

Jędrzej Wieczorkowski

Big Data : aspekt technologiczny i ekonomiczny vs. aspekt społeczny

Ekonomiczne Problemy Usług nr 113, 399-408

2014

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

JĘDRZEJ WIECZORKOWSKI

Szkoła Główna Handlowa w Warszawie

BIG DATA – ASPEKT TECHNOLOGICZNY I EKONOMICZNY VS. ASPEKT SPOŁECZNY

Streszczenie

Termin *big data* jest pojęciem nowym i dopiero kształtującym się. W pracy zaproponowano wyróżnienie trzech podstawowych aspektów tego terminu: technologiczny (obejmujący poza technologią także metody analityczne), ekonomiczny (obejmujący różnorodne zastosowania) oraz społeczny. Ten ostatni, dotyczący głównie zagrożenia prywatności, jest aspektem wtórnym do poprzednich, ale to jemu poświęcono w artykule najwięcej uwagi.

Słowa kluczowe: *big data*, analiza danych, dane osobowe

Wprowadzenie

Pojęcie *big data* jest obecnie często wykorzystywane w bardzo różnym kontekście. Dotyczy ono nie tylko technologii informatycznych, metod ilościowych oraz ich zastosowań, lecz także zagadnień społecznych. Mimo że termin jest prawie nowy, o jego popularności świadczy fakt, że chętnie jest wykorzystywany i dyskutowany co najmniej na poziomach: akademickim, komercyjnym oraz politycznym. Celem artykułu jest próba systematyzacji pojęcia *big data* w różnych jego aspektach, ze szczególnym naciskiem na zagadnienia społeczne

1. Zrozumieć *big data*

Podstawową kwestią jest próba zrozumienia zjawiska *big data*. Na dziś trudno o jedną powszechnie akceptowaną definicję. Być może pojęcie jest na tyle nowe,

wieloaspektowe i wciąż intensywnie rozwijające się, że nie można go (jeszcze?) jednoznacznie zdefiniować. Tradycyjnie próbując scharakteryzować *big data*, posługuję się „kilkoma v”, czyli cechami opisującymi pojęcie. Są to co najmniej 3 podstawowe „v”: *volume*, *velocity*, *variety*, często jednak uzupełnia się cechy o kolejne „v”, przede wszystkim: *value*, *veracity*.

META Group (2001) – firma doradcza przejęta w następnych latach przez Gartner Inc. – opublikowała raport, w którym scharakteryzowano ówczesny wpływ e-commerce, globalizacji i innych trendów w gospodarce, na rozwój technologii informatycznych. Przewidywane zmiany w IT przedstawiono w modelu 3V w postaci trzech wymiarów. Nie użyto wtedy jeszcze pojęcia *big data*, lecz prognoza dobrze pasuje do rozwoju tego zjawiska, które miało się pojawić kilka lat później.

Zgodnie z modelem 3V, jako charakterystyczne dla *big data* zalicza się wymienione poniżej cechy: a) bardzo duży wolumen przetwarzanych danych (*volume*). Związany jest on przede wszystkim z coraz szybszym przyrostem ilości dostępnych w postaci cyfrowej danych, które mogą potencjalnie zostać wykorzystane; b) zmienność i dynamikę przetwarzanych danych (*velocity*). Dotyczy to w szczególności zagadnienia danych szybkozmiennych i generowanych w bardzo krótkich odstępach czasu, w szczególności takich których wykorzystanie powinno nastąpić wkrótce po ich pojawieniu lub aktualizacji. Cecha ta jest typowa dla danych sensorycznych, strumieniowych, a także danych związanych z analizą zachowań w internecie, w tym w sieciach społecznościowych. Analiza takich danych jest konieczna w czasie zbliżonym do rzeczywistego; c) różnorodność przetwarzanych danych (*variety*). Cecha ta może wiązać się z jednej strony z bardzo dużą liczbą różnego typu atrybutów ustrukturyzowanych danych w bazach relacyjnych. Przede wszystkim jednak dotyczy różnorodnych danych nieustrukturyzowanych, takich jak obrazy, materiały video, dźwięk, odręczne notatki, i inne dane nienadające się do przechowywania w klasycznych bazach relacyjnych.

