

dr hab. Krystyna Leśniak-Moczuk, prof. UR¹

Zakład Socjologii Zbiorowości Terytorialnych i Wirtualnych
Uniwersytet Rzeszowski

Spółeczeństwo równości czy zniewolone danetyzacją?

WSTĘP

Hasłem corocznych konferencji są nierówności ujmowane w różnych kontekstach. Z kolei niniejszy artykuł traktuje o równości, ponieważ w obliczu występowania nierówności społecznych i ekonomicznych można wskazać niszę równości, nadchodzącą za sprawą postępu technicznego w dziedzinie nośników informacji i środków przekazu. Nie jest to prorocstwo, gdyż w społeczeństwie informacyjnym doświadczamy nowych zjawisk w zakresie wolności informacji. Chociaż staramy się zaczarowywać, zaklinać rzeczywistość, doświadczamy dysonansu poznawczego lub buntujemy się przeciw nowościom.

Żyjemy w świecie ustawicznych zmian, a od zarania dziejów występował lęk, obawa przed nowościami. Ludy pierwotne buntowały się przeciw wprowadzeniu pisma w obawie, że pamięć ludzka nie będzie się rozwijać, ponieważ zapis wyeliminuje konieczność ćwiczenia pamięci w zapamiętywaniu treści przekazu. Bunt przeciwko czczonej drukarskiej wynikał z obawy o zanik autorytetu uczonego posiadającego manuskrypt, do którego tylko on miał dostęp. Przy braku powszechności źródeł pisanych nauczano dlatego, by mieć adwersarzy do dyskusji. Jeśli jest dostęp powszechny do wiedzy, to traci sens nauczanie instytucjonalne, każdy może to robić na własną rękę (vide szkoły domowe dla dzieci). Pokolenie X buntuje się przeciw zatopieniu pokolenia Y i Z w Internecie, wskazując czarne listy zagrożeń. Tymczasem zmiany są nieuchronne. Jedynie typ i tempo postępu technicznego jest w stanie ukierunkować je w inny sposób, ale nie można zmianom zapobiec, ani odwrócić ich biegu według własnych koncepcji. Powinniśmy zatem dążyć do naukowego wyjaśnienia procesów przemian.

Zrozumienie otaczającej rzeczywistości wymaga prowadzenia badań naukowych, które tym różnią się od wiedzy potocznej, że przy gromadzeniu danych wymagana jest metodologia naukowa. Wiedza oznaczająca kiedyś zrozumienie

¹ Adres korespondencyjny: Uniwersytet Rzeszowski, Al. Rejtana 16 C, 35-959 Rzeszów; e-mail: kles@autograf.pl.

przeszłości będzie umiejętnością przewidywania przyszłości [Mayer-Schonberger, Cukier, 2014, s. 249]. Ulegają więc zmianie nie tylko funkcje wiedzy, ale muszą także zostać zmienione zasady wartościowania wiedzy. Nieaktualne jest na przykład przeświadczenie, że lepiej powinni być wynagradzani ludzie posiadający specjalistyczną wiedzę, o eksperckich kompetencjach, niż osoby znające się pobieżnie na wszystkim. Jednak doświadczenie, tak jak dokładność sprawdza się w świecie małych zbiorów danych, gdzie kluczową wiedzę zdobywa się dzięki długotrwałemu gromadzeniu uspiętej wiedzy (nabytej nie z książek), której nie jesteśmy czasem świadomi, pozwalającej podejmować mądre decyzje. Błyskotliwość nie zależy od danych, ale posiadacze dostępu do danych, potrafiący analizować ich zbiory wyzbędą się uprzedzeń i konwencjonalnego myślenia z przeszłości nie dlatego, że są inteligentniejsi, ale że dysponują danymi. Poza tym są oni bezstronni wobec sporów w danej dziedzinie, które mogą mieć wpływ na ekspertów. Zmienia się to, co firmy cenią u pracowników – zmienia się to, co trzeba wiedzieć, zmieniają się ci, których trzeba znać, zmienia się to, czego trzeba się nauczyć. Zmianie podlegać musi także proces metodologii naukowej stworzonej w epoce społeczeństw tradycyjnych, dostosowanej do ówczesnych możliwości technicznych i operacyjnych zbierania i przetwarzania danych. W epoce społeczeństwa informacyjnego nabierają znaczenia umiejętności korzystania z dobrodziejstw dużych zbiorów danych (*Big Data*), a nie tylko specjalistyczna i głęboka wiedza wywodząca się niejednokrotnie z ograniczonego dostępu do niezbędnych zasobów danych [Mayer-Schonberger, Cukier, 2014, s. 186–187]. Już Francis Bacon wiedział, że siłę stanowi wiedza będąca podstawą władzy. Dążenie do poszerzania wiedzy i nierozłącznej z nią dokładności zaczęło się w Europie w XIII wieku, kiedy to astronomowie mierzyli rzeczywistość kwantyfikując czas i przestrzeń. W XIX wieku we Francji jako najbardziej zaawansowanym naukowo państwie powstał system precyzyjnie zdefiniowanych jednostek do mierzenia czasu i przestrzeni. Po odkryciu mechaniki kwantowej, niweczącej dokładność mierzenia świata, czyli pół wieku później poza grupą fizyków nadal dominowało wśród naukowców i inżynierów dążenie do bezbłędnego mierzenia, a nawet w kołach biznesowych zostało nasilane, gdy matematyka i statystyka wpływała na gospodarkę. W XXI wieku następuje redefinicja czasu i przestrzeni i zmniejszenie ich znaczenia w przepływie informacji [Mayer-Schonberger, Cukier, 2014, s. 52]. Nastaje era *Big Data*, pojawiają się bowiem techniczne możliwości gromadzenia, przetwarzania i przechowywania zbiorów danych o gigantycznych rozmiarach, w których z jakości przechodzi się w ilość (metaforycznie od – większy może więcej do – większy pokonuje lepszego). Dane dla społeczeństwa informacyjnego są traktowane tak, jak materiałny wsad, surowce, czy paliwo napędzające innowacje w społeczeństwie industrialnym. Bez obfitego ich dostarczania zadławi się kreatywność i produktywność ponowoczesnego świata. Potrzebne jest także zabezpieczenie konkurencyjności w obszarze informacji, by nie powstał „baroni danych”, będący odpowiednikami monopolistów dominujących w danych branżach [Mayer-Schonberger, Cukier, 2014, s. 239].

