

ВИКОРИСТАННЯ «ВЕЛИКИХ ДАНИХ» ДЛЯ СТАТИСТИЧНОГО ЗАБЕЗПЕЧЕННЯ УПРАВЛІННЯ РОЗВИТКОМ «РОЗУМНИХ» СТАЛИХ МІСТ

©2018 КОРЕПАНОВ О. С.

УДК 311.22:[316.32:004.73]

Корепанов О. С. Використання «великих даних» для статистичного забезпечення управління розвитком «розумних» сталих міст

Мета статті полягає у проведенні аналізу основних проблем використання «великих даних» для статистичного забезпечення управління розвитком «розумних» сталих міст, а також в обґрунтуванні шляхів їх подолання. Проаналізовано основні визначення суті концепції «великих даних», з якими погоджується більшість науковців. Обґрунтовано необхідність розроблення додаткових методів і процесів аналізу на базі оновлених технологій і методів для того, щоб ефективно аналізувати «великі дані» та діяти на підставі отриманих результатів. Здійснено групування джерел «великих даних», які доцільно використовувати в інформаційному забезпеченні моделі «розумних» сталих міст, наведено їх характеристику та приклади. Також здійснено порівняльний аналіз традиційних і «великих даних» за їх основними ознаками, зазначено їх відмінності. Визначено переваги «великих даних», які обумовлюють можливість їх використання для інформаційного забезпечення розвитку «розумних» сталих міст, а також шляхи подолання проблем їх використання.

Ключові слова: «великі дані», «розумне» стале місто, статистичне забезпечення управління, аналіз розвитку.

Рис.: 1. **Табл.:** 3. **Бібл.:** 10.

Корепанов Олексій Сергійович – кандидат економічних наук, доцент, доцент кафедри статистики, обліку та аудиту, Харківський національний університет ім. В. Н. Каразіна (пл. Свободи, 4, Харків, 61022, Україна)

E-mail: o.s.korepanov@karazin.ua

УДК 311.22:[316.32:004.73]

Корепанов А. С. Использование «больших данных» для статистического обеспечения управления развитием «умных» устойчивых городов

Цель статьи состоит в проведении анализа основных проблем использования «больших данных» для статистического обеспечения управления развитием «умных» устойчивых городов, а также в обосновании путей их преодоления. Проанализированы основные определения сути концепции «больших данных», с которыми согласно большинство научных деятелей. Обоснована необходимость разработки дополнительных методов и процессов анализа на базе обновленных технологий и методов для того, чтобы эффективно анализировать «большие данные» и действовать на основании полученных результатов. Осуществлено группирование источников «больших данных», которые целесообразно использовать в информационном обеспечении модели «умных» устойчивых городов, приведены их характеристика и примеры. Также осуществлен сравнительный анализ традиционных и «больших данных» по их основным признакам, указаны их отличия. Определены преимущества «больших данных», которые обуславливают возможность их использования для информационного обеспечения развития «умных» устойчивых городов, а также пути преодоления проблем их использования.

Ключевые слова: «большие данные», «умный» устойчивый город, статистическое обеспечение управления, анализ развития.

Рис.: 1. **Табл.:** 3. **Библ.:** 10.

Корепанов Алексей Сергеевич – кандидат экономических наук, доцент, доцент кафедры статистики, учета и аудита, Харьковский национальный университет им. В. Н. Каразина (пл. Свободы, 4, Харьков, 61022, Украина)

E-mail: o.s.korepanov@karazin.ua

UDC 311.22:[316.32:004.73]

Korepanov O. S. Using «Big Data» for Statistical Support of Management of Development of «Smart» Sustainable Cities

The article is aimed at analyzing the main problems of using «big data» for statistical support of management of development of «smart» sustainable cities, as well as substantiating ways of overcoming them. The basic definitions of the essence of the conception of «big data» are analyzed in accordance with consensus of the majority of scientific figures. The necessity of elaboration of additional methods and processes of analysis on the basis of the updated technologies and methods is substantiated in order to efficiently analyze «big data» and to act on the basis of the obtained results. The groupings of «big data» sources, which are expedient to use in information support of the model of «smart» sustainable cities, are accomplished, their characterization and examples are provided. A comparative analysis of traditional and «big data» by their basic attributes is also carried out, their differences are indicated. The advantages of «big data», which stipulate possibility to use them for information support of development of «smart» sustainable cities, and also ways of overcoming problems as to their use are defined.