Spoza klasycznego modelu 3V pozostałe cechy dotyczą: d) dużej wartości przetwarzanych danych (*value*). Cecha ta odnosi się do użyteczności danych i praktycznej możliwości ich wykorzystania przy podejmowaniu decyzji i w konsekwencji generowania istotnej wartości dla organizacji; e) wiarygodności przetwarzanych danych (*veracity*). Dotyczy zagadnienia jakości danych, w szczególności takich problemów, jak ich niekompletność, błędy, szumy i inne cechy wpływające na jakość, które są szczególnie istotne przy przetwarzaniu dużej ilości danych pochodzących z różnorodnych źródeł.

Tego typu definiowanie pojęcia jest w dużym stopniu zabiegiem marketingowym, lecz w praktyce tak najczęściej dzisiaj wyjaśnia się znaczenie *big data*. Mimo problemów z nowością i wieloaspektowością pojęcia można oczywiście odnaleźć także różne definicje *big data* formułowane przez naukowców, firmy analityczne lub twórców encyklopedii internetowych. Przykładowo organizacja McKinsey Global Institute (2011) pojęcie *big data* odnosi do zbiorów danych, których rozmiary prze-

kracząc możliwości typowych narzędzi bazodanowych w zakresie gromadzenia, przechowywania, zarządzania i analizowania tych danych. Z kolei w encyklopedii PcMag (2013) pojęcie *big data* zostało określone jako przetwarzanie ogromnej ilości zbieranych w czasie danych, które są trudne do analizy i obsługi za pomocą typowych narzędzi do zarządzania bazami danych. *Big data* obejmuje transakcje biznesowe, wiadomości e-mail, zdjęcia, wideo monitoringu, logi oraz inne dane generowane maszynowo, a także nieustrukturyzowany tekst umieszczany w internecie – w blogach i mediach społecznościowych. Boyd i Crawford (2012) wykraczają poza typowe podejście, określając *big data* jako kulturowe, technologiczne i naukowe zjawisko, które opiera się na technologii, analizie i mitologii. O ile aspekt technologii i analizy jest zrozumiały, o tyle ciekawe jest wyróżnienie aspektu mitologii jako powszechnego przekonania, że duże zbiory danych oferują wyższą formę inteligencji i wiedzy, pozwalającą generować spostrzeżenia, które wcześniej były niemożliwe.

Ze względu na komercyjną wartość pojęcia, próbują je przybliżyć potencjalnym klientom firmy oferujące rozwiązania wspomagające *big data*. Ponieważ powstanie koncepcji związane jest w dużym stopniu z rozwojem możliwości technologicznych, szczególnie istotnie wydaje się więc to, jak rozumieją to pojęcie czolowe firmy IT, gdyż pokazuje to prawdopodobne kierunki rozwoju *big data*. Przykładowo *Big data* oznacza tendencję do poszukiwania i wykorzystania wartości biznesowej drzemiącej w dostępnych coraz większych wolumenach danych (SAS 2013a). Termin *big data* używany jest do opisanego wykładniczego wzrostu i dostępności danych, zarówno ustrukturyzowanych, jak i nieustrukturyzowanych (SAS 2013b). Termin *big data* odpowiada rzeczywistości, w której każdego dnia powstaje 2,5 trylionów (ang. *quintillion*) bajtów danych i w konsekwencji 90% istniejących dziś danych powstało w ostatnich dwóch latach. Dane pochodzą zewsząd: z sensorów wykorzystywanych do zbierania informacji na temat klimatu, wpisów na portalach społecznościowych, cyfrowych zdjęć i filmów, rekordów z transakcji handlowych, sygnałów GPS z telefonów komórkowych, aby wymienić tylko kilka źródeł. Te dane to *big data* (IBM 2013). Termin *big data* używany jest do opisu procesu zastosowania poważnej mocy obliczeniowej (w uczeniu maszynowym i sztucznej inteligencji) do przetwarzania ogromnych i często bardzo skomplikowanych zbiorów informacji (Microsoft 2013). Pojęcie *big data* odnosi się do kilku podstawowych grup danych, takich jak: typowe dane przedsiębiorstwa (pochodzące przykładowo z systemów ERP, CRM i in.), dane pozyskiwane automatycznie (na przykład dane sensoryczne), dane pochodzące z internetu i mediów społecznościowych (Oracle 2013).

