

Полтораков О.Ю.
кандидат політичних наук,
радник при дирекції
Національний інститут стратегічних досліджень
Щиголь І.В.
старший лаборант
кафедри методології та методів соціологічних досліджень
факультету соціології
Київський національний університет імені Тараса Шевченка

«ВЕЛИКІ ДАНІ» (BIGDATA) ТА ПЕРСПЕКТИВИ ПОЛІТИЧНОЇ СОЦІОЛОГІЇ: УКРАЇНСЬКИЙ ВИМІР

Світло відкриває нам те, про що ми у темряві лише здогадувалися.
Григорій Сковорода

Безпрецедентні технологічні новації останніх десятиліть можна вважати повноцінною «інформаційною революцією», зважаючи перш за все на експоненціальне зростання можливостей збору, зберігання та обробки величезних масивів інформації, в т. ч. емпіричної. Технології Big Data («великі дані») - в їх якнайширшому розумінні - відкривають принципово нові можливості для розвитку ключових напрямків соціогуманітарного знання (до педагогіки та археології включно), а надто - суспільствознавства, в т. ч. політології та соціології.

(Вперше термін Big Data ще в 2001 р. запропонував Даг Ланей, експерт аналітичної компанії Gartner: «Великі дані - це інформація великого обсягу, що збільшується з великою швидкістю і має велику різноманітність форматів. Це цінність, яка вимагає нових підходів для поліпшення прийняття рішень, розуміння суті та оптимізації процесів». Великий резонанс щодо Big Data пішов після того, як в червні 2011 року консалтингова компанія McKinsey випустила доповідь «Великі дані: наступний рубіж в інноваціях, конкуренції та продуктивності», в якій оцінила потенційний ринок Big Data в мільярди доларів. В технологічній компанії Gartner, яка також відома поняттям «Цикл надочікувань» (Hype Cycle), що прогнозує та пояснює тенденції, пов'язані з появою нових технологій стверджують - ера великих даних добігає кінця. Але, напевно, відбувається радше своєрідний перехід до нової стадії Big Data: з «технології, що розвиваються» - до відносно усталених технологій. Великі дані стали настільки широко поширеними, що їх вже не можна назвати інновацією. (Тим більше, що дедалі загострюється певний парадокс: зростання спроможностей створювати та накопичувати «великі дані» аж ніяк не корелюється зі зростанням спроможностей їх обробляти та інтерпретувати, а тим більше - не синхронізується, перш за все через принципову неспівставність швидкостей їх зростань.) У нинішніх умовах коректно говорити вже про проблематику не тільки та не стільки суто «Великих Даних» (Big Data), скільки вже про проблематику «Відкритих Даних» (Open Data) та «Розумних Даних» (Smart Data), а зокрема про методи швидкої обробки цих даних і про ту користь, яку можна отримати з наявної інформації).

Своєрідна «Епоха Великих Даних» - це лише один аспект перетворення глобального суспільства, його еволюції, проте саме ця епоха може стати основним визначником функціонування майбутньої економіки, «мейнстрімом» розвитку таких суспільствознавчих дисциплін як соціологія та політологія. (Втім, принципові значущі властивості цієї епохи науковці - починаючи з В. Вернадського та закінчуючи Е. Тоффлером чи М. Кастельсом - виокремили ще в контексті досліджень становлення

інформаційного суспільства та інформаційної економіки.) Окремі дослідники називають епоху Великих Даних «модернізацією» економіки та соціополітичної думки в глобалізаційному руслі. Однак певний відхід від «мейнстріму» досліджень глобалізації в її геоекономічних та суміжних аспектах, дозволяє розглядати та ставити перспективні акценти також на соціально-політичних та соціально-психологічних специфіках розвитку та самоосмислення людства.

Окремого згадування вартий також практичний досвід використання «Великих даних» в контекстах «соціально-політичної інженерії». Показовим, зокрема, є їх внесок у перемогу Б. Обама на президентських виборах в США 2012 р. Адже аналітики виборчого штабу Б. Обама активно використовували саме ресурси Big Data для аналізу настроїв виборців та коригування програми кандидата, і саме це багато в чому обумовило беззаперечну перемогу останнього.

