

УДК 316.004

ВЕЛИКІ ДАНІ В КОНТЕКСТІ ДОСЛІДЖЕННЯ ПРОБЛЕМ СУЧАСНОГО СУСПІЛЬСТВА

Кислова Ольга Миколаївна – кандидат соціологічних наук, доцент кафедри методів соціологічних досліджень соціологічного факультету Харківського національного університету імені В. Н. Каразіна, майдан Свободи, 4, Харків, 61122, Україна, olga.kyslova@karazin.ua, ORCID ID <https://orcid.org/0000-0003-2241-1313>

Kyslova Olga – PhD in Sociological Science, Associate Professor, Department of Methods of Sociological Research V. N. Karazin Kharkiv National University, 4, Svobody Sq., Kharkiv, 61122, Ukraine, olga.kyslova@karazin.ua, ORCID ID <https://orcid.org/0000-0003-2241-1313>

Стаття присвячена аналізу феномену великих даних та ролі, яку вони відіграють у дослідженні сучасного суспільства. Схарактеризовано еволюцію трактування великих даних та показано, що початкова технічна конотація великих даних пов'язана з підкресленням надзвичайності їхніх об'ємів, які не дозволяють здійснювати обробку звичними методами та інструментами, зазнала значних трансформацій, що обумовлені включенням людського компонента. Наразі великі дані розуміються як сукупність цифрових слідів, які залишають люди, коли використовують інформаційні технології: здійснюють інтернет-серфінг, завантажують мобільні додатки чи музику, спілкуються з друзями в соціальних мережах, користуються GPS, купують товари в інтернет-магазинах тощо. Великі дані генеруються інтернетом, проте вони містять інформацію не стільки про інтернет, скільки про суспільство та суспільні процеси, які відображаються в інтернеті. Іншими словами, великі дані – даніфікована інформація про все і про всіх. Показано, що великі дані є новим джерелом інформації як про навколишній світ, так й про розвиток суспільних процесів, що робить їх цінною базою емпіричних соціологічних досліджень. Проте емпіричні дослідження на основі великих даних неможливі без вирішення низки методологічних питань, зокрема питань «перепрофілювання» методів обробки онлайн-даних, які використовуються інтернет-платформами, з метою вирішення соціологічних завдань. Це призводить до необхідності розвитку «цифрових методів» – нового напрямку в сфері методології соціологічного аналізу, що формується у зв'язку з широким поширенням великих даних. Підкреслено, що повсюдна даніфікація змінює суспільство та перевизначає людське існування в епоху великих даних, у зв'язку з чим неможна розглядати великі дані поза контекстом їхнього «темного боку». Зроблено висновок, що сьогодні нагальною проблемою є залучення соціологічної спільноти до розробки справедливої політики даних.

Ключові слова: великі дані, даніфікація, даніфіковане суспільство, цифрові методи

The article is devoted to the analysis of the phenomenon of big data and the role that they play in the study of modern society. The evolution of the interpretation of big data is characterized and it is shown that the initial technological connotation of big data associated with emphasizing the extremeness of its volume, which does not allow processing by traditional methods and tools, has undergone significant transformations due to the inclusion of the human component. Now, big data is understood as a set of digital tracks that people leave when they use information technology: they surf the Internet, download mobile applications or music, chat with friends on social networks, use GPS, buy goods in online stores, etc. Big data is generated by the Internet, but it contains information not only about the Internet, but about society and the social processes that are displayed on the Internet. In other words, big data is datafied information about everything and everyone. It is shown that big data is a new source of information both about the world around us and about the development of social processes, which turns them into a valuable base of empirical sociological research. However, empirical research based on big data is impossible without solving a number of methodological problems, in particular, questions of "re-profiling" of online data processing methods used by Internet platforms in order to solve sociological problems. This necessitates the development of "digital methods" – a new direction in the field of methodology of sociological analysis, which is formed in connection with the wide dissemination of big data. It is emphasized that widespread datafication changes society and redefines human existence in the era of big data, and therefore it is impossible to consider big data outside the context of their "dark side". It is concluded that today the urgent problem is the involvement of the sociological community in the development of a fair data policy.

Keywords: Big Data, Datafication, Datafied Society, Digital Methods

Останнє десятиріччя ознаменовано кардинальними змінами у сфері інформаційних технологій, які призвели до формування концепції «великих даних», проголошення «цифрової революції» та «революції

даних», що, за думкою Віктора Майера-Шенбергера та Кеннета Кук'є, «змінює те, як ми живемо, працюємо та мислимо» [23]. Сучасне суспільство переповнене даними, які постачаються безліччю комп'ютерів, смартфонів, соціальних мереж тощо. Наші повсякденні дії та взаємодії безперервно створюють машиночитані інформаційні масиви, які прийнято називати «великими даними». Датчики є всюди: всі рухи, дії, транзакції не можуть відбутися, не залишивши десь цифрові сліди. Активність у соціальних мережах, пошукові запити та покупки в інтернеті, оцифровані бібліотеки та архіви, навіть керування розумним телевізором зараз перетворюються на «великі дані», які збираються, накопичуються та можуть бути проаналізовані. Таким чином, *актуалізується* необхідність вивчення великих даних з позицій можливості їхнього застосування в контексті досліджень сучасного суспільства, чому й присвячена дана стаття.

Еволюція трактування терміну «великі дані»

Незважаючи на широку розповсюдженість терміну «великі дані», його значення досі оповито великою кількістю міфів. Термін використовується для опису широкого спектру понять: від технологічної здатності вилучати, зберігати та обробляти дані до культурних зрушень, що відбуваються внаслідок поширення великих даних. У зв'язку з цим передусім розглянемо контексти застосування та еволюцію трактування терміну «великі дані», без чого, на наш погляд, складно зрозуміти, яким чином їхнє існування сприятиме дослідженню сучасного суспільства.