2. Technologia, metody analityczne i zastosowanie *big data*

Patrząc na fenomen *big data* zdaniem autora można wyróżnić trzy jego podstawowe aspekty:

- technologiczny (technologia informatyczna i metody analityczne),

- ekonomiczny (zastosowanie koncepcji),
- społeczny (konsekwencje zastosowań).

Aspekt technologiczny, obejmujący właściwą technologię informatyczną wraz ze stosowanymi ilościowymi metodami analitycznymi, należy traktować jako pierwotny. W pewnym sensie oczywiście jest on odpowiedzią na zapotrzebowanie (głównie biznesu), lecz potrzeby nie znajdujące pokrycia w możliwościach są czysto teoretyczne. Aspekt, tutaj nazwany ekonomicznym, traktowany jest więc jako rzeczywiste zastosowanie koncepcji *big data*. Dotyczy on głównie szeroko rozumianej ekonomii (w tym biznesu, a także administracji publicznej), lecz również innych nauk. Ponieważ zastosowania ekonomiczne wydają się na chwilę obecną najszerwsze, tak został nazwany ten aspekt. Aspekt społeczny jest wtórny, obejmujący kwestie wynikające z konsekwencji zastosowań koncepcji *big data*, głównie w zakresie zagrożenia prywatności.

2.1. Aspekt technologiczny

W zakresie technologii informatycznych praktyczne wykorzystanie koncepcji *big data* wymaga zastosowania komputerów o większej niż dotychczas mocy obliczeniowej oraz coraz częściej przetwarzania typu in-memory. Koniecznością staje się także rozproszenie przetwarzania, w tym zastosowanie koncepcji przetwarzania w chmurze (*cloud computing*) oraz paradygmatu MapReduce wraz z rozwiązaniem Apache Hadoop. Zazwyczaj stosuje się także inne niż przy przetwarzaniu transakcyjnym podejście do baz danych, wykorzystując rozwiązania typu NoSQL i bazy kolumnowe.

Odpowiadając na pytanie, co jest naprawdę nowego w koncepcji *big data* w porównaniu z wcześniej stosowanymi rozwiązaniami (jak *business intelligence* i *data mining*), zdaniem autora należy zwrócić szczególną uwagę na przetwarzanie w czasie rzeczywistym lub do niego zbliżonym, a także przetwarzanie danych słabo ustrukturyzowanych. Wzrost ilości przetwarzanych danych można uznać za ewolucyjny efekt postępu technologicznego i coraz większych dostępnych zbiorów danych. Nie ma więc sensu, definiując *big data*, wskazywanie konkretnej wielkości wykorzystywanych zbiorów. Należy zaznaczyć, że wcześniej także wykorzystywano bardzo duże zbiory danych w przetwarzaniu transakcyjnym, a do przetwarzania analitycznego wykorzystywano wielkie hurtownie danych, będące kopią danych transakcyjnych. Równolegle rozwijane były narzędzia zaawansowanej analizy danych oparte przede wszystkim na uczeniu maszynowym i koncepcji *data mining*. Jedną z cech *data mining* jest operowanie na dużych wolumenach danych, jednak, w porównaniu z *big data*, często takie analizy opierają się na próbie danych, co w wielu zastosowaniach nie jest wystarczające. Zdaniem autora termin *big data* w aspekcie technologicznym należy traktować jako opis nowych trendów w IT i metodach analizy, ze szczególnym uwzględnieniem analiz prowadzonych bezpośrednio na danych źródłowych w czasie rzeczywistym.