Biorąc pod uwagę ilościowe i jakościowe zmiany w dostępie i analizie danych zasadne staje się postawienie następujących pytań. Kiedy posiadziemy zdolność skwantyfikowania w postaci cyfrowej wszystkiego, co jest w nas i nas otacza, stworzymy narzędzie magazynowania i przetwarzania gigantycznej liczby danych, na podstawie tych korelacji wyjaśnimy co jest, a co nie jest przyczyną, dzięki temu przewidzimy co może się wydarzyć, będziemy mieć nieograniczoną wolność dostępu i tworzenia informacji, to czy staniemy się niewolnikami swojej wolności, pozbawieni wolnej woli, czy intuicja, kreatywność, dziwactwa, szczęśliwy traf zachowa w nas człowieczeństwo? Czy jeśli będziemy się zachowywać zgodnie z prognozami naukowymi, wyznaczającymi trajektorie życiowe, to tracimy spontaniczność, swobodę i przestaniemy czerpać radość z przygody? Ale czy zatem dzięki przewidywalności będziemy mogli mieć poczucie sprawiedliwości dziejowej, jeśli nasze doświadczenia nie będą wynikiem indywidualnych wyborów, lecz świat zostanie ukształtowany według przewidywalnych wzorów? Czy nie jest to namiastka zapowiedzi powrotu do znanego pokoleniu X systemu równości w zamian za zniewolenie?

BIG DATA

Zmiana skali liczby danych, rosnąc w tempie przekraczającym ludzką wyobraźnię, kumulowanych w elektronicznych nośnikach doprowadziła do zmiany ich jakości, co powoduje zmianę ich statusu. Pojawił się niejednoznaczny termin *Big Data* (łac. *dane, podane*) oznaczający wykorzystanie dużej skali danych do uzyskania nowej wiedzy lub stworzenia nowych wartości wpływających na zmiany myślenia, pracy i życia, pojmowania i eksplorowania świata. Podejście to jest początkiem wielkiej transformacji pozwalającej porzucić poszukiwanie przyczynowości (dlaczego?) na rzecz korelacji (co?) w celu rozumienia świata. Jest to potwierdzenie teorii koincydencji Grzegorza Kołodko [2009, s. 40] „rzeczy dzieją się tak jak się dzieją, ponieważ wiele rzeczy dzieje się naraz”.

Korelacje (siła korelacji mówi o tym, że kiedy wartości jednych danych się zmieniają, to istnieje prawdopodobieństwo, że wartości drugich danych też się zmienią) używane w zbiorach małych liczb nie dają takich rezultatów o statystycznej relacji między wartościami dwóch porównywanych danych jak w *Big Data*. Silna korelacja świadczy o prawdopodobieństwie (a nie pewności) wysokiego występowania zależności między zjawiskami. Istotą korelacji jest zatem właściwa identyfikacja odpowiednika (korelata) dla danego zjawiska, bowiem czasem tylko „ślepy trafem”, w wyniku zbiegu okoliczności dwa zjawiska mogą się zachowywać podobnie. Aby sprawdzić, czy korelat jest właściwy stawia się poparte teoriami hipotezy – abstrakcyjne idee o funkcjonowaniu zjawisk, na podstawie których gromadzono dane. Jeżeli korelat nie był właściwie dobrany, zależność między zjawiskami nie potwierdzi się, obala się hipotezę, uznaje się jej błędność. Wie-

dza powiększa się więc dzięki opartemu na hipotezach procesowi prób i błędów. Ten żmudny proces wykonalny na małej ilości danych był powolny, ponieważ na tworzenie hipotez mogą wpływać negatywnie uprzedzenia, błędna intuicja, mała wiedza. Ograniczenia te nie mają znaczenia w *Big Data*, ponieważ w zbiorze bardzo dużej liczby danych nie ma potrzeby dobierania do analizy pojedynczych korelatów i tworzenia skonkretyzowanych hipotez. Zamiast domyślenia się zależności między danymi zjawiskami poddajemy analizie korelacji duże zbiory danych i pozwalamy w tej analizie odpowiedzieć, jakie są odpowiednie korelaty [Mayer-Schonberger, Cukier, 2014, s. 76–80, 87].