Виникнення поняття «великі дані». Аналіз ресурсів цифрової бібліотеки ACM (Association for Computing Machinery) дає можливість стверджувати, що вперше термін «великі дані» був застосований у 1997 р. Майклом Коксом і Девідом Еллсвортом для позначення даних, яким бракує ємності комп'ютерної пам'яті для виконання певних операцій з їхньої обробки [27]. У 1998 році Джон Меші (John R. Mashey), керівник дослідницьких робіт у відомій компанії SGI (Silicon Graphics, Inc.), на конференції USENIX у доповіді «Великі дані та наступна хвиля інфраструктури» («Big Data and the Next Wave of Infrastrucure») окреслив роль та місце Big Data в розвитку науки того часу. У 2001 р. аналітик компанії Gartner Дуг Лейні запропонував трактувати великі дані шляхом виділення їхніх головних рис та розробив відому концепцію «3V», згідно з якою Big Data визначались як такі, що мають надзвичайний об'єм (Volume), мінливі внаслідок швидкого накопичення (Velocity) та різноманітні, тобто представлені у різних форматах (Variety). Отже, початкова конотація великих даних пов'язана, по-перше, з підкресленням їхніх надзвичайних об'ємів, які не дозволяють здійснювати обробку звичними методами та інструментами; по-друге, з виділенням їхніх специфічних рис. Цими питаннями займалося лише обмежене коло комп'ютерних учених. Термін «великі дані» у той час не зазнав широкого розповсюдження.

Розповсюдження. Галас навколо великих даних почався з виходу у вересні 2008 року спеціального випуску журналу «Nature», присвяченого Big Data, який привернув увагу всього наукового співтовариства. Тим більше, що в 2011 році компанія Gartner заявила, що великі дані представляють другий за значимістю тренд інформаційно-технологічної інфраструктури, що впровадження цієї технології спровокує революційні зміни практично у всіх сферах людської діяльності. З даного моменту великі дані почали привертати увагу всієї наукової спільноти, у тому числі й соціологічної [10; 11]. Вчені прагнули осягнути феномен великих даних у всій повноті та спрогнозувати їхній вплив не лише на науку, а й на розвиток суспільства. Розуміння великих даних розширилося та зазнало певних трансформацій, пов'язаних з включенням людського компоненту. Так, наприклад, Б. Хопкінсоном і Б. Евельсоном було запропоновано визначення великих даних як поєднання методів і технологій, які витягують сенс з даних на екстремальній межі практичності, а також людей з відповідними навичками аналізу, які здатні зробити доступною інформацію, приховану в різноманітні об'ємних і різнорідних даних [20]. У цей час активізуються заклики до «нвої грамотності», виникають радикальні думки стосовно того, що соціологія втрачає свої пріоритетні позиції у дослідженні суспільства, одночасно комп'ютерні вчені беруться за вирішення суто соціологічних завдань (див., наприклад: [18; 25]).

Дійсно, тенденція «приватизації» соціології представниками комп'ютерних наук актуалізувалась з появою великих даних. Наведемо висловлювання авторів книги «Big Data and Social Science: A Practical Guide to Methods and Tools»: «Світ змінився для емпіричних соціологів. Нові типи «великих даних» породили цілу нову наукову сферу – науку про дані. У цьому світі переважають комп'ютерні вчені, які створили нові способи створення та збору даних, розробили нові аналітичні та статистичні методи, запропонували нові способи візуалізації та подання інформації. Ці нові джерела даних та методів трансформують прикладні суспільні науки» [18]. Такі висловлювання не можуть не турбувати соціологічну спільноту, реальна загроза існує і потребує як артикуляції, так й пошуку шляхів її подолання.

Сучасне розуміння. Наразі найпоширенішим трактуванням великих даних є розуміння їх як величезних масивів інформації, що генеруються інтернетом у процесі користування людьми сучасними цифровими технологіями. Інакше кажучи, великі дані ототожнюються з сукупністю цифрових слідів, які залишають люди, коли здійснюють інтернет-серфінг, завантажують мобільні додатки чи музику, спілкуються з друзями в соціальних мережах, користуються GPS, купують товари в інтернет-магазинах тощо.

В Оксфордському англійському словнику термін «Big Data» визначається таким чином: «Надзвичайно великі набори даних, які можна обчислювати для розкриття моделей, тенденцій та асоціацій, особливо стосовно поведінки людей та їхньої взаємодії» [37]. Технологічний термін у даній інтерпретації тісно пов'язаний з соціальним компонентом та акцентує увагу на практичних можливостях аналізу людської діяльності. Тлумачення Кембриджського словника підкреслює важливість методів та інструментів, що застосовуються для роботи з великими даними, але в ньому також постулюється роль людини у зв'язці з технологіями: «Великі дані – дуже великі масиви даних, вироблені людьми в процесі користування інтернетом, які можуть бути збережені, зрозумілі та використані тільки за допомогою спеціальних інструментів і методів» [12]. Big Data розуміються комплексно, тобто як «інформаційні активи, що характеризуються такими великими обсягами, швидкістю та різноманітністю, що вимагають особливих технологій та аналітичних методів для перетворення їх у цінність» [14]. Дана Бойд і Кейт Кроуфорд у статті «Критичні питання великих даних» пропонують визначати великі дані як культурне, технологічне та наукове явище, що базується на взаємодії: (1) технології; (2) методах аналізу; (3) міфах та людських уявленнях [10]. Отже, останній (на даний час) етап еволюції розуміння великих даних пов'язаний з відходом від фокусування уваги на обсязі даних та складності їхньої обробки. Відбувається зміщення фокусу уваги з технологічних питань на соціогуманітарні аспекти, особливо на проблеми етичності використання великих даних.

Великі дані – це онлайн дані, проте вони криють у собі інформацію не тільки про інтернет-середовище. Великі дані наразі розуміються як соціальні сигнали інтернету, що відбивають стан суспільства та спроможні вимірювати його тенденції [30; 31], як «маніфестація соціальної поведінки», а також «прояви соціальних процесів та діяльності людини» [29], як «прояв соціальної та культурної реальності» [35]. Такий підхід до великих даних «...кидає виклик соціологам, спонукаючи узагальнити поняття соціального знання з урахуванням нових його контекстів, зумовлених формуванням цифрового середовища існування людини» [15].

Головні аспекти великих даних: інформація, технології та методи, соціальні наслідки

Аналіз еволюції розуміння великих даних сприяв виділенню головних аспектів даного поняття: інформація, технології та методи, соціальні наслідки. На кожному з виділених нами етапів превалує підвищений інтерес до одного з цих аспектів, проте всі вони необхідні для розуміння сутності великих даних.

