

**Jarosław Woźniczka**

Uniwersytet Ekonomiczny we Wrocławiu, Wydział Nauk Ekonomicznych, Katedra Zarządzania Marketingowego

jaroslaw.wozniczka@ue.wroclaw.pl

**BIG DATA W MARKETINGU: SZANSE  
I ZAGROŻENIA**

**Streszczenie:** Dzięki rozwojowi technologii informatycznych i telekomunikacyjnych, cyfryzacji danych oraz rosnącej skłonności nabywców do korzystania z internetu i urządzeń skomunikowanych z internetem obserwuje się szybki wzrost ilości danych gromadzonych w systemach informatycznych. Big data uważa się obecnie za jeden z kluczowych zasobów dysponujących tego typu danymi przedsiębiorstw i za źródło przełomu w zarządzaniu i marketingu, a szczególnie w badaniach marketingowych i analizie danych. Jednocześnie wykorzystywanie wielkich zbiorów danych rodzi różnego rodzaju wyzwania zarządcze i metodologiczne oraz uzasadnione wątpliwości etyczne.

Celem artykułu jest przedstawienie wybranych szans i zagrożeń związanych z wykorzystaniem big data w marketingu. Zawiera on przegląd współczesnej literatury światowej dotyczącej tej problematyki i syntezę wybranych wątków na ten temat.

**Słowa kluczowe:** big data, marketing, badania marketingowe, analiza danych, etyka, ochrona danych osobowych.

**Klasyfikacja JEL:** M1, M3.

**BIG DATA IN MARKETING: OPPORTUNITIES AND THREATS**

**Abstract:** Thanks to the development of information and telecommunications technologies, data digitalization and the growing tendency to use the Internet and various devices connected to the Internet, the rapid increase of the amount of data collected in IT systems can be observed. Big data is currently considered to be one of the key resources of enterprises having access to them and a breakthrough in management

and marketing, especially in marketing research and data analysis. At the same time, the use of large data sets raises various management and methodological challenges and ethical dilemmas. The aim of the paper is to present selected opportunities and threats related to the use of big data in marketing. It contains the current world literature review and a synthesis of the selected topics.

**Keywords:** big data, marketing, marketing research, data analysis, ethics, personal data protection.

## Wstęp

Według jednej z ocen era wielkich zbiorów danych rozpoczęła się wtedy, gdy koszt przechowywania danych stał się niższy od kosztu ich usuwania (Hofacker, Malthouse i Sultan, 2016, s. 89). Jej nastanie miało jednak na pewno związek z postępującym wzrostem dostępności danych, wynikającym z ich cyfryzacji i rosnących zdolności do rejestrowania aktywności użytkowników internetu, mediów społecznościowych, mobilnych urządzeń telekomunikacyjnych, innych urządzeń elektronicznych itp. (Post i Edmidson, 2014, s. 18–19; Skiera, 2016, s. 11–12). Jak się szacuje, obecnie ok. 98% wszystkich informacji jest przechowywanych w formie cyfrowej, podczas gdy jeszcze w 2000 roku udział danych cyfrowych w globalnych zasobach informacji wynosił tylko 25% (Xu, Frankwick i Ramirez, 2016, s. 1563). Nabywcy nieprzerwanie generują zarówno strukturyzowane dane transakcyjne, jak i nieustrukturyzowane dane dotyczące ich zachowań (Erevelles, Fukawa i Swayne, 2015, s. 2; Hofacker i in., 2016, s. 89). Trudno nie dostrzec, że w tych danych tkwi duży potencjał analityczny i nic dziwnego, że w ostatnich latach w coraz większym stopniu próbuje się go eksploatować, tym bardziej że szybki rozwój technologii informatycznych pozwala na budowanie coraz doskonalszych systemów *business intelligence* przetwarzających big data, identyfikujących występujące w nich związki, zależności lub wzorce zachowań i opracowujących na ich podstawie algorytmy przyszłych działań marketingowych. Wielkie zbiory danych są traktowane jako nowa forma kapitału, a ich rosnące zasoby i znaczenie prowokują do mówienia o „megatrendzie” czy „rewolucji big data” (Erevelles i in., 2015, s. 897; Mayer-Schönberger i Cukier, 2013; Bosch, 2016, s. 57). Według niektórych analityków wartość globalnego rynku big data ma wzrosnąć z ok. 18 mld dol. w 2014 roku do ok. 92 mld dol. w 2026 roku (Connick, 2017).

Korzystanie z wielkich baz danych transakcyjnych, a tym bardziej rozproszonych danych nieustrukturyzowanych, wymaga znacznych inwestycji i pokonywania wyzwań o charakterze metodologicznym, technicznym

i organizacyjnym, ale może prowadzić do pozytywnych efektów zarządczych i finansowych, o czym świadczą opisywane przykłady (np. Marr, 2016; Vriens i Kidd, 2014, s. 26–29; Bean, 2017). Entuzjazm, z jakim przynajmniej część analityków i wiele firm, np. handlowych czy ubezpieczeniowych, podchodzi do eksploracji big data, świadczy o opłacalności wspomnianych inwestycji, mimo że tkwiące w nich możliwości są dopiero rozpoznawane (Rose, 2015, s. 108–109; Smith, 2015, s. 54). Wyniki badania firm z listy Fortune 1000 w okresie 2012–2017 dowodzą, że rośnie liczba przedsiębiorstw dostrzegających korzystne skutki inwestycji w systemy analizy big data. W 2017 roku menedżerowie najwyższego szczebla ponad 80% największych amerykańskich firm ocenili tego typu inwestycje jako udane, a niemal 50% z nich wskazało ich konkretne, mierzalne efekty (Bean, 2017). Z kolei według B. Hopkinsa, wiceprezesa firmy badawczej Forrester Research, ok. 75% amerykańskich przedsiębiorstw uznaje dziś zasadnicze znaczenie big data dla swojej działalności, choć tylko mniej niż 30% z nich przyznaje, że jest w stanie wykorzystać uzyskiwaną dzięki nim wiedzę, co oznacza, że problemem są zdolności i umiejętności zaawansowanej analizy wielkich zbiorów danych (Hopkins, 2016; Connick, 2017).