2.2. Aspekt ekonomiczny

Nowe możliwości technologiczne, a także gwałtowny spadek cen dostępu do zaawansowanej infrastruktury informatycznej, pozwoliły na nowe zastosowania analizy danych, w szczególności w szeroko rozumianej ekonomii. Błędem byłoby twierdzenie, że wszystkie zastosowania *big data* to wyłączna zasługa tej koncepcji. Przykładowo w zakresie zastosowań komercyjnych, rozwijający się od wielu lat *data mining* umożliwiał prowadzenie wielu podobnych analiz. Metoda ta była i jest powszechnie stosowana do rozwiązywania takich problemów, jak modelowanie ryzyka, wykrywanie oszustw, przewidywanie odejść klientów (*churn*) (Szupiluk 2013). Tym niemniej nowe możliwości analizy danych w czasie rzeczywistym oraz oparcie na całych dostępnych zbiorach danych, zamiast na ich próbie, znacząco rozszerzają potencjalne zastosowania. Nowe metody nie wymagają na wstępie stawiania tezy i budowy modelu, tylko umożliwiają poszukiwanie nieznanych wcześniej korelacji bezpośrednio na danych źródłowych. Dotąd organizacje projektujące struktury danych, w szczególności wdrażając hurtownie danych, oceniały z góry, które dane mają dla nich istotne znaczenie biznesowe. Nadmiar przetwarzanych danych mógł prowadzić do braku możliwości ich obróbki, a co najmniej znacząco zwiększał koszty i czas przetwarzania. Model stosowany w *big data* opiera się na założeniu niemal bezkosztowości gromadzenia i przetwarzania danych. Jak zauważa Płoszajski (2013), w historii zawsze było tak, że kiedy powszechność zastępowała rzadkość, to potrzebne były nowe modele biznesowe. Drastycznie spadający koszt zbierania i przetwarzania danych powoduje, że jesteśmy w stanie rozwiązywać zadania obliczeniowe, które wczoraj nie były ekonomicznie uzasadnione, a nawet wydawały się niemożliwe.

Można traktować rozwój podejścia *big data* jako odpowiedź na opisany przez Wieczorkowskiego i Dalka (2013) problem przeciążenia informacyjnego (*information overload*). Polega on na sytuacji, w której organizacja posiada dostęp do wystarczającej ilości istotnych dla niej danych, nie ma jednak możliwości ich przetworzenia i w konsekwencji przeprowadzenia procesu wnioskowania. Organizacja może więc mieć nadmiar informacji prawdziwej, lecz nieprzydatnej w praktyce ze względu na brak możliwości jej wykorzystania.

Rozważając aspekt ekonomiczny *big data* należy, poza kwestią udoskonalenia metod zarządzania, wziąć pod uwagę zagadnienia gospodarcze w skali makro. Nowe metody przetwarzania masowych danych dają możliwości rozwoju gałęzi gospodarki zajmujących się dostarczaniem usług dostępu do infrastruktury informatycznej (na przykład w modelu *cloud*), a także usług doradczych na podstawie analizy dużych wolumenów danych. Analizy mogą dotyczyć m.in. danych osobowych, ale to zagadnienie znajduje się na styku aspektu ekonomicznego i aspektu społecznego.

3. Społeczny aspekt *big data*

Aspekt społeczny dotyczy przede wszystkim zagadnienia przetwarzania i wykorzystywania danych osobowych, czyli zgodnie z Ustawą z dnia 29 sierpnia 1997 r. o ochronie danych osobowych – danych dotyczących zidentyfikowanej lub możliwej do zidentyfikowania osoby fizycznej, przy czym osobą możliwą do zidentyfikowania jest osoba, której tożsamość można określić bezpośrednio lub pośrednio, w szczególności przez powołanie się na numer identyfikacyjny albo jeden lub kilka specyficznych czynników określających jej cechy fizyczne, fizjologiczne, umysłowe, ekonomiczne, kulturowe lub społeczne.

Szczególą rolę wśród danych osobowych odgrywają te pochodzące z Internetu, który może być źródłem potencjalnie niemal niewyczerpanych danych o użytkownikach, traktowanych na przykład jako potencjalni konsumenci. Przykładowo banki i firmy pożyczkowe już dziś coraz szerzej zaczynają analizować portale społecznościowe w celu profilowania klientów i lepszej oceny ich wiarygodności kredytowej. Firmy ubezpieczeniowe mogą dzięki poznaniu klienta lepiej szacować indywidualne ryzyko. Reklamy, które ogląda użytkownik Internetu zaczynają opierać się na danych profilowanych w czasie rzeczywistym. Nie należy jednak pomijać danych osobowych pochodzących z innych źródeł, na przykład dane pochodzące z telefonii (*roaming*, dane lokalizacyjne), dane o prywatnych transakcjach finansowych (płatności kartami, transakcje z rachunków bankowych, kredyty), dane dotyczące dokonywanych zakupów powiązane z programami lojalnościowymi, dane wideo z monitoringu, dane medyczne, edukacyjne.