W 2008 roku Chris Anderson opublikował swoją tezę „końca teorii” oznaczającą przestarzałość metod naukowych z powodu zalewu danych. Teorie w naukach empirycznych nie będą bowiem potrzebne, gdy świat będziemy objaśniać danymi bez wyjaśniających hipotez. Stawianie hipotez testowanych w realnym świecie z wykorzystaniem modelu przyczynowo-skutkowego będzie zastąpione analizą statystyczną korelacji bez posilkowania się teorią. Gdy danych jest wystarczająco dużo, mówią same za siebie. Teoria pozostanie w świecie abstrakcyjnym, gdzie będą dziedziny czysto teoretyczne (fizyka kwantowa) niemające nic wspólnego z rzeczywistością. Teza o końcu teorii jest za daleko posunięta, ponieważ sama metodologia *Big Data* opiera się na teoriach statystycznych, informatycznych i matematycznych, na podstawie których buduje się modele do stawiania skutecznych prognoz, sekcjonuje dane, wybiera narzędzia do analiz, interpretacji rezultatów. Nasze wybory wpływają na teorie – co z kolei wpływa na to co znajdujemy. Natomiast *Big Data* nie jest obciążona konwencjonalnym myśleniem i wrodzonymi uprzedzeniami ukrytymi w teoriach z określonej dziedziny. *Big Data* nie oznacza końca teorii, ale zmienia sposób nadawania sensu światu i oznacza wyzwanie dla intuicji [Mayer-Schonberger, Cukier, 2014, s. 99–101].

Odkrywanie korelatów w społecznych kontekstach to jeden ze sposobów wykorzystania technik *Big Data*. Drugim zastosowaniem jest analiza prognostyczna do przewidywania zdarzeń zanim one wystąpią. Nie wyjaśnia ona przyczyny zaistnienia problemu, ale wskazuje na jego istnienie. Aby odkryć ukrytą zależność trzeba zastosować metody statystyczne do olbrzymiej ilości danych i dzięki zaawansowanej analizie zidentyfikuje się nieliniowe związki między danymi. A analiza sieciowa pozwala wykryć wielowymiarowe zależności poprzez odwzorowanie, pomiar i obliczanie węzłów i połączeń między wszystkich ze wszystkimi (vide znajomi na Facebooku). Analizy takie dostarczą użytecznej wiedzy, która zwiększy skuteczność przewidywań, ponieważ pozwolą dostrzec społeczne relacje, technologiczne związki i inne zależności, o których nie wiedzieliśmy wcześniej, umykały naszemu zrozumieniu. Czy zadowala nas jednak badanie związków koicydencji, a nie przyczynowości? Ale z filozoficznej debaty nie wynika wprost, czy przyczynowość istnieje, bowiem gdyby wszystko było przez coś powodowane, to nie posiadalibyśmy wolności, by o czymkolwiek decydować. Nie istniałaby wolna wola, bo każda decyzja byłaby skutkiem wcześniejszej przyczy-

ny. Filozofowie przeciwstawiają przyczynowość wolnej woli. Ludzie wyjaśniają i rozumieją świat przez iluzoryczną szybką analizę przyczynowości i wolne eksperymenty odkrywające związki przyczynowo-skutkowe. *Big Data* zmienia role obu z nich. Relacje przyczynowe dostrzegamy intuicyjnie i zakładamy, że istnieją nawet tam gdzie jej nie ma. Wynika to z działania ludzkiej percepcji, wiązania związkiem przyczynowym następujących po sobie wydarzeń. Widząc wymagowane związki przyczynowe nie rozumiemy świata, bo przez skrót poznawczy intuicyjne dostrzeganie przyczynowości daje iluzję wiedzy, a nie objaśnia świata. Dostrzeganie przyczynowości jest skrótem, z którego korzystają mózgi leniwe do metodycznego, pogłębionego, wolnego myślenia (takim skrótem jest korzystanie z próby losowej).