Nieuchronnym ubocznym skutkiem powstawiania big data, a w jeszcze większym stopniu zarządzania nimi, jest obawa przed negatywnymi skutkami społecznymi dysponowania dostępem do nich i ich przetwarzania. Obecne prawodawstwo w tym zakresie ocenia się jako niewystarczające, a zastrzeżenia etyczne, dotyczące m.in. ryzyka naruszenia prawa do ochrony danych osobowych i wrażliwych, a w konsekwencji manipulowania nimi w sposób niezgodny z ich przeznaczeniem, a nawet na szkodę użytkowników internetu, mediów społecznościowych i urządzeń mobilnych, wydają się uzasadnione (Boyd i Crawford, 2012, s. 678; Rose, 2015, s. 166–169; Lee, James, Zhang, Chester i Williams, 2016, s. 83–86).

## 1. Big data a marketing

Termin „big data” odnosi się do dużych, różnorodnych i zmiennych zbiorów danych generowanych za pomocą nowoczesnych urządzeń informatycznych i telekomunikacyjnych, których przechowywanie, przetwarzanie i analizowanie wymaga stosowania zaawansowanej technologii i dzięki którym można uzyskać rozległą wiedzę służącą m.in. dostarczaniu wartości nabywcom, uzyskiwaniu przewagi konkurencyjnej i pomiarowi efektów prowadzonych

działań (por. Chen, Chang i Storey, 2012, s. 1166; Bosch, 2016, s. 57; Xu i in., 2016, s. 1562; Zhan, Tan, Ji, Chung i Tseng, 2017, s. 519).

Zainteresowanie sposobami, uwarunkowaniami i efektami wykorzystania big data w marketingu gwałtownie wzrasta, o czym świadczy liczba publikacji na ten temat, których szybki przyrost nastąpił po 2012 roku (por. np. bazy danych Web of Science, Scopus, JSTOR czy EBSCO). W publikacjach, które się ukazały w tym okresie, można znaleźć konceptualne propozycje teoretyczne i badawcze lub sugestie pytań i hipotez badawczych dla przyszłych badań (por. Zhan i in., 2015; Erevelles i in., 2015, s. 900–901; Wedel i Kannan, 2016, s. 97–102; Xu i in., 2016, s. 1564–1565; Hofacker i in., 2016, Saboo, Kumar i Park, 2016). W świetle współczesnej literatury przedmiotu problematykę big data w marketingu można uznać za nowo otwarte pole badań, na którym identyfikuje się luki badawcze dotyczące m.in. wpływu big data na procesy marketingowe, możliwości, uwarunkowań i skutków ich wykorzystania oraz metodologii prowadzenia badań z użyciem big data.

Posiadanie dostępu do wielkich zbiorów danych i zdolności ich analizy powszechnie uznaje się za użyteczne w procesie budowania podstaw informacyjnych dla podejmowania decyzji w zakresie marketingu i w procesie wdrażania działań marketingowych. Duża część danych tego rodzaju dotyczy zachowań nabywców, dlatego według wyników badań już obecnie stanowią one źródło pogłębionej informacji na ich temat, wykorzystywane przez coraz większą liczbę podmiotów rynkowych podejmujących przedsięwzięcia marketingowe (Hofacker i in., 2016, s. 89). Wraz z rozwojem zdolności technologicznych i analitycznych, rosnącym doświadczeniem firm badawczych i wyspecjalizowanych jednostek w przedsiębiorstwach oraz coraz większym zaawansowaniem systemów *data mining* i *business intelligence* rozszerza się zakres ich wykorzystania.

Analitikę wielkich zbiorów danych przeciwstawia się tzw. tradycyjnej analizie marketingowej, wskazując, że w przeciwieństwie do niej koncentruje się na przetwarzaniu wielkich strumieni danych napływających w czasie rzeczywistym lub niemal rzeczywistym zamiast posługiwania się określonymi miernikami pozwalającymi w sposób pośredni i z pewnym opóźnieniem oceniać poszczególne inwestycje marketingowe (Xu i in., 2016, s. 1563).

Wśród zasadniczych możliwości zastosowania analizy wielkich zbiorów danych w marketingu wymienia się m.in.:

- prognozowanie rynku,
- profilowanie docelowych nabywców,
- kwantyfikowanie potrzeb nabywców i ich motywów zakupu,
- rozpoznawanie sposobów postrzegania, zakupu i użytkowania produktów,

- identyfikowanie czynników sukcesu marki,
- projektowanie innowacji,
- wprowadzanie nowych produktów,
- optymalizację działań marketingowych i produkcyjnych,
- zarządzanie relacjami z klientem (Vriens i Kidd, 2014, s. 26-29; Bean, 2017; Kalkowski, 2016).

Zaawansowane systemy analityczne pozwalają na lepsze niż dotąd przewidywanie popytu na rynkach krajowych i lokalnych oraz w segmentach rynku wyróżnianych na podstawie kryteriów innych niż geograficzne. Dzięki temu firmy mogą efektywniej alokować swoje zasoby (Vriens i Kidd, 2014, s. 26). Wielkie zbiory danych są najczęściej analizowane w celu dokładniejszej niż dotąd identyfikacji cech nabywców, wzorców ich zachowań zakupowych i użytkowych w czasie i w przestrzeni oraz ich relacji z produktami, markami, firmami itp. Systemy analityczne służą m.in. opracowaniu modeli predykcyjnych pozwalających na przewidywanie dynamiki zachowań nabywców – całkowitej wartości klientów, wielkości i częstości ich zakupów, migracji między dostawcami itp. Za pomocą uzyskiwanej w ten sposób wiedzy próbuje się przewidywać ich przyszłe wybory zakupowe i dostosowywać do nich operacyjne działania marketingowe, osiągając wymierne efekty finansowe. Poza kreowaniem pomysłów nowych produktów, przyspieszaniem procesu innowacyjnego, zmniejszaniem ryzyka zmian cen i modyfikacją oferty rynkowej analizy big data mogą być przydatne w programach efektywnościowych, służąc optymalizacji kosztów i przychodów (Vriens i Kidd, 2014, s. 28; Bean, 2017).