Głównym problemem w przypadku przetwarzania danych osobowych jest kwestia prywatności. Wymienione źródła dzięki technologiom *big data* mogą posiadać wielką wartość dla analiz prowadzonych w celach komercyjnych i publicznych. Istnieje jednak niebezpieczeństwo przekroczenia umownej bariery prywatności. Stąd ochrona prawna danych osobowych, w tym w szczególności danych wrażliwych. Przepisy dotyczące przetwarzania takich danych regulują także zasady przechowywania ich poza granicami kraju, co w przypadku rozproszenia przetwarzania, może stanowić ważną przeszkodę dla koncepcji *big data*. Społeczny aspekt *big data* jest więc zagadnieniem na styku wolności gospodarczej, bezpieczeństwa publicznego, praw człowieka, a także systemu prawnego.

W przypadku wielu analiz stosuje się metody anonimizacji. Tym niemniej często można odnaleźć symboliczne, oczywiście nie w pełni poważne, porównania *big data* do Big Brother Orwella. Według Urbanictionary (2013) *big data* to współczesna wersja Big Brother, a wyszukiwania online, zakupy w sklepach, wpisy na Facebooku, Tweeterze czy Foursquare, wykorzystanie telefonu komórkowego itp. tworzy zalew danych, które zorganizowane, skategoryzowane i przeanalizowane, ujawniają trendy i zwyczaje dotyczące nas samych i społeczeństwa. Jednakże, jak zauważył Simon (2013), Big Brother karał niewinnych za samodzielne myśle-

nie, traktowane jako zbrodnie myślowe, lecz również łapał rzeczywistych przestępców. Podobnie może być z *big data*, z jednej strony grozi totalną inwigilacją, z drugiej może przeciwdziałać rzeczywistej przestępczości, jak wyludzenia w sektorze ubezpieczeniowym. Z kolei (Mayer-Schonberger, Cukier 2013, s. 150) już w 2007 r. brytyjskie media zauważyły, że w odległości 200 jardów od londyńskiego mieszkania, w którym Orwell pisał *1984*, znajdowało się ponad 30 kamer monitoringu.

Z jednej więc strony prawo powinno zapewniać ochronę prywatności, z drugiej zaś nadmierna ochrona danych blokuje rozwój wykorzystania możliwości współczesnej technologii, mogąc negatywnie wpływać na rozwój gospodarczy. W USA toczy się obecnie dyskusja, czy informacje, które nie naruszają prywatności i bezpieczeństwa, a zostały zgromadzone przez instytucje publiczne, powinny być powszechnie dostępne i traktowane jako ogólnie dobro. Czy i w jakim stopniu, prawo powinno więc zezwalać na ponowne wykorzystywanie tego, co powstało za pieniądze publiczne (Economist 2013)? Ogólnie należy dążyć do sytuacji, w której osoby prywatne mogą chronić swoje dane osobowe, jednocześnie całkowicie nie ograniczając możliwości analizy danych. Podobnie sprawa jest widziana przez administrację publiczną w Polsce. W październiku 2013 r. ówczesny minister administracji i cyfryzacji Michał Boni mówił w kontekście *big data* o zachowaniu równowagi między możliwościami biznesowymi, a sposobami wyrażania zgody na przetwarzanie danych. Twierdził, że ministerstwo nie jest przeciwne profilowaniu, ale muszą być zachowane określone parametry czy warunki, na jakich się odbywa (MAC 2013a). W innej wypowiedzi w tym samym okresie twierdził, że od jakiegoś czasu obserwuje się zupełnie nowe zjawisko – nowe postrzeganie zagadnienia prywatności rośnie tak szybko jak rośnie znaczenie *big data* i tworzy wartość dodaną dla całej gospodarki (MAC 2013b).

Inna sytuacja dotyczy wykorzystania koncepcji *big data* przez policję, służby specjalne i agencje wywiadowcze. Dotyczy to nowych możliwości w zakresie infiltracji środowisk mogących stanowić zagrożenie dla państwa i porządku publicznego, a także wykrywania naruszeń prawa przez poszczególne jednostki. Do tych możliwości zalicza się przykładowo śledzenie Internetu, a także zbieranie danych z monitoringu. Analiza tego drugiego przypadku przeprowadzona przez Stępnia (2013) podkreśla duże znaczenie nie tylko bezpośredniej obserwacji ludzi, ale także identyfikacji pojazdów na podstawie tablic rejestracyjnych. Tego typu zastosowania *big data* mają na celu poprawę bezpieczeństwa, lecz sprawą otwartą pozostają kwestie etyczne i granica dozwolonej inwigilacji. Jak pokazał ostatni r. i ujawnienie w USA programu PRISM, koncepcja *big data* jest już od kilku lat w pełni szeroko wykorzystywana w USA w dziedzinie bezpieczeństwa.