Psycholog Daniel Kahneman, laureat Nagrody Nobla 2002 z ekonomii, wykrył w mózgu dwa sposoby myślenia. Jeden szybki działający bez wysiłku, dostrzegający związki przyczynowe tam gdzie ich nie ma, potwierdzający nasze przekonania, niezbędny w dawnych czasach pozwalający przetrwać w niebezpiecznym środowisku, gdzie konieczne były błyskawiczne decyzje przy ograniczonej liczbie informacji. Drugi wolny wymagający pracy, zmuszający do dogłębnego myślenia. Ten wolny sposób myślenia zostanie poddany transformacji przez badanie korelacji w dużych zbiorach danych, które rutynowo będą podważać nasze intuicje relacji przyczynowych (z trybu szybkiego myślenia), pokazując, że nie ma statystycznego związku między skutkiem a przyczyną. Nie ma matematycznych sposobów udowodnienia przyczynowości (tak jest to możliwe dla korelacji) za pomocą standardowych równań, dlatego nawet przy wolnym i wnikliwym myśleniu trudno jest odkryć jednoznaczne związki przyczynowe. Ponieważ umysły są przyzwyczajone do ograniczonej liczby informacji, wnioskujemy na podstawie małej liczby danych, nawet gdy bierzemy pod uwagę zbyt dużo czynników. Prawdopodobieństwo wystąpienia związków przyczynowych można udowadniać za pomocą eksperymentów, obserwując kontrolną grupę odniesienia niepoddaną oddziaływaniu czynników, którym poddana jest grupa eksperymentalna. W sferze społecznej eksperymenty są nieetyczne, kosztowne i czasochłonne. Natomiast do zbadania korelacji istnieją matematyczne i statystyczne tanie sposoby. Korelacje wskazujące kierunek poszukiwań relacji przyczynowych, uświadamiają istnienie ważnych zmiennych wykorzystywanych do eksperymentów badających przyczynowość. Dlatego *Big Data* pomaga w poszukiwaniu przyczynowości wskazując prawdopodobne związki korelacyjne, na które trzeba zwrócić uwagę po wykorzystaniu metod *Big Data*. Jest ona zatem turboładowaniem do analizy nieszukającej związków przyczynowych. Nawet jeśli nie zrezygnujemy z poszukiwań przyczynowości, to nie stoi ona na piedestale źródeł poznania

Rewolucja metodologiczna w świecie danych nie zostanie spowodowana przez maszyny przetwarzające, ale przez sposób użycia dużych zbiorów danych. Idee rewolucji cyfrowej pączkujące w latach 60. XX wieku zostają w zaskakującym tempie wcielane w życie. Tempo digitalizacji informacji (przekształcanie

analogowych danych w zero-jedynkowy kod binarny przetwarzany przez komputery) rośnie zaskakująco szybko. W 2000 roku jeszcze 75% informacji było zapisanych analogowo (papier, taśmy, płyty, kasety), a w 2007 roku już tylko 7%, a w 2013 roku – 2%. Gdyby te zdigitalizowane dane zapisać w książkach, to pokryłyby one 52 warstwami powierzchnię USA, a zapisane na płytach CD ułożonych jedna na drugiej tworzyłyby 5 kolumn z ziemi do Księżyca. Olbrzymia ilość informacji nie jest jeszcze przekształcana w dane cyfrowe, choć tempo liczby informacji rośnie cztery razy szybciej niż gospodarka światowa, a moc obliczeniowa komputerów dziewięć razy szybciej [Mayer-Schonberger, Cukier, 2014, s. 84–85, 87–96].

DANETYZACJA

Drugim procesem wzrostu dostępu do dużej skali danych jest danetyzacja polegająca na zbieraniu informacji o wszystkim, nawet o czymś, czego obecnie nie bierzemy pod uwagę jako źródło danych. Jej warunkiem jest chęć do kwantyfikowania (liczenia, mierzenia, ważenia) i zapisywania informacji. Czynniki sprzyjające danetyzacji sięgają trzeciego tysiąclecia p.n.e., kiedy to w dolinie Indusu, Egipcie i Mezopotamii pojawiła się idea zapisywania informacji, i istniały one w epoce analogowej, ale były zbyt kosztowne i czasochłonne. Dopiero cyfryzacja (zapisywanie danych w zero-jedynkowy kod binarny) w epoce komputerów dająca narzędzia pomiaru i przechowywania danych uczyniła danetyzcję efektywną. Digitalizacja zbiorów analogowych poprzez skanowanie daje czytelnikom tylko cyfrowy obraz stron, którego nie można przeanalizować ani wyszukać w nim słów, zmienić wielkości i kroju liter. Dopiero po danetyzacji (przekształcenie słów w dane – rozpoznanie w cyfrowym obrazie liter, słów, zdań, akapitów) informacje cyfrowe są użyteczne dla komputerów do przetwarzania i dla algorytmów do analizowania. Danetyzacja dotyczy także relacji, doświadczeń nastrojów, czyli przetwarzania interakcji w dane. Dlatego idea danetyzacji jest istotą firm z branży mediów społecznościowych. Platformy społecznościowe dla użytkowników Internetu służą do kontaktów, a firmy wychwytyją i przekształcają w dane nienamacalne elementy naszego życia, które mogą być wykorzystane do wielu badań. Facebook danetyzuje relacje, Tweeter danetyzuje uczucia przez dzielenie się myślami i emocjami, LinkedIn danetyzuje doświadczenia zawodowe bez świadomości podmiotów, o których gromadzone są te dane. Gdy cały świat zostanie zdanetyzowany potencjalnie, to wykorzystanie informacji o namacalnych i nienamacalnych aspektach życia dla społeczeństwa, nauki i biznesu, przez narzędzia (statystyka, algorytmy) i wyposażenie (procesory, cyfrowe dyski), będzie ograniczać tylko wyobraźnia. Od początku XX wieku fizycy wiedzieli, że podstawą wszystkiego jest informacja a nie atom. Danetyzacja to nie tylko przepływy i analiza danych, ale wzbogacanie pojmowania rzeczywistości, postrzeganej