Zasadniczy wymiar związku między big data a marketingiem przejawia się w ich wpływie na proces badań marketingowych. Dość szybko okazało się, że rola badań opartych na wielkich zbiorach danych przynajmniej w części obszarów zainteresowania marketingu staje się podstawowa w stosunku do badań prowadzonych innymi metodami. Tak jest m.in. z badaniem konsumpcji dóbr i usług oraz badaniem korzystania z mediów, w związku z powszechną digitalizacją tych procesów. Co ważniejsze, za sprawą big data nastąpiła zasadnicza zmiana paradygmatu badawczego, polegająca na zastąpieniu aktywnego pozyskiwania informacji w drodze bezpośrednich lub pośrednich interakcji z uczestnikami badania przez podejście pasywne, oznaczające rejestrowanie przejawów zachowań osób badanych bez ich udziału. Ten bierny (z ang. *passive*) rodzaj pomiaru jest szybszy, praktycznie nieograniczony co do skali, bardziej odporny na błędy, a przede wszystkim uważany za bardziej obiektywny i dokładny, pozwalający odkryć zjawiska, związki i zależności niedostrzegalne w badaniach prowadzonych tradycyjnymi metodami (Bosch, 2016, s. 58–59).

Równocześnie zwraca się uwagę, że wartość wyników badań marketingowych wzrasta, gdy w procesie badawczym łączy się analizę big data z innymi metodami badawczymi, przy czym wskazuje się raczej na badania antropologiczne i etnograficzne niż na tradycyjne metody gromadzenia danych jakościowych i ilościowych, jak wywiady czy ankiety (Rasmussen i Hansen, 2015; Connick, 2017). Na razie nie przewiduje się więc całkowitego zastąpienia tradycyjnych metod badawczych analityką wielkich zbiorów danych, a raczej ich koegzystencję, zawierającą potencjał synergiczny, określane jako możliwość wzbogacenia wyników badań i uzyskania dodatkowych korzyści aplikacyjnych (Krajicek, 2015, s. 8).

Według opinii analityków z firm świadczących usługi w zakresie przetwarzania big data tradycyjna segmentacja rynku i profilowanie nabywców stają się mało użyteczne w obliczu możliwości stwarzanych przez wielkie zbiory danych. Dzięki nim uzyskuje się rzeczywiste, aktualne i wolne od deklaracyjności dane o charakterze behawioralnym, mówiące o konkretnym nabywcy znacznie więcej niż segmentacyjne uogólnienia i pozwalające na stosowanie marketingu zindywidualizowanego. Można się spotkać z odważnym stwierdzeniem, że na podstawie systematycznej analizy danych na temat indywidualnych danych zakupowych lub innych danych behawioralnych, szczególnie w długim, np. kilkuletnim okresie, dysponuje się pełniejszą i obiektywniejszą wiedzą na temat nabywcy od tej, jaką on sam ma o sobie (Rose, 2015).

Dzięki wielkim zbiorom danych i ich analizie przedsiębiorstwa mogą także przyspieszać proces rozwoju produktu i obniżać jego koszty (Manyika i in., 2011; Zhan i in., 2017). W modelu opisującym zasadnicze czynniki sukcesu w procesie rozwoju produktu, za które uznaje się zdolność przyspieszania procesu innowacyjnego (z ang. *accelerated development process*), zdolność zapewnienia łączności z nabywcami w toku tego procesu (z ang. *customer connection*) i zdolność tworzenia tzw. ekosystemu innowacji w przedsiębiorstwie (z ang. *ecosystem of innovation*), big data mogą ułatwiać uzyskiwanie każdej z tych zdolności (Zhan i in., 2017, s. 522–530).

Wyniki badań dowodzą pozytywnego wpływu big data na proces rozwoju produktu. Poprzez dostęp do licznych źródeł danych w czasie rzeczywistym firmy stosujące analitykę wielkich zbiorów danych mogą w sposób ciągły obserwować zachowania nabywców i działania konkurentów, monitorować reakcje nabywców na nowe dobra i usługi oraz uzyskiwać szczegółową wiedzę na temat ich stosunku do cech produktów, ich cen czy sposobów sprzedaży (Xu i in., 2016, s. 1563). Wychodząc od generalnego podziału metod analizy informacji na tradycyjną analizę marketingową (z ang. *traditional marketing analysis*, TMA) i analizę wielkich zbiorów danych (z ang. *big data analysis*,

BDA), Xu i in. (2016, s. 1563–1564) sugerują, że korzystanie z obu rodzajów analizy pozwala na tzw. fuzję wiedzy uzyskiwanej tymi dwoma drogami (z ang. *knowledge fusion*), a w zależności od stopnia korzystania z jednego lub drugiego podejścia można mówić o czterech zasadniczych strategiach w procesie rozwoju produktu. Strategie te są przedstawione na poniższym rysunku.