В таких умовах нового звучання набуває загальна проблематика гносеологічного плану - починаючи від концептуальних підходів до загальної методології та конкретних методів пошуку, зберігання та обробки великих обсягів «сирої» інформації (переважно емпіричного та/або «кількісного» характеру), яка є або може бути/стати відносно новим джерелом отримання політологічно та соціологічно значущих «якісних» знань. Адже «великі дані» (BigData) пред'являють, відповідно, і «великі критерії» до роботи з ними - вже у їх якості «розумних даних» (SmartData), тобто перш за все у площині їх соціальних та політичних сенсів/смислів - і відповідних соціологічних та політологічних інтерпретацій. Подібний перехід полягає у зміні співвідношення між конкретними методами збору даних та загальною методологією їх аналізу в контексті загальної зміни підходів до самого аналізу. Характерна, зокрема, поява неологізму «libratory», утворюваного від слів «library» (бібліотека) та «laboratory» (лабораторія), - як відображення змін, які стосуються уявлень про те, що можна вважати результатом дослідження.

По-перше, виникає потреба в нових техніках аналізу, орієнтованих на все більш загальніші великі, а не вузькі вибіркові масиви даних. («...вибіркове дослідження - інструмент, який не стоїть «по-за історією», - констатують дослідники як М. Савідж та Р. Бероуз. - Його слава залишилася у минулому». Вони це пояснюють перш за все тим, що «у нинішній ситуації, коли дані по всій генеральній сукупності регулярно поступають у якості побічного продукту інституційних транзакцій, вибіркове дослідження вбачається дуже вбогим інструментом».) Через це актуалізується запит на відносно новий методологічний інструментарій оцінки надійності зв'язків, що мінімізують ризики артикуляції випадкових закономірностей, які неминуче виникають при застосуванні традиційних підходів до аналізу масивів безпрецедентних обсягів. (Разом із тим не можна не обійти увагою ту обставину, що навіть номінально «генеральна» сукупність є фактично лише хоча й дуже якісною, але лише вибіркою, яка спостерігається, із усіх можливих реалізації процесу, що генерує дані. Це, відповідно, означає, що вибіркові методи статистичного аналізу слід застосовувати також при дослідженнях генеральних сукупностей.)

По-друге, актуалізується потреба в нових технологіях підготовки даних до соціологічного аналізу - сучасному соціологу все частіше стають необхідними навички прикладного програмування - роботи, наприклад, з високорівневою мовою програмування Python, у програмному середовищі R (або ROOT, якщо доводиться оперувати дійсно величезними обсягами даних), із статистичним пакетом SPSS або Statistica тощо. Адже базові дані (ті ж масиви первинної статистики), які знаходяться в розпорядженні дослідника-суспільствознавця, зазвичай не є а ні однорідними (принаймні в плані їх форматів та якості), а ні належно структурованими. (За експертними оцінками, зараз щодня генерується близько 7 тисяч петабайт нових

даних, з яких структурованими є лише близько 10%, причому ця частка постійно зменшується.) Необхідні досліднику «великі дані» зазвичай знаходяться у форматах, що потребують «перекладу» в придатний для машинного аналізу вигляд, а також систематичної (автоматизованої) роботи по пошуку та видаленню/виправленню помилок.

Додатковими проблемами є питання доступу як до вже наявних масивів даних, так і до технологічних (обчислювальних) ресурсів для їх глибокої обробки, а також до синхронізації доступних досліднику масивів даних, пов'язаних між собою.

У разі вирішення окресленого кола проблем - що, втім, вбачається радше технічним, аніж концептуальним завданням - перед такими соціологічними дисциплінами як політична (електоральна) соціологія чи соціологія еліт тощо відкриваються нові дослідницькі горизонти та аналітичні перспективи.

В якості прикладу можна взяти соціологію політичних еліт сучасної України. Залишаючись в системно-функціональних рамках стратифікаційної теорії та альтиметричного підходу (Дж. Сарторі), можна стверджувати, що за чверть століття незалежності України членство у своєрідному «елітному клубі» її державності (розуміючи під ним перш за все обіймання найвищих посад загальнодержавного масштабу) набули, в певному наближенні, десятки сотень осіб - народних депутатів, міністрів та ін.