nie jako ciąg wydarzeń, ale jako uniwersum składające się głównie z informacji. W przeciwieństwie do przepływów stwarzanych przez przełomowe wynalazki z poprzednich epok – przepływ wody w akweduktach umożliwiał rozwój miast, przepływ słowa pisanego dzięki drukowi przyczynił się do rozpowszechniania idei Oświecenia, przepływ informacji w gazetach przyczynił się do powstania państw narodowych, przepływ informacji przez telefon, Internet przyczynił się do rozprzestrzenienia kontaktów interpersonalnych.

Dzięki danetyzacji można ujawnić ukrytą niedostrzegalną wartość informacji, wykorzystywanych wielokrotnie na różne sposoby do wyjaśniania i prognozowania zjawisk. Umiejętność przetworzenia wszystkich danych danego zjawiska dzięki wysokowydajnym technologiom cyfrowym zwalnia z wykorzystywania prób losowych, stosowanych od XIX wieku w erze analogowej, jako następstwa ograniczeń w przetwarzaniu informacji w postaci braku narzędzi do analizy zebranych danych. Nie dostrzegano wówczas sztuczności próby losowej, ale ograniczała ona badaczy do zawężonych lub małych ilości danych, tracąc szczegóły. Struktura próby losowej jest podobna do analogowej fotografii, która z większej odległości wygląda dobrze, a przy powiększeniu staje się nieostra. Akceptując 3-procentowy margines błędu przy ekstrapolacji próby do całej populacji gubi się dokładność i detale w podgrupach próby. W próbie losowej ważne było zapewnienie wysokiej jakości danych przy redukcji liczby błędów, ponieważ nawet pojedynczy błąd zmniejszał dokładność wyników. Decyzje podejmowane na podstawie takich badań opierają się na małej liczbie dokładnych informacji opisujących zjawiska, których przyczynę byliśmy w stanie poznać. Przetworzenie przez maszyny gigantycznych rozmiarów zbiorów zwalnia człowieka z podejmowania decyzji. Dostęp do dużego spektrum danych zwalnia też z dużej dokładności, bowiem ze wzrostem skali rośnie niedokładność. Taksonomia i indeksy sprawdzają się tylko na małych liczbach i rozpadają się przy dużej skali, dlatego są zastąpione mechanizmem mniej uporządkowanym ale bardziej elastycznym „otagowywania”. Tagi pozwalają serfować w przepastnych zasobach Internetu. Nowatorsko zaprojektowane bazy rozproszonych dużych zbiorów danych łamią stare zasady rekordów i predefiniowanych pól odzwierciedlających starannie zdefiniowaną hierarchię informacji. Wymagają w zamian za to większej mocy obliczeniowej do ich przetwarzania. Tolerancja zniekształconych, niedokładnych, niedoskonałych informacji i błędnych liczb jest mimo wszystko korzystna, bo chociaż nie znamy każdego szczegółu, to zyskujemy poznanie ogólnego trendu, kierunku rozwoju zjawiska, zwiększenie dokładności prognozy i lepsze zrozumienie fenomenu zjawisk. Dokładność nie będzie celem, a brak precyzji, bezład może być postrzegany za pozytywną cechę, a nie wadę, kiedy w zamian za poluzowanie standardów dopuszczalnych błędów można otrzymać więcej danych. Kompletny zbiór danych ma większą wartość, bo zsumowanie wszystkich, w tym nieprecyzyjnych odczytów części danych daje całościowy obraz rzeczywistości, która jest przecież obciążona błędami. Duża liczba mniej dokładnych informacji, brak porządku na

skali na rzecz zwiększenia skali, pozwala zapomnieć o idealnej precyzji i korzystniejsze jest tolerowanie błędów niż zapobieganie im. *Big Data* zmienia liczbę w coś bardziej probabilistycznego niż precyzyjnego, a korzyści z niej nie wynikają z lepszych procesorów i algorytmów, ale z większej liczby danych, bowiem prostszy model na dużej liczbie danych jest skuteczniejszy od skomplikowanego modelu na małej liczbie danych. Nieuporządkowanie nie jest cechą *Big Data*, lecz wynika z niedoskonałości narzędzi używanych do mierzenia, zapisywania i analizy danych.