Stopień wykorzystania tradycyjnej analizy marketingowej	wysoki	Strategia perfekcjonisty	Strategia pioniera
	niski	Strategia widza	Strategia odkrywcy
		niski	wysoki
		Stopień wykorzystania analizy big data	

### Fuzja wiedzy a proces rozwoju produktu

Źródło: Na podstawie: (Xu i in., 2016, s. 1564)

W przedstawionej klasyfikacji strategia widza (z ang. *bystander*) oznacza opieranie się na mało złożonej wiedzy o charakterze heurystycznym, a stosunkowo niski stopień wykorzystania zarówno tradycyjnej analizy marketingowej, jak i analizy big data prowadzi do niskiego poziomu fuzji wiedzy, który rokuje relatywnie niski stopień prawdopodobieństwa sukcesu nowego produktu. Strategia perfekcjonisty (z ang. *perfectionist*) polega na posługiwaniu się stosunkowo mało złożoną wiedzą o charakterze deklaratywnym, a opieranie się głównie na tradycyjnej analizie marketingowej umożliwia osiągnięcie średniego poziomu fuzji wiedzy i średniego poziomu sukcesu nowego produktu. Podobny wynik, jeśli chodzi o poziom fuzji wiedzy i sukces nowego produktu, przypisuje się strategii odkrywcy (z ang. *explorer*), która polega na posługiwaniu się wiedzą złożoną i zautomatyzowaną, opierającą się głównie na analizie big data. Za najskuteczniejszą uznaje się strategię pioniera (z ang. *pioneer*), która oznacza intensywne wykorzystanie zarówno tradycyjnej analizy marketingowej, jak i analizy big data, co pozwala na osiągnięcie wysokiego poziomu fuzji wiedzy i uzyskanie złożonej wiedzy o zindywidualizowanym charakterze, a w konsekwencji rokuje wysokie prawdopodobieństwo sukcesu

nowego produktu (Xu i in., 2016, s. 1564–1565). Za przykład stosowania strategii pioniera uważa się m.in. serwis internetowy Netflix, który buduje złożone algorytmy działania na podstawie kompleksowej analizy danych na temat zachowań milionów swoich użytkowników w internecie, a jednocześnie prowadzi rozległe badania marketingowe dotyczące m.in. testowania nowych produktów czy strategii cenowych, osiągając wysoki poziom sukcesu przy wprowadzaniu nowych ofert (Netflix, 2011; Xu i in., 2016, s. 1564).

## 2. Większe nie zawsze znaczy lepsze

Jak się powszechnie uważa, big data rozbudziły uzasadnione oczekiwania co do ich możliwości wykorzystania, ale spełnienie tych nadziei wydaje się trudniejsze niż zakładano (Bosch, 2016, s. 57). W dyskusji nad szansami i zagrożeniami związanymi z big data zwraca się uwagę na ryzyko nadmiernego zawierzania wynikom analizy statystycznej wielkich zbiorów danych, wynikające z częściowej niepewności danych, ich braków lub tkwiących w nich błędów. Wbrew obiegu opinii nawet dysponowanie bardzo dużą liczbą obserwacji określonych zmiennych może nie wystarczyć, aby nadać badaniu walor reprezentatywności, a tym bardziej nie zwalnia z obowiązku prowadzenia go w sposób zgodny ze standardami metodologicznymi. Co więcej, badania dowodzą, że wielkie zbiory danych rzadko są reprezentatywne, a regułą jest, że mimo licznych obserwacji nie pozwalają na bezpośrednią identyfikację wielu zmiennych, nierzadko kluczowych z punktu widzenia celu badania (Bosch, 2016, s. 60–62). Ponadto nieuważna analiza, abstrahująca od kontekstu, pochodzenia danych i przyczyn ich występowania, wraz z nadmierną swobodą interpretacji może prowadzić do efektywnych, ale nieuzasadnionych wniosków. Przykładowo: czy statystyczna analiza liczby kontaktów między użytkownikami Facebooka może stanowić wystarczającą podstawę do oceny siły łączących ich relacji? Analiza big data znacznie częściej pozwala na identyfikację współzmienności zjawisk niż na dowodzenie związków przyczynowo-skutkowych między nimi. W dodatku w internecie występuje wiele fałszywych lub zmanipulowanych informacji, np. w postaci pozytywnych lub negatywnych opinii o produktach, zdarzeniach, osobach itp., które nietłwo wyłowić i usunąć z zestawu analizowanych danych (Boyd i Crawford, 2012, s. 668–672; Leinweber, 2007, s. 19; Hofacker i in., 2016, s. 93–94; Hu, Bose, Koh i Liu, 2012, s. 674–684; Bosch, 2016, s. 60).

Stawiając pytanie, czy wielkie zbiory danych to dane lepsze (od innych), Boyd i Crawford (2012) analizują przykład serwisu społecznościowego Twitter



jako tego typu zbiór. Wpisy na Twitterze są przedmiotem wielu badań socjologicznych, opierających się nierzadko na wynikach analizy statystycznej wielu tysięcy obserwacji. Interpretując je, należy jednak brać pod uwagę liczne ograniczenia wynikające z natury tego portalu, wpływające na poprawność wnioskowania. Struktura użytkowników Twittera jest niereprezentatywna dla jakiegokolwiek populacji – ani lokalnej, ani globalnej. Pojęcia „użytkownika Twittera” i „konta na Twitterze” nie są tożsame, ponieważ niektórzy użytkownicy zakładają wiele kont, inni – żadnego, a niektóre konta są zaprogramowane. To samo dotyczy pojęcia „aktywnego uczestnika”, którego aktywność może oznaczać zarówno zamieszczenie wpisów, jak i jedynie ich śledzenie (Crawford, 2009, s. 532). Z kolei wyniki analizy częstości poruszania we wpisach określonej tematyki mogą być zniekształcone za sprawą polityki automatycznego usuwania wpisów zawierających określone słowa lub treści. Ponadto część wpisów jest zastrzeżona na życzenie użytkownika.