Keywords: «big data», «smart» sustainable city, statistical support of management, analysis of development.

Fig.: 1. **Tbl.:** 3. **Bibl.:** 10.

Korepanov Olexsiy S. – PhD (Economics), Associate Professor, Associate Professor of the Department of Statistics, Accounting and Auditing, V. N. Karazin Kharkiv National University (4 Svobody Square, Kharkiv, 61022, Ukraine)

E-mail: o.s.korepanov@karazin.ua

Останнім часом поширюється дискусія, як громади, уряди та підприємства можуть максимально використовувати переваги «великих» даних. Приватний сектор спочатку використовував «великі» дані для збільшення доходів та підвищення ефективності. Ця практика поширилася у світовій статистичній спільноті.

П'ятий принцип Фундаментальних принципів щодо офіційної статистики стосується доступу до да-

них. Він наголошує, що «дані для статистичних цілей можуть братися з усіх типів джерел, будь то статистичні обстеження або адміністративні картотеки. Статистичні установи повинні вибирати джерело з урахуванням міркувань якості, оперативності, витрат і тягаря, що лягає на респондентів». Цей принцип може тлумачитися як включення доступу до «великих» даних [4].

Єдиного унікального визначення цього відносно нового явища, відомого як «великі дані», не існує.

На базовому рівні це розуміється як набори даних, об'єм, швидкість або різноманітність яких є дуже високими порівнянн з типами наборів даних, які традиційно використовувались.

Доцільно послатися на такі визначення суті цієї концепції, з якими погоджується більшість науковців.

Перше визначення запропонував Мерв Адріан у 2011 р. в статті для журналу *Teradata Magazine*: «Великі дані – це дані, збір, управління і обробку яких неможливо здійснити з допомогою найбільш часто використовуваних апаратних середовищ і програмних інструментів протягом допустимого для користувача часу» [3].

«Великі дані» можна визначити на основі тлумачення, яке наведено у глосарії компанії «Гартнер», провідної світової компанії, що здійснює наукові дослідження та консультації в галузі інформаційних технологій [5]: «Великі дані представляють собою джерела даних, які в загальному сенсі можна описати як «інформаційні активи великого обсягу, високої швидкості і/або високого рівня, що вимагають ефективних і інноваційних форм обробки інформації та забезпечують більш глибоке розуміння, прийняття рішень і автоматизацію технологічних процесів».

У доповіді McKinsey Global Institute в травні 2011 р. представлено таке визначення: «Великі дані – це набори даних, розміри яких виходять за межі можливостей звичайного програмного забезпечення зі збирання, зберігання, управління й аналізу даних» [2].

З цих визначень випливає: об'єкт, що вважається «великими даними», змінюється з розвитком технологій. Дані, що колись були «великими», або дані, що вважаються «великими» сьогодні, будуть відрізнятися від «великих даних» завтрашнього дня. Наведені визначення означають, що суть «великих даних» може відрізнятися залежно від галузі або навіть організації, якщо існує значна різниця в можливостях інструментів і технологій.

Поняття «великі» стосується не тільки обсягу даних. Хоча, перш за все, мається на увазі наявність великої кількості даних, водночас це передбачає присутність деяких інших ознак. «Великі дані» характеризуються збільшеною швидкістю їх передачі, складністю і різноманітністю джерел порівняно з джерелами даних минулого. Такі фактори ускладнюють роботу з «великими даними», оскільки доводиться мати справу не просто з великою кількістю даних, а з тим, що вони надходять дуже швидко, у складних формах і з різноманітних джерел.

Необхідно розробити додаткові методи і процеси аналізу на базі оновлених технологій і методів для того, щоб ефективно аналізувати «великі дані» та діяти на підставі отриманих результатів.

Метою статті є аналіз основних проблем використання «великих даних» для статистичного забезпечення управління розвитком «розумних» сталих міст, а також обґрунтування шляхів їх подолання.

Поряд із загальними досягненнями в галузі інформаційно-комунікаційних технологій важливим компонентом, який сприяє суттєвому прискореному розвитку «розумних» сталих міст, є «великі дані». Їх використання забезпечує оперативне інформування громадян про нові форми споживання та попиту на послуги, які спираються на «розумні» системи.