Analizy danych z próby losowej ograniczały się do sprawdzenia hipotez formułowanych przed zebraniem tych danych. Stosując do analiz dane z całej populacji można odkryć zależności, których nie przewidywaliśmy i nie bralibyśmy pod uwagę. Posiadanie i potaniecie dostępnych metod do gromadzenia danych, mocy obliczeniowej, miejsca do przechowywania i narzędzi do ich analizy pozwoli zrezygnować z niedoskonałości próby losowej. Może to doprowadzić do utraty monopolu na wyjaśnianie danych empirycznych przez przedstawicieli nauk społecznych, kiedy aktywne zbieranie przez nich danych z próby losowej za pomocą kwestionariuszy zostanie zastąpione pasywnym pozyskiwaniem danych podczas codziennej działalności. Schematy przeprowadzania badań na próbie losowej okażą się za drogie i nie sprostają standardom zagwarantowanym w badaniach realizowanych na dużą skalę za pomocą *Big Data* w świecie nadmiaru informacji. Stosowanie prób losowych może całkowicie nie zaniknąć, ale będzie porównywalne do chwytania bata w erze samochodów, gdy celem będzie korzystanie z dużych zbiorów danych [Mayer-Schonberger, Cukier, 2014, s. 19–24, 28–32, 42–45, 49–72, 76, 109–118, 124–126, 131–132].

ZNIEWOLENIE WOLNOŚCIĄ

Firmy dostosowują politykę prywatności do rodzajów danych, jakie chcą wykorzystać. Przy wykorzystywaniu zbiorów dużej liczby danych, dla prywatnych użytkowników zagrożeniem przestanie być naruszenie prywatności, a zagrożeniem stanie się prawdopodobieństwo wykrycia zaistnienia jakiegoś niekorzystnego zdarzenia, za skutki którego będzie trzeba ponieść odpowiedzialność. Powstanie etyczny problem wolnej woli w obliczu dyktatury danych. Jak po powstaniu prasy drukarskiej zaistniała konieczność ochrony wolności słowa, szeroko rozpowszechnianego, tak w epoce *Big Data* trzeba pragnie uregulować nienaruszalność praw jednostki [Mayer-Schonberger, Cukier, 2014, s. 33].

Rządy państw jako pierwsze na masową skalę gromadziły dane i teraz posiadają konkurentów z prywatnych firm. Dane zbierane przez rząd powinny być udostępniane społeczeństwu dla celów prywatnych i komercyjnych (Open Government Data w USA od 2009 roku – strona data.gov) z wyłączeniem przypadków zagrażających bezpieczeństwu i naruszania prywatności. Podobnie uczynił rząd

Wielkiej Brytanii, gdzie wcześniej do wykorzystania danych potrzebna była kosztowna licencja. Tymi śladami idą rządy Australii, Brazylii, Chile, Kenii. Mark Zuckerberg, założyciel Facebooka 18 maja 2012 roku przekształcił go w spółkę publiczną. Bank Światowy i Eurostat czerpie dochody z odsprzedaży danych pochodzących od podmiotów komercyjnych. Firmy umieszczone w środku strumienia przepływu informacji mogą wykorzystać efekt skali i przechwytywać wartość ukrytą w danych [Mayer-Schonberger, Cukier, 2014, s. 155–161, 167].

Wartość danych nie zmniejsza się po ich wykorzystaniu (w przeciwieństwie do rzeczy materialnych) i mogą być przetwarzane wielokrotnie w jednym celu oraz w kontekstach, dlatego pełna wartość danych jest większa niż wartość wydobyta podczas pierwotnego użycia. Uśpiona wartość może być wyzwolona dzięki połączeniu zbioru z odmiennym zbiorem. Powtórne wykorzystanie danych może przybrać ukrytą inteligentną formę. Umiejętności wymaga przeczesywanie zasobów danych i usuwanie przestarzałych danych, ponieważ nie wszystkie w takim samym tempie i w ten sam sposób tracą wartość. Biorąc pod uwagę te aspekty powstają trudności oszacowania wartości danych sprzedawanych. Korzystne dla ich właścicieli jest określenie w umowie zapłaty jako odsetka wartości wydobytej z danych, niż ustalonej wcześniej kwoty. Z uwagi na wielokrotność wykorzystania danych i możliwość niewydobycia z nich całej opcyjnej wartości właściciel może nie udostępnić ich na wyłączność. Bardzo rzadka jest wiedza jak wydobyć cenne informacje z powszechnie dostępnych danych. Błędnie uznaje się, że wartość powszechnie dostępnych za darmo danych jest niewielka. Sukces odnoszą ci, którzy posiadają wyjątkowe umiejętności komplementarne wobec tego, co jest wszechobecne i tanie [Mayer-Schonberger, Cukier, 2014, s. 136–138, 144, 148–149, 160–162, 166].