W kontekście sygnalizowanych ograniczeń i problemów analitycznych i interpretacyjnych, które należy uznać za słabość big data, uzasadniony wydaje się wniosek, że w przypadku określonych problemów badawczych większą wartość może mieć analiza jednostkowych przypadków (z ang. *small data*) (Veinot, 2007, s. 171). Według V. Boscha, dyrektora Departamentu Marketingu i Analizy Danych firmy badawczej GfK, istotą big data wcale nie musi być szczególnie duży rozmiar danych, w związku z czym należałoby mówić nie tyle o „wielkich zbiorach danych”, ile o „danych nowego typu” (Bosch, 2016, s. 59). Z kolei D. Krajicek, dyrektor generalny GfK Consumer Experiences North America, już w 2014 roku postulował posługiwanie się terminem „inteligentnych danych” (z ang. *smart data*), odnoszącym się do wielkich zbiorów danych przetwarzanych w sposób profesjonalny i użyteczny dla decydentów. W tym ujęciu inteligentne dane to te, których pochodzenie jest znane, które są zarządzane odpowiedzialnie, tzn. z poszanowaniem prawa do prywatności i bez nadużywania zaufania publicznego i które są analizowane z zachowaniem zasady reprezentatywności, z dbałością o istotność identyfikowanych zależności i z odpowiednim rygoryzmem w interpretacji wyników (Krajicek, 2014, s. 10).

Wyniki analizy wielkich zbiorów danych mogą wiele powiedzieć o cechach i zachowaniach nabywców lub innych użytkowników internetu, lecz znacznie mniej o ich motywach i przyczynach tych zachowań (Rasmussen i Hansen, 2015). W tym zakresie wymagają uzupełnienia o wyniki tradycyjnych badań socjologicznych lub marketingowych, choć jednocześnie można próbować wnioskować na temat potrzeb, motywów i postaw nabywców na podstawie analizy ich zachowań, i to zarówno w odniesieniu do przeszłości, jak i przyszłości (Hofacker i in., 2016, s. 90–91).

W oryginalnej definicji big data sformułowanej przez D. Boyda i K. Crawford zwraca się uwagę nie tyle na techniczno-pragmatyczny wymiar wielkich zbiorów danych, ile na sposób ich traktowania i zdolności do ich analizy. Według nich big data to zjawisko polegające na „wzajemnym oddziaływaniu na siebie technologii, analizy i mitu” (Boyd i Crawford, 2012, s. 662). Technologia oznacza w tym wypadku maksymalizację mocy obliczeniowej i precyzji algorytmów służących gromadzeniu i analizowaniu danych oraz poszukiwaniu związków występujących w ich wielkich zbiorach, analiza – korzystanie z tych zbiorów w celu identyfikacji zależności upoważniających do wyciągania wniosków o charakterze ekonomicznym, społecznym, technicznym lub prawnym, a mit – przekonanie, że na podstawie tego typu danych można uzyskać niedostępną dotąd wiedzę, umożliwiającą poznanie rzeczywistości w sposób „prawdziwy, obiektywny i dokładny” (Boyd i Crawford, 2012, s. 662).

### 3. Big data a ochrona prywatności

Korzystanie z wielkich zbiorów danych rodzi wątpliwości o charakterze etycznym. Właściciele mediów społecznościowych, przedsiębiorstwa, instytucje badawcze lub inni użytkownicy mający dostęp do wielkich zbiorów danych przetwarzają je w dowolny sposób, zwykle bez świadomej zgody osób, których te dane dotyczą (Lee i in., 2016, s. 85). Mimo że są one wykorzystywane w sposób anonimowy, to po pierwsze dzieje się to zwykle bez wiedzy np. użytkowników portali internetowych czy sieci komórkowych, a po drugie trudno mówić o gwarancji pełnej i trwałej anonimowości (Zimmer, 2008; Nunan i Di Domenico, 2013, s. 508–509).

D. Nunan i M.L. Di Domenico (2013, s. 508–510) wskazują kilka zasadniczych zagrożeń dla ochrony prywatności w związku z gromadzeniem wielkich zbiorów danych: 1) zagrożenie wynikające z możliwości identyfikacji osób na podstawie początkowo anonimowych zbiorów danych za pomocą rozwijających się nowych technologii informatycznych, 2) zagrożenie wynikające z ryzyka wykradzenia danych osobowych przez hakerów, 3) zagrożenie wynikające z automatyzacji procesów gromadzenia i analizy danych i rozwoju kontrolującej te procesy sztucznej inteligencji, 4) zagrożenie wynikające z biernego gromadzenia wszelkich danych, nie tylko tych, które są potrzebne do konkretnych celów, 5) zagrożenie wynikające z długotrwałego przechowywania danych osobowych. Ponieważ przedsiębiorstwa i instytucje gromadzące big data same nie są w wielu przypadkach w stanie stwierdzić,

kiedy i w jaki sposób mogą one być wykorzystywane, ochrona prywatności staje się iluzoryczna. Na przykład zdolność wielkich systemów informatycznych do przechowywania danych przez wiele lat, stwarzająca nieosiągalne dotąd możliwości śledzenia zmienności zjawisk i zachowań w czasie, niesie ze sobą ryzyko wykorzystywania informacji o zdarzeniach z przeszłości ze szkodą dla określonych osób, a więc i ryzyko łamania prawa do zapomnienia i bycia zapomnianym.

Big data w większości przypadków są analizowane po to, aby na podstawie uzyskanych wyników podejmować określone działania, np. o charakterze marketingowym, skierowane do tych samych osób, których dotyczą, lub oddziałujące na nich w sposób bezpośredni lub pośredni. Niezależnie od tego, jak korzystne dla nabywców, obywateli lub inaczej opisywanych członków określonej społeczności byłyby te działania, powinni oni być w pełni świadomi użycia informacji na ich temat do ich przygotowania, tym bardziej że korzyści czerpane przez wykorzystujące je firmy lub instytucje są zwykle znacznie większe. Wątpliwości etyczne rosną, gdy korzystanie z danych osobowych dotyczy tzw. wrażliwych grup odbiorców, w tym m.in. dzieci, młodzieży, osób starszych lub chorych czy różnego rodzaju mniejszości społecznych, lub polega np. na prowadzeniu atrakcyjnej, personalizowanej komunikacji marketingowej na rzecz produktów szkodliwych lub o ograniczonej dostępności, takich jak alkohol, wyroby tytoniowe, produkty farmaceutyczne, broń itp. (Lee i in., 2016, s. 84).