Громадяни «розумного» міста спонукають органи місцевого самоврядування до розгляду та інвестування у відповідні процеси та системи для підтримки сталого розвитку. У цих процесах реалізується найкраща практика та з'являються нові проекти та пропозиції, оскільки це динамічне середовище.

Міста, що використовують сучасні інноваційні підходи до аналізу «великих даних», які можуть бути порівняні з міжнародними стандартами, позиціонуються для майбутнього успіху і мають можливість просувати суспільство вперед, у більш «розумне» майбутнє.

«Великі дані» характеризуються такими ознаками, що наведено на *рис. 1*.

Одне з найчастіше використовуваних визначень, що представлено компанією Gartner, включає, крім обсягу (*Volume*), такі характеристики даних, як швидкість (*Velocity*) та різноманітність (*Variety*) [8]. «Швидкість» означає швидкість створення, оцінювання та аналізу даних, тоді як термін «різноманітність» включає той факт, що дані можуть існувати як у різних формах (текст, аудіо, відео), так і в різних форматах (структуровані та неструктуровані). Визначення трьох *V* поширюється і розширюється. Четверте *V* – достовірність (*Veracity*) – була введена для визначення аспектів, що стосуються якості та походження даних, а також невизначеності, яка може існувати при їх аналізі [6]. П'яте *V* – значення (*Value*) – включено деякими дослідниками, щоб відмітити потенційно високу соціально-економічну цінність, яка може бути отримана «великими даними» [7].

За наявності зазначених вище п'яти ознак накопичені обсяги даних можна відносити до «великих».

Статистична комісія Організації Об'єднаних Націй (СК ООН) і національні статистичні організації (НСО) вивчають шляхи використання великих джерел даних для доповнення офіційної статистики та кращого виконання їх завдань для надання своєчасних і достовірних даних для формування політики. СК ООН визнала, що «великі дані» представляють собою джерело інформації, яке не можна ігнорувати і яке необхідно оцінити належним чином. Для цього СК ООН на своїй 45-й сесії, що відбулася 4–7 березня 2014 р., заснувала Глобальну робочу групу з питань використання «великих даних» для підготовки офіційної статистики, яка мала скласти перелік уже здійснюваних заходів і прикладів використання «великих даних», розглянути питання, пов'язані з методологією, людськими ресурсами, якістю і конфі-

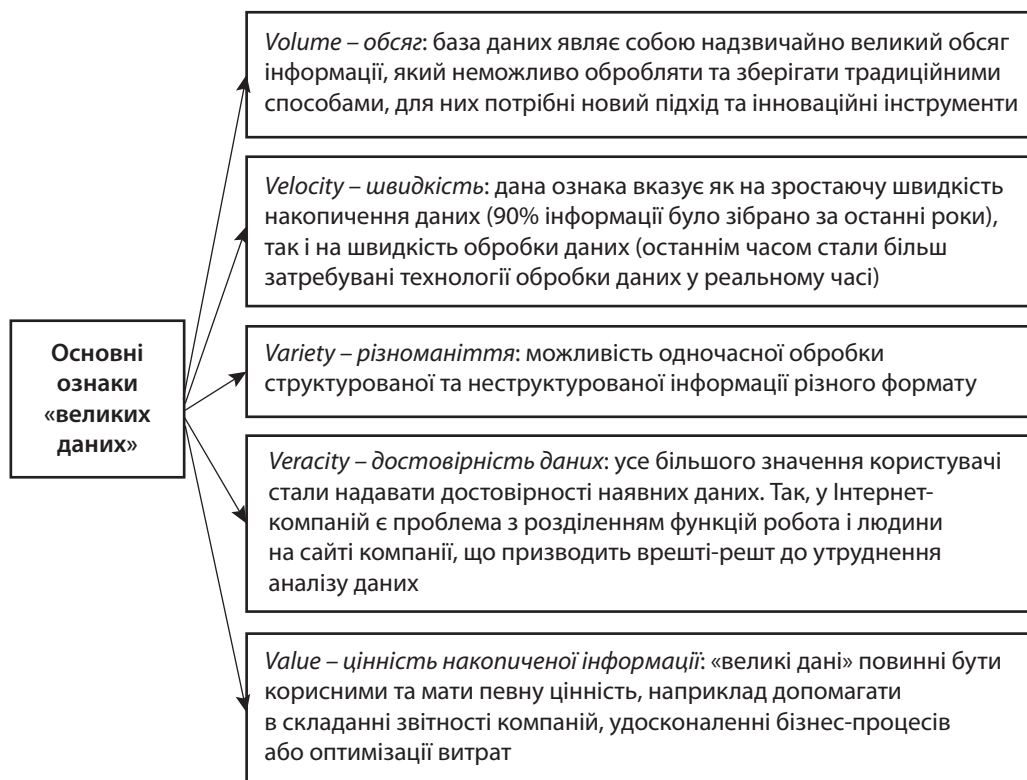


Рис. 1. Основні ознаки «великих даних»

Джерело: складено за [6–8].