Nasza percepcja i stworzone przez nas instytucje są przystosowane do świata z niedoborem a nie nadmiarem informacji. Jak *Big Data* wpłynie na naszą prywatność i poczucie wolności? Internet sprawił, że śledzenie jest łatwiejsze, tańsze i bardziej użyteczne, bo szpiegują nas nie tylko tajne służby, ale Tweeter wie, co nam chodzi po głowie, Facebook wie o naszych relacjach towarzyskich, operatorzy komórkowi wiedzą gdzie, z kim rozmawiamy i kto jest w naszym pobliżu. Jeśli Internet zagraża prywatności, to czy *Big Data* nie naraża nas jeszcze bardziej? Kiedy dziś czujemy, że zostaje naruszane nasze prawo do prywatności, to czy *Big Data* nie uczyni tego powszechniejszym? Firmy gromadzą o nas gigantyczne ilości informacji z wszystkich aspektów naszego życia i bez naszej wiedzy dzielą się z innymi, wykorzystują je na różne sposoby. Czy nasze dane może wykorzystać wymiar sprawiedliwości? Za pomocą prognozowania w *Big Data* można karać ludzi za przewidziane skłonności do popełnienia czynów karalnych, przed ich popełnieniem. Czy większą wagę będą mieć korzyści z profilaktyki przestępczości, czy negowanie idei wolnej woli i podważanie godności ludzkiej? Czy potrafimy tak kontrolować *Big Data*, by nie zostać poddanym kontroli?

Big Data ogranicza prywatność i zagraża wolności. Pozwala na inwigilowanie naszego życia i sprawia, że prawne środki ochrony prywatności są przestarzałe,

techniczne metody zapewnienia anonimowości przestają być skuteczne. Każdy wynalazek w dziedzinie komunikowania wymagał nowych reguł prawnych. Wynalezienie czcionki drukarskiej powodujące szeroki strumień przepływu informacji, z których ludzie korzystali i sami je tworzyli, wprowadziło cenzurę kontrolującą słowo drukowane. Uchwalono prawo autorskie, które stało się dla autorów prawnym i ekonomicznym bodźcem do tworzenia. Dopiero w XIX wieku w rosnącej liczbie państw konstytucja zagwarantowała wolność słowa. Uświadomienie siły słowa pisanego stwarza potrzebę ochrony przed niewłaściwym jego wykorzystaniem, ale nie pozwoliliśmy na cenzurę, lecz stworzyliśmy odpowiednie prawa. W przeciwieństwie do długiego okresu, umasowienia druku, nie mamy setek lat na dostosowanie się, lecz błyskawiczne tempo danetyzacji daje nam kilka lat. W tym czasie trzeba zdefiniować ideę prawa, gwarantującą wolność działania z ponoszeniem odpowiedzialności za czyny. Muszą powstać instytucje i zawody objaśniające złożone algorytmy będące podstawą odkryć dokonywanych z wykorzystaniem dużych zbiorów danych, a także wspierające osoby skrzywdzone przez *Big Data*. Przez dekady prawo do prywatności kontrolowane było przez pojedyncze osoby, decydujące jak i komu udzielić zgody na przetwarzanie własnych danych osobowych. W epoce Internetu ta idea przeradza się w formalizm „podpisz i wyraż zgodę”. A epoce *Big Data* taki mechanizm nie spełni swojej roli, bo większa część wartości pochodzi z wtórnego użycia danych, a w momencie ich zbierania nie sposób sobie wyobrazić, do czego będą wykorzystane. Prywatność nie będzie chroniona przepisami wymuszającymi uzyskanie zgody na przetwarzanie danych osobowych, lecz zapewnić ją będzie przejęcie przez ich posiadaczy odpowiedzialności za ich wykorzystywanie. Powstaną innowacyjne technologie chroniące dane w pewnych przypadkach oraz idea „zróżnicowania prywatności”, czyli celowego rozmycia danych, by kwerendy w bazach danych ujawniały tylko przybliżone rezultaty. Powstanie inny model ochrony prywatności, skoncentrowany bardziej na odpowiedzialności działań użytkowników danych, niż na uzyskaniu zgody poszczególnych osób. Firmy dostaną prawo do wykorzystywania danych osobowych w zamian za odpowiedzialność za ich używanie. Same muszą dokonać oceny wpływu na pojedyncze osoby, jakie będzie miało powtórne użycie informacji zawierających dane osobowe. Przesunięcie odpowiedzialności ze społeczeństwa na użytkowników danych jest zasadne, bo to oni wiedzą więcej niż ustawodawcy i konsumenci, do czego mogą dane wykorzystywać. Prawo nie będzie wymagało usunięcia danych osobowych po ich wykorzystaniu. Użytkownicy muszą sami wyważyć korzyści z powtórnego wykorzystania danych i ryzyko ujawnienia zbyt dużej ilości informacji. Aby zbilansować korzyści i zagrożenia ustawodawcy muszą uzależnić czas wykorzystania danych od stopnia ryzyka i wartości wyznawanych w poszczególnych państwach.

Większość, choć nie wszystkie generowane dane zawierają dane osobowe. Gdy ktoś posiada wystarczającą liczbę danych nie ukryjemy przed nim naszej tożsamości [Mayer-Schonberger, Cukier, 2014, s. 195–200, 204–207, 214, 220, 223–230, 252].