Na pytanie jak wygląda analogiczna sytuacja w Polsce częściowo pomaga odpowiedzieć raport Fundacji Panoptikon (2013), analizujący na podstawie czterech dużych polskich firm świadczących usługi drogą elektroniczną, skalę otrzymywanych

od różnych organów państwa zapytań o dane użytkowników. Liczba takich zapytań w badanym okresie od stycznia 2012 roku do czerwca 2013 roku systematycznie rosła, lecz nie przekraczała kilku tysięcy rocznie. Brakuje wiedzy o charakterze tych zapytań, lecz analiza ich struktury wg organów pytających, w której dominowały zapytania od prokuratury (63%) i policji (33%), a znacząco mniejszy był udział takich organów jak ABW, CBA, Służba Celna i inne służby (łącznie 2%), pozwala przypuszczać, że zapytania głównie dotyczyły konkretnych prowadzonych postępowań karnych. Potwierdzają to informacje uzyskane bezpośrednio od służb w trybie dostępu do informacji publicznej. Przykładowo ABW w analizowanym okresie skierowała do firm świadczących usługi drogą elektroniczną 1149 zapytań. Zakładając wiarygodność danych i przyjmując, że państwo nie stosuje metod niejawnego sięgania po dane, można przypuszczać, że zapytania organów państwa kierowane do dostawców usług internetowych nie mają na razie wiele wspólnego z masową inwigilacją i koncepcją *big data*. Powyższy raport dotyczył przede wszystkim kontroli danych pozyskiwanych z Internetu, z kolei NIK (2013) opublikował raport dotyczący pozyskiwania i przetwarzania przez uprawnione organy państwowe danych telekomunikacyjnych (m.in. o bilingach i danych lokalizacyjnych) w okresie od stycznia 2011 roku do czerwca 2012 roku. Mimo generalnie pozytywnej oceny przestrzegania obowiązujących zasad, w ocenie NIK aktualne przepisy w zakresie pozyskiwania przez uprawnione podmioty danych telekomunikacyjnych nie chronią w stopniu wystarczającym praw i wolności obywatelskich przed nadmierną ingerencją ze strony państwa. Niejednolitość i ogólnikowość przepisów uprawniających do pozyskiwania danych telekomunikacyjnych, może nasuwać wątpliwości, co do współmierności stosowanych ograniczeń praw i wolności obywatelskich w sferze wolności komunikacji z zasadami określonymi w Konstytucji RP. Na podstawie powyższych raportów można więc wysunąć wniosek, że nadmierna inwigilacja Internetu i usług telekomunikacyjnych ze strony państwa nie była dotychczas znaczącym problemem, jednak wraz z upowszechnieniem metod *big data* niezbędne jest doprecyzowanie odpowiednich zasad prawnych.

Podsumowanie

Powyższe rozważania prowadzą do pytania, na ile zjawisko *big data* jest zagrożeniem z obszaru technologii i ilościowych metod analitycznych, na ile ekonomii, a na ile zaś z obszaru kwestii społecznych. Naturalne jest zainteresowanie naukowców nowościami i badanie wpływu nowych zjawisk technologicznych i ekonomicznych na otoczenie społeczne. W historii często wielkie wynalazki i postęp techniczny prowadziły do znaczących przemian społecznych. Czy rzeczywiście nowe możliwości dotyczące przetwarzania danych określane jako *big data* mają podobne znaczenie? Trudno dziś odpowiedzieć na to pytanie, natomiast na

pewno poprzez aktualizowanie systemu prawnego należy zapobiegać niekorzystnym wpływom zmian technologicznych na stosunki społeczne.