Великі дані – небувале раніше джерело інформації. Наразі великі дані сприймаються як новий потужний ресурс досліджень сучасного суспільства, оскільки вони надають надзвичайні можливості для просування нашого розуміння стосовно поведінки людей та розвитку соціальних процесів способом, який раніше був неможливим [2; 5; 7; 10; 11]. «Порівняно з традиційними джерелами даних, «великі дані» характеризуються такими перевагами, як відносна дешевизна, висока своєчасність та новизна, що відкриває можливість оцінки характеристик та процесів, які неможливо було оцінити на основі звичайних джерел даних» [9].

Кількість інформації, що містять великі дані, вибухоподібно збільшується завдяки поширенню процесів оцифрування (digitization) та даніфікації (datafication). Оцифрування та даніфікація є двома взаємопов'язаними частинами одного процесу – переведення всієї спадщини людства в машиночитаний формат, тобто у масиви Big Data. Даніфікація наразі розглядається як концепція перетворення в формат кількісних онлайн даних всього, що оточує нас, навіть того, що, на перший погляд, не сприймається як інформація, наприклад місце розташування людини, вібрації двигуна або навантаження на міст. При цьому можливість кількісного аналізу є необхідним компонентом отриманого масиву даних.

Розбіжності між оцифруванням та даніфікацією можна висвітлити на прикладі відомого проекту Google Books стосовно повного оцифрування друкованої спадщини¹ і надання можливості людям всього світу шукати та безкоштовно переглядати книги через інтернет, див., напр., [23]. Перший текст, оцифрований Google, був представлений у вигляді відсканованих сторінок, які були перетворені в цифрові копії, доступні будь-кому через інтернет з будь-якої точки світу. Однак при цьому текст неможливо було ані знайти за ключовими словами, ані аналізувати його. Доступними були лише зображення тексту, які люди могли перечитати та перетворити в корисну інформацію. Другий крок – даніфікація, яка була здійснена Google завдяки програмі оптичного розпізнавання символів, що могла розпізнати літери, слова, речення та абзаци в цифровому зображенні. В результаті було отримано даніфіковані тексти замість оцифрованих зображень сторінок. Тепер інформація зі сторінок стала доступна не тільки для читання людьми, а й для обробки на комп'ютерах і для аналізу за допомогою алгоритмів. Крім того, тексти стали індексованими, а значить, доступними для пошуку. Служба Google Books NgramViewer надала можливості текстового аналізу, що дозволило по-новому поглянути на процес поширення ідей та розвиток людського мислення протягом століть (див., наприклад, нове застосування контент-аналізу з використанням даних безпрецедентного розміру для аналізу соціокультурної динаміки [13]).

¹ Проект Google Books стартував у 2004 році, а сьогодні ми вже користуємось наданими можливостями.

Даніфікація перетворює наші щоденні дії та взаємодії у формат даних, які можуть бути проаналізовані та надати інформацію про окремих людей, спільноти та суспільство в цілому. Наприклад, соціальні платформи (Facebook, Instagram тощо) збирають і зберігають інформацію про нас та наші переваги. Наявність цих даних, накопичених сайтами соціальних мереж і пошуковими системами, провокує прагнення використовувати їх для цілей соціологічного аналізу, оскільки традиційний збір соціологічних даних є важким процесом, який вимагає багато часу і ресурсів. Проте доступ до таких даних можуть отримати не всі, соціальні медіакомпанії обмежують доступ (особливо до транзакційних даних), що створює нову інформаційну нерівність. «Антрополог, що працює на Facebook, або соціолог, який працює в Google, матимуть доступ, а ось інша частина наукового співтовариства ні» [10]. Так, опитування понад 9000 учених-соціологів, проведене в 2016 році з метою вивчення ступеня доступності великих даних академічним дослідникам, показало, що для 32% опитаних отримання доступу до комерційних або запатентованих даних було «великою проблемою» [24].

Деякі компанії обмежують доступ до даних у всій повноті, інші продають привілей доступу за дуже високу плату; а треті пропонують невеликі набори даних для університетських дослідників. Таким чином, не всі «великі дані» рівноцінні. У зв'язку з чим Д. Бойд і К. Кроуфорд [10] пропонують їх класифікувати за критерієм відкритості: «багаті великі дані», що містять повну інформацію, але доступні лише обмеженому колу дослідників; «бідні великі дані», які не можуть претендувати на повне охоплення інформації про досліджуваний феномен, але доступні широкій аудиторії дослідників. Отже, великі дані, з одного боку, мають великий потенціал для задоволення інформаційних потреб дослідників сучасного суспільства, з іншого боку, створюють нові проблеми.

Технології та методи – другий неодмінний атрибут великих даних. Доступ до великих даних, як і їхня обробка, неможливі без застосування особливих інструментів і методів. Мабуть, тому терміни «Big Data» і «Big Data Analytics» часто ототожнюють. Всі методи великих даних засновані на концепції інтелектуального аналізу, тобто на автоматичному пошуку закономірностей у наявних даних [4]. Відрізнення інтелектуального аналізу даних та Big Data Analytics полягає в тому, що Big Data Analytics є застосуванням технологій Data Mining до нового виду даних. Можна сказати, Big Data Analytics – це поєднання онлайн-даних (цифрових слідів), технологій доступу до них та інтелектуального аналізу даних.

Перші ж спроби дослідження великих даних з метою соціологічного аналізу призвели до необхідності вирішення низки методологічних питань, а також питань «перепрофілювання» методів обробки онлайн-даних, які використовуються інтернет-платформами, з метою вирішення соціологічних завдань. У результаті сформувався підрозділ соціологічної науки – цифрова соціологія, яка фокусується на розумінні використання великих даних як частини повсякденного життя.

Цифрова соціологія виникла у відповідь на певний ажітаж навколо того, як «нові дані» перетворюють способи пізнання суспільства. Вона розуміється як «обчислювальна соціальна наука». Деякі критично налаштовані автори ототожнюють її з певною формою аналізу даних. Але цифрова соціологія пропонує альтернативу вузьким визначенням цифрових соціальних досліджень. Навпаки, цифрові соціологи прагнуть досліджувати набагато ширший набір взаємодій між даними, людьми та технологіями, які переповнюють, перевищують і не «вписуються» у просту історію про нові форми аналізу даних, що займають місце таких старих методів соціальних досліджень, як опитування чи польові роботи [22].