Bycie celem inwigilacji jest większym naruszeniem prywatności, ponieważ liczy się zdobycie nie tylko informacji o osobie, ale o jej związkach, relacjach z innymi. Jeśli w *Big Data* nie zachowamy prawa do kształtowania naszej przyszłości, zniszczy ona istotę człowieczeństwa – racjonalność i wolność wyboru. W świecie powszechnego prognozowania wolna wola musi pozostać nienaruszalną świętością i musimy zachować prawo do podejmowania wyborów moralnych i ponoszenia osobistej odpowiedzialności za własne czyny. Nowa klasa specjalistów algorytmików musi stworzyć algorytmy zapewniające obiektywne oceny analiz wykonanych z wykorzystaniem *Big Data*, by stworzony dzięki temu świat mniej przypadkowy nie stał się nieprzewidywalnym mechanizmem, w którym jedne niewiadome zostały zastąpione innymi [Mayer-Schonberger, Cukier, 2014, s. 252–253].

ZAKOŃCZENIE

Big Data wpłynie na sposób myślenia o przyszłości. Przed 500 laty czas był cykliczny – każdy dzień był podobny do poprzedniego, przeszłość, teraźniejszość, przyszłość zlewały się ze sobą, później czas był linearny składał się z sekwencji dni, w ciągu których wpływaliśmy na trajektorie swojego życia, ludzie przeszłość mieli za sobą, kształtowali teraźniejszość jako panowie swoje losu i oczekiwali przyszłości. *Big Data* może spowodować, że staniemy się więźniami naszych wcześniejszych działań, które mogą być wykorzystane przeciwko nam, gdy stworzymy system przewidujący nasze przyszłe zachowania, nie będziemy móc uciec od tego, co wcześniej się stało. Za pomocą algorytmów może się urzeczywistnić szekspirowska maksyma: „Przeszłość jest prologiem”. A może stanie się odwrotnie, wiedza o skutkach naszych działań pozwoli podjąć kroki by do nich nie doszło? Nic nie będzie z góry ustalone, bo możemy odpowiednio zareagować na uzyskane informacje. *Big Data* nie będzie lodowatym światem automatów i algorytmów, bo dziwactwo, błędy, przeoczenia idą w parze z kreatywnością, instynktem i genialnością. Dzięki procesom psychicznym będącym źródłem upokorzenia i niedoskonałości osiągamy sukcesy i ocieramy się o wielkość. Jeśli nauczymy się wykorzystywać nieuporządkowane dane, to uznamy, że niedokładność jest cechą człowieczeństwa, a bezład cechą otaczającego świata (vide obecne rankingi, wyścig, gloryfikowanie najlepszych). W świecie odwołującym się do danych *Big Data* może rozróżniać będą od siebie skłonność do ryzyka, przypadek, instynkt, popełnienie pomyłki. Niezbędne będzie zachowanie przestrzeni dla człowieczeństwa – dla intuicji, zdrowego rozsądku, szczęśliwego trafu. Wielkość człowieka będzie tym, czego nie ujawnią algorytmy i procesory, czego nie będzie można przekształcić w dane. Istota człowieczeństwa nie będzie „tym co jest”, ale „tym czego nie ma”.

Big Data jest zasobem i narzędziem, informuje, nie wyjaśnia, poszerza rozumienie rzeczywistości lub prowadzi na manowce. Ilość danych, które zgromadzimy i przetworzymy będzie niewielką częścią wszechświata, więc nasze przewi-

dywania będą obarczone błędem, nie będą złe, ale niekompletne. Nie unieważnia to wiedzy, jaką poznamy dzięki *Big Data* i musimy z niej korzystać z dozą pokory i człowieczeństwa [Mayer-Schonberger, Cukier, 2014, s. 254–258].

BIBLIOGRAFIA

- Kołodko G., 2009, *Wędru(ó)jący świat*, Pruszyński i S-ka, Warszawa.
Mayer-Schonberger V., Cukier K., 2014, *Big Data. Rewolucja, która zmieni nasze myślenie, pracę i życie*, przekł. M. Głatki, MT Ltd. Biznes, Warszawa.

Streszczenie

W epoce społeczeństwa informacyjnego postęp techniczny stwarza możliwości gromadzenia, przechowywania i analizowania dużych zbiorów danych (*Big Data*). Nie są one zbierane podczas prowadzenia odrębnych badań naukowych, lecz za pośrednictwem narzędzi elektronicznych, często poza świadomością ich dostarczycieli. W artykule dokonano charakterystyki nowej metodyki w naukach społecznych w postaci *Big Data*, procesu danetyzacji polegającego na zbieraniu informacji o wszystkim, nawet o czymś, co nie jest brane pod uwagę jako źródło danych. Podjęto także zagadnienie konsekwencji tej wolności informacyjnej dla jednostki żyjącej w społeczeństwie informacyjnym.

Słowa kluczowe: duże zbiory danych, danetyzacja, społeczeństwo informacyjne

Society of equality or enslaved by datafication?

Summary

In the age of information society the technological progress creates opportunities for collecting, storing and analyzing large sets of data (*Big Data*). They are not collected when conducting separate research, but through electronic tools, often beyond the awareness of their providers. The article presents the characteristics of a new methodology in the social sciences in the form of the *Big Data*, which is a datafication process involving collecting information about everything, even about something that is not taken into account as a data source. There is also an issue of consequence of such freedom of information for individuals living in the information society.

Keywords: *Big Data*, datafication, information society

JEL: O10