Jak pokazują konkretne przykłady, rejestrowane dane na temat szczegółowej historii zakupów i zachowań sklepowych poszczególnych klientów mogą służyć nie tylko do przygotowania spersonalizowanych ofert i ukierunkowanej komunikacji marketingowej, ale i do identyfikacji i oceny ich nawyków, stanu zdrowia, trybu życia, sytuacji materialnej czy cech osobowościowych. W 2012 roku ujawniono, że w amerykańskiej sieci detalicznej Target na podstawie analizy zakupów dokonywanych przez jej klientki oceniano, które z nich są w ciąży, a nawet w którym jej trymestrze, przewidując na tej podstawie terminy porodów i przysyłając im odpowiednio sprofilowane oferty. Sprawa stała się głośna najpierw w amerykańskiej prasie, a potem w mediach na całym świecie z powodu jednej z takich ofert, która trafiła do domu pewnej będącej w ciąży nastolatki, a właściwie jej nieświadomego niczego ojca (Duhigg, 2012; Hill, 2012; Rose, 2015). Na podstawie analizy danych dotyczących aktywności w sieci osób zarejestrowanych na jednym z portali randkowych identyfikowano ich rzeczywiste preferencje seksualne, poglądy polityczne czy postawy wobec mniejszości etnicznych, nierzadko sprzeczne z deklarowanymi (Rudder, 2014; Rose, 2015). Firma ubezpieczeniowa Aviva już w 2010 roku

zaczęła zlecać prowadzenie analizy danych na temat zwyczajów konsumpcyjnych swoich klientów, pochodzących m.in. z ich kart lojalnościowych, aby na tej podstawie oceniać ich kondycję fizyczną, przewidywać długość życia i szacować ryzyko ubezpieczeń zdrowotnych i ubezpieczeń na życie (Scism i Maremont, 2010; Rose, 2015). W podobnych lub innych celach mogą być wykorzystywane dane z urzędzeń dokonujących ciągłego pomiaru parametrów biometrycznych organizmu – np. ciśnienia krwi, przyswajanych składników odżywczych, poziomu cholesterolu, liczby spalanych kalorii itp. Z kolei analiza haseł wpisywanych do wyszukiwarek internetowych jest wykorzystywana do śledzenia przypadków określonych chorób, np. grypy, i do monitorowania stopnia rozwoju ewentualnej epidemii (Rose, 2015).

Przedstawione przykłady, podobnie jak wiele innych, dowodzą z jednej strony przydatności big data w eksplorowaniu indywidualnych lub grupowych zachowań z przynajmniej potencjalną korzyścią dla ich właścicieli, z drugiej zaś każą stawiać pytanie o moralną odpowiedzialność wynikającą z ingerencji w prywatność klientów lub odbiorców określonych działań z narażeniem ich na jej nieznaną konsekwencje. W czasach, w których komunikacja między ludźmi, a w ślad za nią także prywatne i zawodowe więzi społeczne są w dużym stopniu uzależnione od korzystania z internetu, mediów społecznościowych, technologii mobilnych itp., narodził się trudny dylemat: ochrona prywatności a uczestnictwo w życiu publicznym. Osoby świadome ryzyka pozostawiania w sieci informacji wrażliwych, dotyczących własnej tożsamości, kondycji zdrowotnej czy życia prywatnego i ograniczające w związku z tym swoją aktywność w internecie narażają się na częściowe lub całkowite wykluczenie z życia zawodowego lub towarzyskiego (Nunan i Di Domenico, 2013, s. 510–511).

Wszystko to powoduje konieczność wypracowania odpowiednich regulacji prawnych w tym zakresie, utrzymywania wysokich standardów etycznych w procesie prowadzenia badań i wykształcenia powszechnego poczucia odpowiedzialności za prowadzone działania (Ess, 2002; Boyd i Crawford, 2012, s. 672). Mimo że w ostatnich latach wiele uczyniono dla zwiększenia stopnia bezpieczeństwa użytkowników sieci i ochrony ich prywatności, procesy przetwarzania i wykorzystywania big data wywołują uzasadnione kontrowersje. Z punktu widzenia użytkowników wielkich zbiorów danych służą one tworzeniu wartości dodanej w postaci np. lepszych dóbr i usług, wygodniejszych sposobów komunikacji czy sprawniejszej obsługi, podczas gdy przynajmniej część odbiorców postrzega korzystanie z nich jako przejaw inwigilacji i manipulacji. W tym kontekście zwraca się też uwagę na asymetrię dostępu do big data. Mają go bowiem przede wszystkim wielkie podmioty rynkowe i pozarynkowe, do których należą tego rodzaju zbiory danych, lub te, które są w stanie

za ten dostęp zapłacić. Te same przedsiębiorstwa lub instytucje dysponują też odpowiednimi umiejętnościami i narzędziami do obróbki tego typu danych. Tym samym dostęp do big data zaczyna stanowić kolejne kryterium podziału między mniejszymi a większymi przedsiębiorstwami, agencjami badawczymi czy instytucjami, stawiając te drugie w uprzywilejowanej sytuacji (Boyd i Crawford, 2012, s. 675).