Цифрова соціологія, як вважають її апологети, призначена зрозуміти сучасний даніфікований світ, який потребує нових способів мислення про соціальне. Її послідовники розробляють поняття, інструменти та практики для аналізу перетину соціального та цифрового. Мета цифрової соціології – досліджувати закономірності соціального життя сучасної людини, інтегрованої у цифровий інтернет-простір. Об'єкт – цифрове суспільство як нова соціокультурна реальність. Предмет – соціальні відносини, що виникають у цифровому середовищі, цифрове соціальне життя, що включає в себе різноманітні соціальні феномени, що виникають у цифровому середовищі, а також їхній взаємозв'язок з матеріальною соціальною реальністю. Проте з самого початку свого формування цифрова соціологія зазнавала найбільшої критики у зв'язку з відсутністю теоретичного підґрунтя [5]. Сьогодні можна констатувати, що вона так й не знайшла шляхів перетворення у форму спеціальної соціологічної теорії. Натомість, соціологія поступово акумулює всі методи, розроблені цифровими соціологами, що сприяє широкій зацікавленості до напрямку, що отримав назву «цифрові методи» та по суті є розгорнутою методологією проведення соціологічних досліджень на основі великих даних, яка наразі вже розглядається «не як інновація, а як мейнстрім» [36].

Поняття «цифрові методи» було введено до наукового обігу амстердамським соціологом Р. Роджерсом у 2007 році з метою диференціації нових методів, що з'явилися в останні два десятиріччя та суттєво розширили арсенал методів соціологічних досліджень [30]. Необхідність виділення цифрових методів в якості окремого різновиду методів соціологічного аналізу була зумовлена двома факторами. 1. Соціальна реальність стала розумітися як «доповнена», тобто така, де «реальні» (матеріальні) та

«віртуальні» (онлайн) події взаємопов'язані [6; 8; 21], де цифрове середовище є природнім середовищем розгортання соціальних процесів. 2. Необхідність дослідження доповненої соціальної реальності, особливо її «віртуальної» складової, викликала міграцію соціологічних досліджень з «реального світу» у «віртуальний світ інтернету», з офлайн у онлайн [5; 30]. Ця міграція спровокувала, по-перше, потребу адаптувати традиційні методи отримання соціологічної інформації (опитування, інтерв'ю, спостереження тощо) до цифрового середовища, по-друге, необхідність розробки новітніх методів, заснованих на використанні можливостей цифрового середовища для отримання даних про різноманітні соціальні феномени. Під «віртуальними» (або «оцифрованими») методами почали розуміти методи соціологічних досліджень, перенесені зі звичної реальності (офлайн) у віртуальну реальність інтернету (тобто в онлайн) і адаптовані до онлайн-специфіки інтернет-досліджень. На відміну від них, цифрові методи – це методи, які від самого початку розроблялися для цифрового середовища, повною мірою застосовують можливості інтернету і не можуть використовуватися поза нього.

Особливість цифрових методів полягає в тому, що вони вбудовані в онлайн-пристрої (Google, Facebook, Twitter та ін.) і здатні використовувати як дані, так й обчислювальні можливості онлайн-платформ. При цьому інтернет постає не тільки джерелом даних, а й інструментом їхнього дослідження [30; 31; 39; 40]. Головна проблема, що виникає при застосуванні цифрових методів, визначається труднощами у відокремленні досліджуваних соціальних феноменів від специфіки медіа, в яких вони спостерігаються (як відомо, носій є повідомленням).

Застосування цифрових методів ґрунтується на стратегії «слідування середовищу» [31; 32; 38], етапи якої зумовлюються специфікою цифрових даних, що нерозривно пов'язані як з носієм, де вони були створені, так і з методами, якими вони можуть бути досліджені. Перший етап – відбір даних. Його мета – вибрати з величезних обсягів наявних даних те, що доступно та може сприяти досягненню мети дослідження. По-перше, треба проаналізувати адекватність джерела, що планується використовувати для отримання даних стосовно об'єкту дослідження, тобто необхідно виявити ступінь, в якій спостережуване явище поширене в середовищі, що застосовується для його вивчення. Наприклад, якщо досліджується певна спільнота, слід в'яснити, чи є вона в мережі, чи використовують її учасники певну платформу, простір або пристрій, з якого можуть збиратися дані (Twitter, Facebook, веб-сайт або мобільний додаток). По-друге, слід з'ясувати доступність даних, які містять інформацію про досліджуване соціальне явище (пошукові запити, гіперпосилання, теги, ретвіти, URL-адреси, IP-адреси користувачів, мітки часу, події, акції тощо), та визначити рівень повноти інформації, що міститься в них.

Другий етап стратегії «слідування середовищу» – вибір інструментів аналізу. На цьому етапі належить з'ясувати, які саме інструменти (тобто пристрої інтернету) обробляють обрані об'єкти, та відібрати з них ті, що можуть слугувати аналізу.

Третій етап – перепрофілювання визначених пристроїв з урахуванням специфіки об'єкту дослідження. Тут треба враховувати, що дослідження, засновані на цифрових методах, є різновидом вторинних досліджень, оскільки дані (цифрові сліди), які слугують емпіричною базою, створювалися не для цілей соціологічного аналізу, тому їхнє застосування потребує нетривіальної операціоналізації та узгодження дослідницьких питань з цифровими медіа.

На четвертому етапі вирішується питання стосовно обґрунтованості висновків, зроблених на основі онлайн-даних. Тобто виникає потреба їхньої перевірки традиційними офлайн-методами.

Лідером розвитку цифрових методів є провідна європейська дослідницька група «The Digital Methods Initiative» (DMI), яка складається з соціологів та дослідників нових медіа, що розробляють методи та інструменти перепрофілювання інтернет-пристроїв і платформ (таких як Twitter, Facebook, Google) для дослідження соціальних та політичних питань.

Становлення цифрових методів є результатом еволюції онлайн-методів дослідження середовища, тобто вони повною мірою використовують можливості онлайн-інструментів для цілей соціологічних досліджень завдяки їхньому перепрофілюванню. Перепрофілювання (вторинне використання) – дуже важливе поняття в цифрових методах. Його сутність можна пояснити на прикладі такого поширеного пристрою, як Google Analytics, створеного з метою допомогти веб-майстрам аналізувати інтернет-сайти, наприклад, переглядати наявність посилань на свої сторінки з інших ресурсів та оптимізувати їхню видимість (первинне використання). Проте пошукові запити можна розглядати значно ширше, ніж це роблять веб-майстри. Пошукові запити містять у собі ознаки соціальних проблем, є маркерами соціальної динаміки. Їхнє вивчення дає можливість перейти від дослідження цифрового середовища до використання інтернет-пристроїв для вивчення суспільства [22; 31]. Отже, сьогодні цифрові методи є найбільш розробленою та апробованою методологією дослідження сучасного суспільства на основі великих даних, що підтверджується численними публікаціями, наприклад [34].