денційністю, та розробити керівні принципи класифікації різних видів джерел «великих даних» [10].

У цілому джерела «великих даних», які доцільно використовувати в інформаційному забезпеченні моделі «розумних» сталих міст можна розділити на такі групи, що наведені в *табл. 1*.

Можливості контролю, які є в дослідника, а також потенційні аналітичні потужності варіюються залежно від типу джерела «великих даних». Наприклад, найчастіше в дослідника немає можливості хоч якось керувати даними, отриманими з платформ соціальних мереж, тому аналізувати текст соціальних мереж може бути досить важко. З іншого боку, для отримання адміністративних даних національних статистичних організацій може укласти партнерську угоду з відповідним державним органом і впливати на їх дизайн. Адміністративні дані більш структуровані, визначені та про них більше відомо, ніж, можливо, про всі інші джерела «великих даних».

«Великі дані» відрізняються від традиційних даних наявністю важливих характеристик. Майже кожне джерело «великих даних» має всі перераховані особливості, проте більшості властиві такі ознаки, що наведено в *табл. 2*.

Перевагами «великих даних», які обумовлюють можливість їх використання для інформаційного забезпечення розвитку «розумних» сталих міст, є такі [1]:

- ✦ **цифрове генерування:** дані створюються в цифровому вигляді (на відміну від оцифров-

ки вручну), можуть зберігатися з використанням ряду одиниць і нулів і, таким чином, можуть управлятися комп'ютерами;

- ✦ **пасивне виробництво:** дані є побічним продуктом нашого повсякденного життя або результатом взаємодії з цифровими послугами;
- ✦ **автоматичний збір:** наявність системи, яка витягає і зберігає відповідні дані по мірі їх створення;
- ✦ **географічне або часове відстеження:** наприклад, дані про місце розташування мобільного телефону або час тривалості виклику;
- ✦ **постійний аналіз:** інформація має відношення до добробуту та розвитку людей і може аналізуватися в режимі реального часу.

Використання «великих даних» в офіційній статистиці, а також для статистичного забезпечення управління розвитком «розумних» сталих міст, пов'язано з певними труднощами та проблемами.

Шляхи подолання проблем використання «великих даних» для статистичного аналізу розвитку «розумних» сталих міст представлені в *табл. 3*.

Потенційна придатність використання «великих даних» для статистичного забезпечення управління розвитком «розумних» сталих міст полягає в такому:

- ✦ на загальному рівні, належним чином проаналізувавши ці дані, можливо забезпечити «знімки» рівня життя населення з високою

Джерела «великих даних»

Група «великих даних»	Характеристика	Приклади
Адміністративні дані	Є результатом реалізації державних або інших програм	– Електронні медичні карти; – відвідування лікарень; – бази даних страхових компаній; – податкові записи
Дані транзакцій – <i>transaction-generated data</i> (TGD)	Є результатом комерційних операцій між двома суб'єктами	– Банківські операції (міжбанківські, а також особисті); – операції з кредитними картками; – купівля в супермаркетах; – інтернет-покупки
Сенсорні дані/Датчики та прилади спостереження	Є результатом запису даних датчиками та приладами спостереження	– Дорожні та транспортні датчики; – датчики клімату; – датчики обладнання та інфраструктури; – супутникові/GPS пристрої
Дані мобільного телефонного зв'язку	Є результатом комунікації між абонентами	Мобільні телефони
Інтернет-активність / Соціальні медіа / дані соціальних мереж	Є результатом фіксації запитів в он-лайн пошукових системах, перегляду веб-сторінок, коментарів у соціальних мережах	– Пошукова активність в Інтернеті; – інтернет-перегляди сторінок; – блоги та публікації, а також інший авторський та неавторизований онлайн-контент і діяльність у соціальних мережах; – аудіо/зображення/відео
Персональні дані	Є результатом фіксації даних персональних «розумних» відстежувальних пристроїв	Дані з персональних «розумних» годинників