Nie wydaje się właściwe oparcie społecznego aspektu *big data* wyłącznie na technologicznym i ekonomicznym rozumieniu tego pojęcia. Aspekt społeczny *big data* jest związany ogólnie z problemem prywatności i ochrony danych osobowych. Z jednej strony możliwości współczesnych technologii i metod numerycznych w przetwarzaniu dużych wolumenów szybkozmiennych i nieustrukturyzowanych danych znajdują znacznie szersze zastosowania niż te, które są poruszane w społecznym aspekcie *big data* (poza badaniem zachowań konsumentów, to także na przykład badanie zjawisk fizyki kwantowej, obserwacja klimatu). Z drugiej strony społeczny aspekt *big data* to nie tylko możliwości przetwarzania danych osobowych, ale także kwestie powszechnego dostępu do technologii internetowych, funkcjonującego systemu prawnego, organizacji gospodarki w zakresie handlu i usług detalicznych itp. Tak więc aspekty: technologiczny, ekonomiczny i społeczny *big data* to nie w pełni pokrywające się zbiory, z istotną częścią wspólną, choć aspekt społeczny wydaje się najbardziej odrębny.

Literatura

- Boyd D., Crawford K. (2012), *Critical questions for big data*, „Information, Communication & Society” Vo. 15, No 5, s. 662–679.
- Economist (2013), <http://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innovation-new-goldmine>.
- IBM (2013), <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>.
- Laney D. (2001), *Application delivery strategies*, META Group, Stamford.
- MAC (2013a), <https://mac.gov.pl/dzialania/michal-boni-o-big-data-na-warsawy-international-media-summit/>.
- MAC (2013b), <https://mac.gov.pl/dzialania/prywatnosc-zaufanie-wolnosc-i-bezpieczenstwo-michal-boni-przedstawia-polska-wizje-rozwoju-sieci-na-swiatowej-konferencji-o-cyberprzestrzeni-w-seulu/>.
- Mayer-Schonberger V., Cukier K. (2013), *Big data – A revolution that will transform how we live, work, and think*, An Eamon Dolan Book Houghton Mifflin Harcourt, Boston.
- Microsoft (2013), <http://blogs.msdn.com/b/microsoftenterpriseinsight/archive/2013/04/15/the-big-bang-how-the-big-data-explosion-is-changing-the-world.aspx>.
- McKinsey (2011), *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute.
- NIK (2013), *Informacja o wynikach kontroli „Uzyskiwanie i przetwarzanie przez uprawnione podmioty danych z bilingów, informacji o lokalizacji oraz innych da-*

- nych, o których mowa w art. 180 c i d ustawy Prawo telekomunikacyjne”, Naczelna Izba Kontroli, Warszawa.
- Oracle(2013), <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>.
- PcMag, <http://www.pcmag.com/encyclopedia/term/62849/big-data> (2013).
- Płoszajski P. (2013), *Big Data: nowe źródło przewag i wzrostu firm*, „e-mentor” nr 3 (50), s. 5–10.
- SAS (2013a), http://www.sas.com/offices/europe/poland/actual/press/news2_01_13.html.
- SAS (2013b), <http://www.sas.com/big-data>.
- Simon P. (2013), *Too big to ignore – The business case for big data*, Wiley, Hoboken.
- Stępnia C. (2013), *Kierunki wykorzystania systemów monitoringu miejskiego w zarządzaniu rozwojem miast*, „Roczniki Kolegium Analiz Ekonomicznych SGH” z. 29, Oficyna Wydawnicza SGH, Warszawa.
- Szupiluk R. (2013), *Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli data mining*, Oficyna Wydawnicza SGH, Warszawa.
- Szymielewicz K., Szumańska M. (2013), *Dostęp państwa do danych użytkowników usług internetowych, Siedem problemów i kilka hipotez*, Fundacja Panoptykon, Warszawa.
- Urbandictionary (2013), <http://www.urbandictionary.com/define.php?term=Big%20Data>.
- Ustawa z dnia 29 sierpnia 1997 r. o ochronie danych osobowych (Dz.U. 1997, Nr 133, poz. 883, ze zm).
- Wieczorkowski J., Dalek M. (2013), *Problem przeciężenia informacyjnego a integracja systemów informatycznych*, Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 762, Ekonomiczne Problemy Usług nr 104, Szczecin, s. 439–448.

BIG DATA –TECHNOLOGICAL AND COMMERCIAL ASPECTS VS. SOCIAL ASPECT

Summary

The term big data is a new, shaping up concept. The paper proposes the distinction of three fundamental aspects of this term: technological (including technology and analytical methods), economic (including the variety of uses), and social. The last-named, mainly concerning threats to privacy, is a secondary aspect to the previous, but it was given the most attention in the article.

Keywords: big data, data analysis, personal data

Translated by Jędrzej Wieczorkowski