Соціальні наслідки – третій аспект великих даних. Вплив великих даних на суспільство зазвичай описують через історії успіху стартапів, які впроваджують технології Big Data. Проте цим вплив великих даних не обмежується. Піонер у галузі великих даних Алекс Петланд писав: «3 Big Data ми

можемо почати реально розглядати деталі соціальної взаємодії та те, як вони розігруються і більше не обмежуються такими середніми показниками, як ринкові індекси або результати виборів. Це приголомшлива зміна. Можливість бачити деталі ринку, політичні революції та вміти передбачати та контролювати їх – це, безумовно, можна порівняти з даром Прометею, який може бути використаний як для добра, так й для зла» [25].

Повсюдна даніфікація змінює суспільство: «На відміну від цифровізації, яка дала змогу підвищити продуктивність та підвищити ефективність уже існуючих процесів, даніфікація обіцяє повністю перевизначити майже кожен аспект нашого існування як людей на цій планеті» [17]. Як буде перевизначено людське існування у зв'язку з даніфікацією? Яким стане суспільство? Ми вже бачимо, що явно не «суспільством знання», навіть не «цифровим суспільством». Сьогодні все частіше в науковому дискурсі з'являється твердження, що ми самі того не помітивши перейшли у *даніфіковане суспільство, тобто суспільство, засноване на великих даних* [15; 16; 19; 28; 34].

У літературі багато прикладів позитивних наслідків застосування великих даних, але «темний бік великих даних» до останнього часу дуже мало дискутувався. Чому? Мабуть, тому, що сутність феномену великих даних й досі не повністю осягнута, досі немає розуміння, що великі дані стосуються не тільки комп'ютерних учених, а всіх, навіть тих, хто про них нічого не знає. Ми всі вже перетворилися на «суб'єктів даних». «Дані не лише формують наші соціальні відносини, вподобання та життєві шанси, але й наші демократії» [33].

Останнім часом голосно звучать заклики до встановлення меж втручання великих даних у людське життя. Цьому значно сприяв нещодавній скандал, пов'язаний з Facebook, коли особиста інформація 87 мільйонів користувачів потрапила до компанії Cambridge Analytica (CA), яка використала могутність великих даних з метою впливу на думку виборців шляхом точно розрахованої персоналізованої реклами. Цей скандал має позитивний бік: «Хороша річ, що стосується скандалу з CA, – це те, що він розпочав тривалу дискусію щодо використання даних соціальних медіа» [40].

Існує безліч ентузіазму та оптимізму щодо того, як уряди всіх рівнів можуть використовувати великі дані, алгоритми та штучний інтелект. Проте одночасно зростає стурбованість ризиками, які виникають із цими новими системами. Оскільки уряди широко використовують великі дані, існує низка реальних та потенційних наслідків, про які треба знати та бути напоготові: 1) заохочується перманентний збір даних про громадян, у зв'язку з чим громадяни перетворюються на «суб'єктів даних»; 2) все більше поширюються автоматизовані сервіси прийняття управлінських рішень, алгоритми яких невідомі широкому загалу; 3) громадяни стають «безперервно пізнаними», але мають мало можливостей дізнатися, як їхні дані збираються та використовуються; 4) акценти зміщуються від причинного зв'язку до кореляції, від запобігання до випередження на основі прогнозів, зроблених штучним інтелектом; 5) заохочується розширення партнерства між державним та приватним секторами, а також участь ІТ-корпорацій в управлінні та наданні державних послуг [1].

Зрозуміло, що без належної прозорості, підзвітності та контролю системи великих даних можуть бути використані таким чином, що порушує громадянські права. Ключовим тут є визнання, що при великих даних несприятливі наслідки можуть бути навіть не навмисними зловживаннями, а просто «глюком системи». На даний момент громадяни не мають інформації або ресурсів, необхідних для змістовної взаємодії з цими змінами, які вже відбуваються.

Приклад Китаю, який з 2014 року запроваджує систему соціального рейтингу, яку ще називають системою соціального кредиту, просто жахає. Згідно з «Програмою створення системи соціального кредиту», вже до 2020 року в країні повинна бути запущена система оцінки кожного громадянина, що буде працювати в режимі реального часу, а результати її роботи мають публікуватися в інтернеті. Артикульована мета програми – побудова суспільства, в якому «бути чесним і вартим довіри стане престижним й бажаним». Критерії оцінювання обирає державна влада, яка має право визначати, чи варто вам довіряти. На додачу, рейтинг буде публічною інформацією, від якої залежатиме, чи отримаєте ви кредит, нову посаду, можливість подорожувати і навіть підуть з вами на побачення чи ні.

Алгоритм обчислення рейтингу повністю не розкривається, але оприлюднено 5 головних факторів, що впливають на рейтинг: 1) кредитна історія; 2) здатність виконувати взяті на себе зобов'язання; 3) верифіковані особисті дані, наприклад, адреса чи номер мобільного телефону; 4) особисті переваги та поведінка; 5) стосунки між громадянами.

Як же нас бачить штучний інтелект очима своїх алгоритмів? Якщо ви вчасно не оплатили комунальні рахунки або кредит, то ваш соціальний рейтинг знижується. Система не бачить і не враховує контексти, наприклад, що рахунки не оплачені, оскільки ви лежали в лікарні. Крім того, штучний інтелект оцінює людей за типами продуктів, які вони споживають. Так, особа, що проводить багато часу за відеоіграми, вважатиметься ненадійною. Ті, хто часто купує підгузки, імовірно мають маленьку дитину, отже вважаються відповідальними. П'ятий фактор означає врахування те тільки поведінки певної особи, а й всіх її онлайн-друзів, тобто буде враховуватися, якими словами і про що ви переписуєтеся у чатах та

соціальних мережах, які повідомлення та коментарі залишає (на підтримку уряду, чи ні, схвалює економічний стан країни, чи ні). Навіть ставити «лайк» громадянам Китаю доведеться дуже обережно, оскільки він може змінити їхній особистий соціальний рейтинг довіри. Таким чином, система спрямовано схилитиме громадян уникати сценаріїв поведінки, які не подобаються уряду.