Ochrona danych osobowych jest przedmiotem nowych regulacji prawnych, takich jak uchwalony w kwietniu 2016 roku pakiet legislacyjny Unii Europejskiej i stanowiące jego integralną część Rozporządzenie Ogólne o Ochronie Danych Osobowych (RODO), które obowiązuje od maja 2018 roku (GIODO, 2016). Wydaje się, że szczególne wyzwania o charakterze etycznym stoją przed środowiskiem badaczy rynku, które powinno reagować na zmieniające się możliwości i warunki prowadzenia badań, odpowiednio dopasowując do nich swoje zasady postępowania. Podstawowe kodeksy etyczne w tej dziedzinie, takie jak Międzynarodowy kodeks postępowania w badaniach marketingowych i społecznych Europejskiego Stowarzyszenia Badaczy Opinii Publicznej i Rynku (z ang. *European Society for Opinion and Marketing Research*, ESOMAR) czy Kodeks postępowania Stowarzyszenia Badań Marketingowych (z ang. *Marketing Research Society*, MRS), wielokrotnie aktualizowane, już wcześniej zawierały zapisy chroniące prawa respondenta i regulujące procesy gromadzenia danych i ich wykorzystania. Jedne z kluczowych zasad zapisanych w Kodeksie ESOMAR mówią m.in. o nienadużywaniu zaufania respondentów, niewykorzystywaniu ich braku wiedzy lub doświadczenia, obowiązku informowania ich o celu gromadzenia danych, zbieraniu adekwatnych i nienadmiarowych danych wyłącznie w określonych celach badawczych, nieużywaniu danych niezgodnie z tymi celami, przechowywaniu danych przez czas nie dłuższy niż wymaga tego cel badania, zapewnieniu braku dostępu zleceniodawcy badania do tożsamości respondentów, zapewnieniu adekwatnych środków bezpieczeństwa do zapobieżenia nieautoryzowanemu dostępowi do danych osobowych, ujawnieniu ich i manipulowaniu nimi oraz o zapewnieniu respondentom prawa do żądania nieudostępniania innym swoich danych osobowych (ICC/ESOMAR, 2007, s. 4–7; Nunan i Di Domenico, 2013, s. 512–513).

W 2016 roku opublikowano nowy Kodeks postępowania ESOMAR, który w tytule odnosi się nie tylko do badań, ale i do analityki danych (z ang. *International Code on Market, Opinion and Social Research and Data Analytics*) i zawiera liczne zmodyfikowane lub zupełnie nowe zapisy, szczególnie w zakresie ochrony prywatności. Wśród nich już w artykule 1 nakłada się na badaczy obowiązek zapewnienia, by uczestnicy badania nie ponieśli szkody w związku wykorzystaniem ich danych osobowych, zwraca się uwagę m.in. na powinność

uzyskania zgody osób będących przedmiotem badania w przypadku biernego gromadzenia danych, a gdy jest to niemożliwe – na konieczność posiadania odpowiedniej podstawy prawnej do prowadzenia takiego badania, obowiązek zapewnienia, że podczas korzystania z danych o charakterze wtórnym zawierających dane osobowe ich właściciele nie poniosą żadnej szkody, oraz obowiązek zastosowania środków mających na celu ochronę przed takimi szkodami, a także na konieczność ochrony tożsamości respondentów badania przed jej odkryciem za pomocą dedukcji (ICC/ESOMAR, 2016, s. 8–10). Analogiczne regulacje znajdują się w Kodeksie postępowania Stowarzyszenia Badań Marketingowych z 2014 roku (*MRS*, 2014).

## Zakończenie

Marketing stanowi obszar decyzji, w którym wsparcie ze strony analizy wielkich zbiorów danych ocenia się jako przydatne i oczekiwane. Liczne duże podmioty rynkowe dysponujące dostępem do big data i odpowiednimi zdolnościami kapitałowymi, technologicznymi i ludzkimi już od co najmniej kilku lat dostarczają dowodów na biznesową użyteczność inwestycji w analitykę tego typu danych. Postęp w tej dziedzinie przebiega szybko, choć wymaga pokonywania wielu problemów metodologicznych, technicznych i organizacyjnych.

Wśród pól badawczych związanych z big data, których eksploracji można się spodziewać w przyszłości, wskazuje się m.in. identyfikację związku korzystania z internetu, mediów społecznościowych i urządzeń mobilnych z przebiegiem procesu podejmowania decyzji zakupu przez nabywców, rozpoznanie sposobów nabywania i użytkowania produktów, określenie znaczenia określonych przejawów zachowań nabywców rejestrowanych w internecie dla procesów budowania cech oferty, wyboru kanałów dystrybucji i opracowywania programów komunikacji marketingowej, rozpoznanie możliwości personalizacji ofert stwarzanych przez big data i pomiar jej efektów, określenie wpływu algorytmizacji działań marketingowych na wyniki finansowe, identyfikację czynników wpływających na procesy powstawania i dyfuzji innowacji itp.

Ważnym wątkiem badań na styku big data i marketingu wydaje się także rozpoznawanie mechanizmów i skutków ich wykorzystywania z punktu widzenia ochrony prywatności, bezpieczeństwa danych osobowych i wrażliwych oraz etycznej oceny praktyki badań rynkowych. Niezbędne wydaje się szacowanie ryzyka związanego z użyciem big data oraz posługiwanie się wynikami badań w celu zwracania uwagi na problemy z tym związane.