Тільки суто біографічні та суміжні (майнові, електоральні і т. п.) дані про цю державну та політичну (державно-політичну) еліту України наразі представляють собою достатньо великі масиви різнопланової інформації, більшість із яких знаходиться у відносно вільному відкритому доступі. Ці масиви можуть бути належним чином категоріально уніфіковані та проаналізовані із застосуванням методології кількісної обробки великих обсягів «сирої», відносно розрізненої інформації - тобто із використанням перш за все статистичних і суміжних методів її обробки та соціологічної інтерпретації. (Додатково можна залучити, зокрема, методи репутаційного аналізу - перш за все для залучення до перспективних вибірок тих осіб, які формально не входять до офіційної системи державної влади, проте мають значний політичний вплив на прийняття державно-політичних рішень та здійснення державної політики.)

Результатом може стати, зокрема, своєрідний відносно цілісний, динамічний та багатовимірний «відкритий» соціологічний портрет державної та політичної еліти України в широкій контекстній системі її політико-соціальних та інших зв'язків з українським суспільством. Також це відкриває похідні можливості для проведення суміжних і компаративних досліджень в загальному контексті політичної соціології, зокрема електоральної її складової, а також соціології еліт тощо.

Література

1. Головаха Є.І. Концептуальні й організаційно-методичні засади створення «Українського соціологічного архіву і банку даних соціальних досліджень» // Соціологія: теорія, методи, маркетинг. - 2000. - № 1. - С. 138-151.
2. Горбачик А.П. Архіви соціологічних даних: цілі існування, форми роботи, проблеми створення // Соціологія: теорія, методи, маркетинг. - 2000. - № 3. - С. 130-144.
3. Горбачик О.А. Комп'ютерна обробка текстів у якісних та кількісних соціологічних дослідженнях // Соціологія: теорія, методи, маркетинг. - 2006. - № 1.- С. 124-133.
4. Жукова Т.И. На пути к объединению интеллектуальных ресурсов мира : современный этап развития архивов социологических данных // Социальное

мышление и деятельность : влияние новых интеллектуальных технологий : Труды Института системного анализа Российской академии наук. - М. : Эдиториал УРСС, 2004. - 464 с.

5. Кислова О.Н. Быть или не быть цифровой социологии? // Вісник Харківського національного університету ім. В.Н. Каразіна «Соціологічні дослідження сучасного суспільства: методологія, теорія, методи». - 2013. - № 10(45). - С. 9-15.

6. Кислова О.Н. Социология в контексте вызова «больших данных»: роль интеллектуального анализа данных в становлении новых подходов к социологическим исследованиям // Вісник Харківського національного університету ім. В.Н. Каразіна «Соціологічні дослідження сучасного суспільства: методологія, теорія, методи». - 2015. - № 11(48). - С. 57-62.

7. Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим - М.: Манн, Иванов и Фербер, 2014. - 240 с.

8. Тэйлор Л. К социологии данных [Электронный ресурс] // Webscience. - 20.03.2013. - Режим доступа: <http://webscience.ru/details/k-sociologii-dannyh>

9. Ansolabehere S., Hersch E. Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate // Polit Anal. - 2012. - № 20(4). - 437-459 pp.

10. Buchholts S., Bukowski M., Sniegowski A. Big & Open Data in Europe: A growth engine or a missed opportunity? // Report commissioned by demosEUROPA - Centre for European Strategy Foundation within the «Innovation and entrepreneurship» programme. - Warsa: Mdruk, 2014. - 116 p.

11. Clark W.R., Golder M. Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science? // PS: Political Science & Politics. - 2015. - № 48. - 65-70 pp.

12. Grimmer J. We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together // PS: Political Science & Politics. - 2015. - № 48. - 80-83 pp.

13. Lynch C. Big data: How do your data grow? // Nature. - 2008. - № 455. - 28-29 pp.

14. Monroe B.L. The Five Vs of Big Data Political Science Introduction to the Virtual Issue on Big Data in Political Science // Political Analysis. - 2011. - № 19. - 66-86 pp.

15. Savage M., Burrows R. After the crisis? Big Data and the methodological challenges of empirical sociology [Electronic resource] // Big Data & Society. - 2014. - №

1. - 1-6 pp. - Режим доступа: <http://bds.sagepub.com/content/1/1/2053951714540280.full>

16. Tinati R., Halford S., Carr L., Pope C. Big Data: Methodological Challenges and Approaches for Sociological Analysis [Electronic resource] // Sociology. - 2014. - Vol. 48 (4). - 663-681 pp.