У західних IT-виданнях було багато критики стосовно китайської системи соціального рейтингу. Проте ми не маємо ніяких гарантій, що в інших країнах не створюються передумови для впровадження подібних систем. Китайський уряд хоча би повідомив громадян про свої «соціальні інновації». А тим часом алгоритми Facebook знають наших друзів та можуть розпізнати нас на фотографії. Google взагалі «знає про нас більше, ніж ми знаємо про себе» [3]. Ще у 2015 році OECD (Організація економічного співробітництва та розвитку) підрахувала, що на 100 жителів США припадає не менше 24,9 підключених пристроїв. Всі компанії, які виробляють і обслуговують такі гаджети, збирають безліч інформації про користувачів та їхні звички, щоб розуміти і прогнозувати онлайн-поведінку. «Уряди більшості країн займаються мониторингом. Американське АНБ навіть і не заперечує масштабного стеження за всіма громадянами. Жителі США вже живуть у світі, де алгоритми вирішують, чи є особа небезпечною або походить із зони ризику, чи є вона порядним громадянином та чи можна їй довіряти. Захід набагато ближче до Сходу, який трансформує кредитний рейтинг у соціальну систему оцінювання SCS. От тільки, на відміну від китайців, американців про це офіційно не повідомили» [1].

Якщо не бути пильними та не втручатися у формування даніфікованого суспільства, заснованого на великих даних, ми можемо опинитися в суспільстві безпрецедентної диктатури, суспільстві, де стеження є нормою, а життя пересічних громадян перетвориться на нескінченний конкурс популярності, в якому всі, окрім обраних, змагаються за високий рейтинг, поступаючись власними принципами і вподобаннями.

Як вплинути на ситуацію? Перш за все, потрібна повна відкритість, надійні та прозорі механізми, що управлятимуть процесом і забезпечать належне використання інформації. Щоб довіряти системі, вона не повинна бути «чорним ящиком», в ній не має залишитися жодного нерозкритого елемента, а також не має бути можливостей корегування рейтингів для обраних. Крім того, потрібно встановити рейтинг довіри для самих оцінювачів.

Підводячи підсумок, зазначимо, що великі дані – це не тільки джерело інформації про сучасне суспільство, а й інструмент його дослідження. Цей інструмент поки що далеко не повною мірою опанований соціологами. Проте робота в цьому напрямку йде, лише хотілося би, щоб наша країна активніше залучилася до неї. Безумовно, треба бути відкритими для інновацій, критично осмислювати нові епістемологічні підходи, активніше впроваджувати практику соціологічного аналізу на основі великих даних. Все так, проте, на нашу думку, більш нагальною проблемою на сьогодні є необхідність визнати, що даніфікація немуніча, вона вже відбулась. І якщо дані не використовуються кимось (владою чи корпораціями) з корисною метою, то це не означає, що такої можливості нема. У зв'язку з цим соціологічна спільнота повинна залучитися до розробки справедливої політики даних.

Література:

1. Ботсман Р. Китайське попередження — як працює соціальний рейтинг на основі Big Data, 2017. URL: <https://www.imena.ua/blog/big-data-big-brother-china>
2. Губа К. Больше данные в социологии: новые данные, новая социология? *Социологическое обозрение*. Т. 17 (1), 2018. С. 213-236.
3. Докучаев К. Google знает о вас всё: 14 вещей, которые могут вас удивить, 2017. URL: <https://blog.themarfa.name/google-znaiet-o-vas-vsio/>
4. Кислова О. Н. Интеллектуальный анализ данных: история становления термина. *Український соціологічний журнал*. №1-2, 2011. С. 83-94.
5. Кислова О. Н. Быть или не быть цифровой социологии? *Вісник Харківського національного університету імені В. Н. Каразіна серія «Соціологічні дослідження сучасного суспільства: методологія, теорія, методи»*. № 1045, 2013 С. 9-15.
6. Кислова О. Н. Дополненная реальность сквозь призму интернет-практик современного студенчества. *Методологія, теорія та практика соціологічного аналізу сучасного суспільства*. Випуск 20, 2014. С. 351-356.
7. Кислова О. Н. Социология в контексте вызова «больших данных»: роль интеллектуального анализа данных в становлении новых подходов к социологическим исследованиям. *Вісник Харківського національного університету імені В. Н. Каразіна серія «Соціологічні дослідження сучасного суспільства: методологія, теорія, методи»*. № 1148, 2015. С. 57-62.
8. Кислова О. Н., Николаевская А. М. Концепция дополненной реальности: от технологии к социологии. *Вісник Харківського національного університету імені В. Н. Каразіна серія «Соціологічні дослідження сучасного суспільства: методологія, теорія, методи»*. Т. 37, 2016. С. 49-54.
9. Саріогло В. Г. Великі дані як джерело інформації та інструментарій для офіційної статистики: потенціал, проблеми, перспективи. *Статистика України*. № 4, 2016. С.12-19.
10. Boyd D., Crawford K. Critical Questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon. *Information, Communication & Society*. Vol. 15 (5), 2013. P. 662-679.