## Bibliografia

- Bean, R. (2017). How companies say they're using big data. *Harvard Business Review*.  
Pobrane z <https://hbr.org/2017/04>
- Bosch, V. (2016). Big data in market research: why more data does not automatically mean better information. *GfK Marketing Intelligence Review*, 8(2), 56–63.
- Boyd, D. i Crawford, K. (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication, & Society*, 15(5), 662–679.
- Chen, H., Chang, R. H. i Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS Quarterly* 36(4), 1165–1188.
- Connick, H. (2017). *Turning big data into big insights*. Pobrane z [www.ama.org/publications/MarketingNews](http://www.ama.org/publications/MarketingNews)
- Crawford, K. (2009). Following you: disciplines of listening in social media. *Continuum: Journal of Media & Cultural Studies*, 23(4), 532–33.
- Duhigg, Ch. (2012, 19 lutego). *How companies learn your secrets*. Pobrane z <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>
- Erevelles, S., Fukawa, N. i Swayne, L. (2015). Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904.
- Ess, C. (2002). *Ethical decision-making and Internet research: recommendations from the AoIR ethics working committee*. Association of Internet Researchers. Pobrane z <http://aoir.org/reports/ethics.pdf>
- GIODO. (2016). *Parlament Europejski przyjął nowy legislacyjny pakiet ochrony danych osobowych*. Pobrane z <http://www.giodo.gov.pl/pl/1520147/9228>
- Hill, K. (2012, 16 lutego). *How target figured out a teen girl was pregnant before her father did*. Pobrane z [www.forbes.com/sites/kashmirhill/2012/02/16/460204d16668](http://www.forbes.com/sites/kashmirhill/2012/02/16/460204d16668)
- Hofacker, Ch. F., Malthouse, E. C. i Sultan, F. (2016). Big data and consumer behavior: imminent opportunities. *Journal of Consumer Marketing*, 33(2), 89–97.
- Hopkins, B. (2016). *Think you want to be "data-driven"? Insight is the new data*. Pobrane z <https://go.forrester.com/blogs>
- Hu, N., Bose, I., Koh, N. S. i Liu, L. (2012). Manipulation of online reviews: an analysis of ratings, readability, and sentiments. *Decision Support Systems*, 52(3), 674–684.
- ICC/ESOMAR. (2007). *International code on market and social research*. Pobrane z <https://www.esomar.org>
- ICC/ESOMAR. (2016). *International code on market, opinion and social research and data analytics*. Pobrane z <https://www.esomar.org>
- Kalkowski, J. (2016). *Handling a flood of data*. Pobrane z <https://www.brandpackaging.com/articles/85628>
- Krajciek, D. (2014, January/February). Big data next step. Leveraging the impact of “smart data”. *Marketing Insights*, 10–11.
- Krajciek, D. (2015, September/October). Crashing the big data party. An in-depth look at marketing research 3.0. *Marketing Insights*, 8–9.

- Lee, K., James, J. P., Zhang, M., Chester, J. i Williams, J. D. (2016). Big data and policy issues. W: *Marketing and public policy conference proceedings* (s. 83–87).
- Leinweber, D. (2007). Stupid data miner tricks: overfitting the S&P 500. *The Journal of Investing*, 16(1), 15–22.
- Manyika, J. R., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. i Hung Byers, A. (2011). *Big data: the next frontier for innovation, competition, and productivity*. McKinsey Global Institute. Pobrane z [https://www.mckinsey.com/file:///C:/Users/User/Downloads/MGI\\_big\\_data\\_exec\\_summary.pdf](https://www.mckinsey.com/file:///C:/Users/User/Downloads/MGI_big_data_exec_summary.pdf)
- Marr, B. (2016). *Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results*. New York: Wiley.
- Mayer-Schönberger, V. i Cukier, K. (2013). *Big data: a revolution that will transform how we live, work and think*. New York: Houghton-Mifflin Harcourt.
- MRS code of conduct. *Celebrating sixty years of successful self-regulation*. (2014). Pobrane z <https://www.mrs.org.uk>
- Netflix. (2011). *How we determine product success*. Pobrane z <https://medium.com/netflix-techblog/980f81f0047e>
- Nunan, D. i Di Domenico, M. L. (2013). Market research and the ethics of big data. *International Journal of Market Research*, 55(4), 505–520.
- Post, R. i Edmidson, D. (2014). Challenging big data preconceptions: new ways of thinking about data and integrated marketing communication. *International Journal of Integrated Marketing Communications*, 6(1), 18–24.
- Rasmussen, M. B. i Hansen, A. W. (2015). *Big data is only the half the data marketers need*. *Harvard business review*. Pobrane z <https://hbr.org/2015/11>
- Rose, C. (2015, March). *Big data gets personal*. Pobrane z [www.cincinnati.com](http://www.cincinnati.com)
- Rudder, Ch. (2014). *Dataclysm: who we are (when we think no one's looking)*. New York: Crown House Publishing.
- Saboo, A. R., Kumar, V. i Park, I. (2016). Using big data to model time-varying effects for marketing resource (re)allocation. *MIS Quarterly*, 40(4), 911–939.
- Scism, L. i Maremont, M. (2010). *Insurers test data profiles to identify risky clients*. Pobrane z [www.wsj.com/articles](http://www.wsj.com/articles)
- Skiera, B. (2016). Data, data and even more data: harvesting insights from the data jungle. *GfK-Marketing Intelligence Review*, 8(2), 10–17.
- Smith, K. (2015, November). Big data discoveries: insurers are just scratching the surface of what's possible with big data. *Best's Review*, 53–56. Pobrane z [www.highbeam.com/doc/1G1-433879430.html](http://www.highbeam.com/doc/1G1-433879430.html)
- Veinot, T. (2007). The eyes of the power company: workplace information practices of a vault inspector. *The Library Quarterly*, 77(2), 157–180.
- Vriens, M. i Kidd, P. (2014, November/December). The big data shift: what every marketer needs to know about advanced analytics. *Marketing Insights*, 23–29.
- Wedel, M. i Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97–121.



- 
- Xu, Z., Frankwick, G. L. i Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: a knowledge fusion perspective. *Journal of Business Research*, 69(5), 1562–1566.
- Zhan, Y., Tan, K., Ji, G., Chung, L. i Tseng, M. L. (2017). A big data framework for facilitating product innovation processes. *Business Process Management Journal*, 23(3), 518–536.
- Zimmer, M. (2008). *More on the “anonymity” of the Facebook dataset – it’s Harvard College*. Pobrane z [www.michaelzimmer.org](http://www.michaelzimmer.org).