Таблиця 2

Ознаки традиційних і «великих даних», їх відмінності

Ознака	«Великі дані»	Традиційні дані
Обсяг інформації	Від петабайт (10^{15} байт) до ексабайт (10^{18} байт)	Від гігабайт (10^9 байт) до терабайт (10^{12} байт)
Швидкість накопичення	Велика	Мала
Швидкість передачі	Велика	Мала
Ступінь структурованості	Неструктуровані та напівструктуровані дані	Структуровані
Спосіб генерування	Часто автоматично без участі людини	За участю людини
Джерела даних	Абсолютно нові	Традиційні
Взаємозв'язок даних	Слабкий	Сильний
Модель зберігання	Горизонтальна	Вертикальна
Спосіб зберігання	Децентралізований	Централізований
Наявність правил формування даних	Правила відсутні	Наявні чіткі правила формування даних

частотою, ступенем деталізації та в широкому діапазоні напрямів, зменшуючи прогалини в часі та знаннях;

- ★ у практичному аспекті аналіз цих даних може допомогти з'ясувати аномальні зміни в тому, як спільноти отримують доступ до послуг. Так звані «цифрові сигнали диму» (термін

Global Pulse) можуть служити індикаторами змін основних складових добробуту населення [9];

- ★ розуміння в реальному часі стану населення та зворотного зв'язку щодо ефективності політичних заходів має, своєю чергою, привести до більш гнучкого та адаптивного підходу

Шляхи подолання проблем використання «великих» даних для статистичного аналізу розвитку «розумних» сталих міст

Труднощі використання «великих даних»	Шляхи подолання
Перевантаження «великими даними»	Залучення фахівців певної кваліфікації, здатних впоратися з проблемами, які виникають з появою різноманітних джерел «великих даних», що дозволить уникнути затримання розвитку
Витрати по збору «великих даних» зростають швидше, ніж можливості суб'єктів щодо їх використання	– Забезпечення відповідних темпів розвитку. Немає необхідності братися за все відразу і прагнути збирати 100% інформації, що надходить з кожного нового джерела «великих» даних; – збір і вивчення зразків нових даних. За їх допомогою можна провести експериментальний аналіз, щоб визначити, що дійсно важливо в кожному джерелі та як кожен з них може бути використаний. Грунтуючись на цьому, зацікавлений суб'єкт буде готовим до проведення повномасштабного ефективного аналізу джерел «великих» даних
Ризик порушення конфіденційності, тобто забезпечення довіри громадськості та отримання її згоди на багаторазове використання даних і їх зв'язку із іншими джерелами	– Запровадження правил обробки, зберігання та застосування «великих даних», які повинні розвиватися разом з аналітичними можливостями. З розвитком сфери «великих даних» повинні розвиватися сфери самостійного і правового регулювання їх використання; – вжити заходів, щоб не допустити порушення конфіденційності. Організації повинні чітко пояснити, як вони будуть забезпечувати безпеку даних і як будуть їх використовувати, якщо хочуть отримати дозвіл користувачів на їх збір і аналіз; – вирішення проблеми конфіденційності, пов'язаної з «великими даними», інакше їх потенціал неможливо реалізувати повністю. Без належного обмеження «великі дані» можуть підняти таку хвилю протесту, що деякі їх джерела будуть повністю закриті; – критично важлива наявність саморегулювання. Учасники ринку повинні забезпечити саморегулювання і розробити правила, яких має дотримуватися кожен. Такі правила зазвичай більш ефективні та менш жорсткі, ніж ті, які вводяться державними органами, коли галузь не може контролювати себе самостійно
Законодавчі	Розроблення системи нормативних документів правового регулювання використання «великих даних», які на законодавчому рівні регламентуватимуть правила доступу до даних і їх використання
Кадрові	– Розроблення методів та інструментів для оцінки базових потреб у навичках використання «великих даних»; – інструктування з питань розробки модульної програми навчання спеціалістів; – координація навчальних програм, що стосуються більш поглибленої підготовки фахівців; – сприяння створенню глобальної мережі навчальних і дослідницьких установ для навчання і зміцнення потенціалу в галузі використання «великих даних» при підготовці статистичної інформації
Управлінські	– Формування національної політики і директив у галузі управління даними та їх захисту; – розроблення стратегічного та оперативного плану дій у галузі управління «великими даними» і захисту пропріетарних даних
Методологічні	– Підготовка методологічних матеріалів і проведення семінарів з питань забезпечення якості даних і застосування статистичних методів аналізу «великих даних»; – організація постійно діючих консультативних центрів на базі навчально-наукових інститутів та закладів освіти
Технологічні	Впровадження інноваційних рішень та вирішення питань, пов'язаних із інформаційними технологіями