11. Burrows R., Savage M. After the crisis? Big Data and the methodological challenges of empirical sociology. *Big data & societ.* Vol. 1(1), 2014. URL: <https://journals.sagepub.com/doi/10.1177/2053951714540280>
12. Cambridge Dictionary URL: <http://dictionary.cambridge.org/dictionary/english/big-data>.
13. Chen Y., Yan F. Centuries of sociology in millions of books. *The Sociological Review*. Vol. 64 (4), 2016 P. 872-893.
14. De Mauro A., Greco M., Grimaldi M. What is Big Data? A Consensual Definition and a Review of Key Research Topics. *AIP Conference Proceedings*. Vol. 1644 (1), 2015 P. 97-104.
15. Dencik L. Understanding the Datafied Society by Decentring Data. *iCS Symposium*. 2018. URL: <https://snurb.info/node/2424>.
16. Di Salvo P. (2019). *The role of whistleblowing in the datafied society*. URL: <https://en.ejo.ch/digital-news/technology/the-role-of-whistleblowing-in-the-datafied-society/>
17. Ericsson E. *The Impact of Datafication on Specific Landscapes*. 1st ed. [ebook] London: Ericsson, Imperial College Business School & UK's Sustainable Society Network, 2014. URL: <https://www.ericsson.com/assets/local/news/2014/4/the-impact-of-datafication-on-strategic-landscapes.pdf>
18. Foster I., Ghani R., Jarmin R. S., Kreuter F., Lane J. *Big data and social science: A practical guide to methods and tools*. London, UK/New York, NY: CRC Press, 2016.
19. Hintz A., Dencik L., Wahl-Jorgensen K. *Digital citizenship in a datafied society*. John Wiley & Sons, 2018.
20. Hopkins B., Evelson B. Forrester: Big data – start small, but scale quickly. *Computer Weekly*, 2012. URL: <http://www.computerweekly.com/opinion/Forrester-Big-Data-Start-Small-But-Scale-Quickly>
21. Jurgenson N. When atoms meet bits: Social media, the mobile web and augmented revolution. *Future Interne.* № 4(1), 2012. P. 83-91.
22. Marres N. *Digital Sociology The Reinvention of Social Research*. Cambridge: Polity Press, 2017.
23. Mayer-Schönberger V., Cukier K. *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013.
24. Metzler K., Kim D. A., Allum N., Denman A. *Who is doing computational social science? Trends in big data research* (White paper). London, UK: SAGE Publishing, 2016.
25. Pentland A. *Social Physics: How Good Ideas Spread. The Lessons from a New Science*. New York: The Penguin Press, 2014.
26. Pentland A. Reinventing society in the wake of big data systems, 2012 URL: <http://edge.org/conversation/reinventing-society-in-the-wake-of-big-data>.
27. Press G. *A Very Short History of Big Data*, 2013 URL: <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data>.
28. Redden J. Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*. Vol. 5 (2), 2018. URL: <https://journals.sagepub.com/doi/full/10.1177/2053951718809145>.
29. Resnyansky L. Conceptual frameworks for social and cultural Big Data analytics: Answering the epistemological challenge. *Big Data & Society*. Vol. 6(1), 2019. URL: <https://journals.sagepub.com/doi/10.1177/2053951718823815>
30. Rogers R. A. *The End of the Virtual: Digital Methods*. Amsterdam: Amsterdam University Press, 2009.
31. Rogers R. Digital methods for web research. *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, 2015. P. 1-22.
32. Rogers, R. *Doing Digital Methods*. SAGE Publications Limited, 2019.
33. Ruppert E., Isin E., Bigo D. Data politics. *Big Data & Society*. Vol. 4(2). 2017. URL: <https://journals.sagepub.com/doi/full/10.1177/2053951717717749>
34. Schäfer M. T., Van Es K. F. *The datafied society: Studying culture through data*. Amsterdam University Press, 2017.
35. Shaw R. Big data and reality. *Big Data & Society*. Vol. 1(4), 2015. P. 1-4.
36. Snee H., Hine C., Morey Y., Roberts S., Watson H. *Digital Methods for Social Science: An Interdisciplinary Guide to Research Innovatio*. Palgrave Macmillan, London, 2015.
37. The Oxford English dictionary. URL: https://en.oxforddictionaries.com/definition/big_data
38. Venturini T., Bounegru L., Gray J., Rogers R. A reality check (list) for digital methods. *New media & society*. Vol. 20(11), 2018. P. 4195-4217.
39. Venturini T., Latour B. The Social Fabric: Digital footprints and quali-quantitative methods. *Proceedings of future en seine*, 2009. P. 87-103
40. Venturini T., Rogers R. API-Based Research or How can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach. *Digital Journalism*, 2019. P. 1-9.

Большие данные в контексте исследования проблем современного общества

Статья посвящена анализу феномена больших данных и роли, которую они играют в исследовании современного общества. Охарактеризована эволюция трактовки больших данных и показано, что первоначальная технологическая коннотация больших данных связанная с подчеркиванием чрезвычайности их объемов, которые не позволяют осуществлять обработку привычными методами и инструментами, претерпела значительные трансформации, обусловленные включением человеческого компонента. Сейчас большие данные понимаются как совокупность цифровых следов, которые оставляют люди, когда используют информационные технологии: осуществляют интернет-серфинг, загружают мобильные приложения или музыку, общаются с друзьями в социальных сетях, пользуются GPS, покупают товары в интернет-магазинах и т.п. Большие данные генерируются интернетом, однако они содержат информацию не столько об интернете, сколько об обществе и о социальных процессах, которые отображаются в интернете. Другими словами, большие данные – данифицированная информация обо всем и обо всех. Показано, что большие данные являются новым источником информации как об окружающем мире, так и о

развитии социальных процессов, что превращает их в ценную базу эмпирических социологических исследований. Однако эмпирические исследования на основе больших данных невозможны без решения ряда методологических вопросов, в частности вопросов «перепрофилирования» методов обработки онлайн-данных, используемых интернет-платформами, с целью решения социологических задач. Это обуславливает необходимость развития «цифровых методов» – нового направления в сфере методологии социологического анализа, которое формируется в связи с широким распространением больших данных. Подчеркнуто, что повсеместная данификация изменяет общество и переопределяет человеческое существование в эпоху больших данных, в связи с чем нельзя рассматривать большие данные вне контекста их «темной стороны». Сделан вывод, что сегодня насущной проблемой является привлечение социологического сообщества к разработке справедливой политики данных.

Ключевые слова: большие данные, данификация, данифицированное общество, цифровые методы

References:

1. Botsman, R. (2017) Chinese warning – how social rating based on Big Data works [online]. Available at: <https://www.imena.ua/blog/big-data-big-brother-china/> [in Ukrainian].
2. Guba, K. (2018) 'Big Data in Sociology: New Data, New Sociology?' *Sociological review*. № 17 (1). P. 213-236 [in Russian].
3. Dokuchaev, K. (2017) Google knows everything about you: 14 things that may surprise you [online]. Available at: <https://blog.themarfa.name/google-znaiet-o-vas-vsio/> [in Russian].
4. Kyslova, O. N. (2011) 'Intelligent Data Analysis: the history of the making of the term'. *Ukrainian Sociological Journal*. № 1-2. P. 83-94 [in Russian].
5. Kyslova, O. N. (2013) 'To be or not to be of digital sociology?' *Visnyk of V. N. Karazin Kharkiv National University. Series «6/ sociological studies of contemporary society: methodology, theory, methods»*. № 1045. P. 9-15 [in Russian].
6. Kislova, O. N. (2014) «Augmented Reality» through the Prism of the Internet Practices of Modern Students. *Karazin Kharkiv National University. Series «Sociological studies of contemporary society: methodology, theory, methods»*. № 20. P. 351-356 [in Russian].
7. Kyslova, O. N. (2015) 'Sociology in the context of the "Big Data" challenge: the role of Intelligent Data Analysis in the development of new approaches to sociological study'. *Visnyk of V. N. Karazin Kharkiv National University. Series «Sociological studies of contemporary society: methodology, theory, methods»* № 1148, 57-62 [in Russian].
8. Kyslova, O. N., Nikolaevska A. M. (2016). The concept of augmented reality: from technology to sociology. *Visnyk of V. N. Karazin Kharkiv National University. Series «Sociological studies of contemporary society: methodology, theory, methods»*. № 37. P. 49-54 [in Russian].
9. Sarioglo, V. H. (2016) Big Data as an Information Source and a Toolkit for Official Statistics: Capacities, Problems, Prospects. *Statistics of Ukraine*. № 4. P. 12-19. [in Ukrainian]
10. Boyd, D., Crawford K. (2013) Critical Questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon. *Information, Communication & Society*. Vol. 15 (5). P. 662-679.
11. Burrows, R., and Savage, M. (2014) After the crisis? Big Data and the methodological challenges of empirical sociology. *Big data & societ.* Vol. 1(1) [online]. Available at: <https://journals.sagepub.com/doi/10.1177/2053951714540280>
12. Cambridge Dictionary [online]. Available at: <http://dictionary.cambridge.org/dictionary/english/big-data>.
13. Chen, Y., and Yan, F. (2016) Centuries of sociology in millions of books. *The Sociological Review*. Vol. 64 (4). P. 872–893.
14. De Mauro, A., Greco, M. and Grimaldi, M. (2015) A. What is Big Data? A Consensual Definition and a Review of Key Research Topics. *AIP Conference Proceedings*. Vol. 1644 (1). P. 97-104.
15. Dencik, L. (2018) Understanding the Datafied Society by Decentering Data. *iCS Symposium* [online]. Available at: <https://snurb.info/node/2424>
16. Di Salvo, P. (2019). *The role of whistleblowing in the datafied society* [online]. Available at: <https://en.ejo.ch/digital-news/technology/the-role-of-whistleblowing-in-the-datafied-society/>.
17. Ericsson, (2014) *The Impact of Datafication on Specific Landscapes*. 1st ed. [ebook] London: Ericsson, Imperial College Business School & UK's Sustainable Society Network. Available at: <https://www.ericsson.com/assets/local/news/2014/4/the-impact-of-datafication-on-strategic-landscapes.pdf>.
18. Foster, I., Ghani, R., Jarmin, R.S., Kreuter F. and Lane, J. (2016). *Big data and social science: A practical guide to methods and tools*. Chapman and Hall / CRC Press.
19. Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2018) *Digital citizenship in a datafied society*. Medford: Polity Press.
20. Hopkins, B., & Evelson B. (2012) Forrester: Big data – start small, but scale quickly. *Computer Weekly* [online]. Available at: <http://www.computerweekly.com/opinion/Forrester-Big-Data-Start-Small-But-Scale-Quickly>.
21. Jurgenson, N. (2012) When atoms meet bits: Social media, the mobile web and augmented revolution. *Future Interne.* Vol. 4(1). P. 83-91.
22. Marres, N. (2017) *Digital Sociology The Reinvention of Social Research*. Cambridge: Polity Press.
23. Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. London: SAGE Publishing.
24. Metzler, K., Kim, D. A., Allum, N., & Denman, A. (2016) *Who is doing computational social science? Trends in big data research* (White paper). London: SAGE Publishing [online]. Available at: <http://repository.essex.ac.uk/17679/1/compsocsci.pdf>.
25. Pentland, A. (2014) *Social Physics: How Good Ideas Spread. The Lessons from a New Science*. New York: The Penguin Press.

26. Pentland, A. (2012) Reinventing society in the wake of big data systems [online]. Available at: <http://edge.org/conversation/reinventing-society-in-the-wake-of-big-data>
27. Press, G. (2013) A Very Short History of Big Data [online]. Available at: <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data>.
28. Redden, J. (2018) Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*. Vol. 5 (2) [online]. Available at: <https://journals.sagepub.com/doi/full/10.1177/2053951718809145>.
29. Resnyansky, L. (2019) Conceptual frameworks for social and cultural Big Data analytics: Answering the epistemological challenge. *Big Data & Society*. Vol. 6(1) [online]. Available at: <https://journals.sagepub.com/doi/10.1177/2053951718823815>
30. Rogers, R. (2009) A. *The End of the Virtual: Digital Methods*. Amsterdam: Amsterdam University Press.
31. Rogers, R. (2015) Digital methods for web research. *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*. Hoboken, NJ: John Wiley & Sons [online]. Available at: http://www.govcom.org/publications/full_list/etrds0076.pdf.
32. Rogers, R. (2019) *Doing Digital Methods*. SAGE Publications Limited.
33. Ruppert, E., Isin, E. & Bigo, D. (2017) Data politics. *Big Data & Society*. Vol. 4 (2) [online]. Available at: <https://journals.sagepub.com/doi/full/10.1177/2053951717717749>.
34. Schäfer, M. T., Van Es, & K. F. (2017) *The datafied society: Studying culture through data*. Amsterdam: Amsterdam University Press.
35. Shaw, R (2015) Big data and reality. *Big Data & Society*. Vol.1(4). P. 1–4.
36. Snee, H., Hine, C., Morey, Y., Roberts, S. & Watson, H. (2015) *Digital Methods for Social Science: An Interdisciplinary Guide to Research Innovatio*. London: Palgrave Macmillan.
37. The Oxford English dictionary [online]. Available at: https://en.oxforddictionaries.com/definition/big_data.
38. Venturini, T., Bounegru, L., Gray, J., & Rogers, R. (2018) A reality check (list) for digital methods. *New media & society*. Vol. 20(11). P. 4195-4217.
39. Venturini, T., & Latour, B. (2009) The Social Fabric: Digital footprints and quali-quantitative methods. *Proceedings of Future on Seine*. [online]. Available at: <https://core.ac.uk/download/pdf/35307793.pdf>.
40. Venturini, T., & Rogers, R. (2019) API-Based Research or How can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach. *Digital Journalism*. Vol. 7 (4). P. 1-9 [online]. Available at: <https://www.tandfonline.com/doi/full/10.1080/21670811.2019.1591927>.