rosstat/smi/conf16/prez_Pushkin.pdf
9. Отчёт AAPOR о больших данных: 12 февраля 2015 [Электронный ресурс] / Л. Джапек, Ф. Крейтер и др.; Американская ассоциация исследователей общественного мнения. – Режим доступа: http://wciom.ru/fileadmin/file/nauka/grusha2015/AAPOR_big_data.pdf

10. Измерение информационного общества. Отчет 2014 год. Резюме [Электронный ресурс] / Международный союз электросвязи. – Режим доступа: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2014/MIS_2014_Exec-sum-R.pdf
11. Scannapieco M. Placing big data in official statistics: a big challenge? [Electronic resource] / M. Scannapieco, A. Virgillito, D. Zardetto. – Mode of access: <http://bit.sp.susu.ru/id/10530/view>
12. Daas P. Big data (and official statistics) [Electronic resource] / P. Daas, M. van der Loo. – Mode of access: <http://docslide.net/documents/big-data-and-official-statistics-56ef21422318c.html>
13. Существует ли возможность использования «больших данных» в ходе переписи населения 2020 года? [Электронный ресурс]. – Режим доступа: http://www.cisstat.com/BigData/10_Poland_RUS.pdf
14. Worldwide Big Data Technology and Services 2012–2016 Forecast [Electronic resource]. – Mode of access: <http://www.datascienceassn.org/sites/default/files/Worldwide%20Big%20Data%20Technology%20and%20Services%202012%20-%202016%20Forecast.pdf>
15. «Великі дані»: можливості і виклики [Электронный ресурс] / Асоціація присяжних сертифікованих бухгалтерів. – Режим доступа: https://www.nctbpu.org.ua/userfiles/file/analytics/big_data_its_power_and_perils_ua.pdf
16. Енциклопедія Кібернетики / відп. ред. В. М. Глушков. – Т. 1 (А-Л). – К. : Гол. ред. УРЕ, 1973. – 584 с.
17. Beyer M.A. The Importance of ‘Big Data’: A Definition [Electronic resource] / M.A. Beyer, D.Laney. – Mode of access: <http://www.gartner.com/id=2057415>
18. Чарни Г. Экстремальность и экспоненциальность – атрибуты цифровой эпохи [Электронный ресурс] / Г. Чарни. – Режим доступа: http://www.cisco.com/assets/global/RU/events/cisco-connect/presentation/kon2/17/09_55_10_40.pdf
19. Perez C. Technological revolutions and techno-economic paradigms [Electronic resource] / C. Perez. – Mode of access: <http://technologygovernance.eu/files/main/2009070708552121.pdf>