до міського розвитку та, зрештою, до більшої сталості та отримання кращих результатів.

ВИСНОВКИ

«Великі дані» для статистичного забезпечення управління розвитком «розумних» сталих міст – це перетворення складних, недосконалих, часто не-

структурованих даних у діючу інформацію, що передбачає використання потужних аналітичних інструментів (таких, як машинне навчання), які розробляються в різноманітних галузях для виявлення тенденцій і взаємозв'язків усередині та між великими наборами даних. ■

ЛІТЕРАТУРА

1. Big Data for Development: Challenges & Opportunities // UN Global Pulse. URL: <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulse-June2012.pdf> (дата звернення: 27.05.2018).

2. Big Data: The Next Frontier for Innovation, Competition, and Productivity // McKinsey Global Institute, 2011, May. URL: https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx (дата звернення: 27.05.2018).

3. CEO Advisory: «Big Data» Equals Big Opportunity. URL: <https://www.gartner.com/doc/1614215/ceo-advisory-big-data-equals> (дата звернення: 27.05.2018).

4. Fundamental Principles of Official Statistics (A/RES/68/261 from 29 January 2014) URL: <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx> (дата звернення: 27.05.2018).

5. Gartner IT Glossary: Big Data. URL: <https://www.gartner.com/it-glossary/big-data/> (дата звернення: 27.05.2018).

6. IBM – Smarter Planet – United States. 27-Aug-2013 // IBM. URL: http://www.ibm.com/smarterplanet/us/en/?ca=v_smarterplanet (дата звернення: 27.05.2018).

7. Jones S. Why «Big Data» is the fourth factor of production, Financial Times. Retrieved 27 December 2012 // Financial Times. 2012. URL: www.ft.com/intl/cms/s/0/5086d700-504a-11e2-9b66-00144feab49a.html (дата звернення: 27.05.2018).

8. Laney D. 3D Data Management: Controlling Data Volume, Velocity, and Variety // META Group. 2001. URL: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (дата звернення: 27.05.2018).

9. UN Global Pulse. Tweets About Global Development Topics // Global Pulse. URL: <http://post2015.unglobalpulse.net/#> (дата звернення: 27.05.2018).

10. 45th Session of the United Nations (2014) // United Nations Statistical Commission. URL: <https://unstats.un.org/unsd/statcom/45th-session/> (дата звернення: 27.05.2018).

REFERENCES

“Big Data for Development: Challenges & Opportunities” UN Global Pulse. <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>

“Big Data: The Next Frontier for Innovation, Competition, and Productivity” McKinsey Global Institute, 2011, May. https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx

“CEO Advisory: «Big Data» Equals Big Opportunity”. <https://www.gartner.com/doc/1614215/ceo-advisory-big-data-equals>

“Fundamental Principles of Official Statistics (A/RES/68/261 from 29 January 2014)”. <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>

“Gartner IT Glossary: Big Data”. <https://www.gartner.com/it-glossary/big-data/>

“IBM – Smarter Planet – United States. 27-Aug-2013”. IBM. http://www.ibm.com/smarterplanet/us/en/?ca=v_smarterplanet

Jones, S. “Why «Big Data» is the fourth factor of production, Financial Times. Retrieved 27 December 2012”. Financial Times. 2012. www.ft.com/intl/cms/s/0/5086d700-504a-11e2-9b66-00144feab49a.html

Laney, D. “3D Data Management: Controlling Data Volume, Velocity, and Variety” META Group. 2001. <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

“45th Session of the United Nations (2014)” United Nations Statistical Commission. <https://unstats.un.org/unsd/statcom/45th-session/>

“UN Global Pulse. Tweets About Global Development Topics” Global Pulse. <http://post2015.unglobalpulse.net/#>