Nierówności Społeczne a Wzrost Gospodarczy, nr 44 (4/2015), część 2
DOI: 10.15584/nsawg.2015.4.2.30 **ISSN 1898-5084**

dr Jędrzej Wieczorkowski[1]

Instytut Informatyki i Gospodarki Cyfrowej, Kolegium Analiz Ekonomicznych
Szkoła Główna Handlowa w Warszawie

Zagadnienia społeczne i prawne w koncepcji *big data*

WPROWADZENIE

Termin *big data* jest ostatnio bardzo chętnie stosowany w dość różnorodnych znaczeniach. Wykorzystują go równie chętnie naukowcy zajmujący się przykładowo nowymi technologiami, metodami analizy danych i zagadnieniami społecznymi, a także przedstawiciele biznesu oraz dziennikarze. Znajduje się on na pograniczu zagadnień technologii informatycznych, analitycznych metod ilościowych oraz społecznych. Trwa dyskusja, czy jest to nowe zjawisko, czy też stopniowa ewolucja trendów w przetwarzaniu danych. Czy *big data* nie jest tylko hasłem marketingowym, dzięki któremu próbuje się sprzedać nowej generacji produkty i usługi analizy danych? Czy *big data* to nie jest odświeżona koncepcja business intelligence z wykorzystaniem nowych możliwości technologicznych?

Celem artykułu jest próba zrozumienia pojęcia *big data* w oparciu o jego powszechną interpretację z wykorzystaniem badania przeprowadzonego przez autora, a następnie przeprowadzenie analizy wynikających z niego zagadnień społecznych i prawnych *big data*.

IDEA KONCEPCJI *BIG DATA*

Już w 2001 roku firma doradcza META Group opublikowała raport dotyczący wpływu handlu elektronicznego, zjawiska globalizacji i innych trendów na rozwój technologii informatycznych [Laney, 2001]. Powstał wówczas model, określany jako 3V, który stał się podstawą wykrystalizowanej później koncepcji *big data*. Jako jej cechy wynikające z powyższego modelu powszechnie więc uważa się:

- volume, czyli duży wolumen przetwarzanych danych,

[1] Adres korespondencyjny: Instytut Informatyki i Gospodarki Cyfrowej, Kolegium Analiz Ekonomicznych, Szkoła Główna Handlowa w Warszawie, ul. Madalińskiego 6/8, 02-513 Warszawa, e-mail: jedrzej.wieczorkowski@sgh.waw.pl, tel. 22 564 92 80.

- velocity, czyli zmienność przetwarzanych danych,
- variety, czyli różnorodność przetwarzanych danych.

W późniejszym czasie różni autorzy starali się dobrać inne cechy charakterystyczne dla *big data*, które mogłyby rozszerzyć model „v", w szczególności:
– value, czyli dużej wartości przetwarzanych danych,
– veracity, czyli wiarygodności przetwarzanych danych,
– visualisation, czyli możliwości wizualizacji danych.

Ogólnie pojęcie *big data* można odnosić do zbiorów danych, których rozmiary przekraczają możliwości typowych narzędzi bazodanowych w zakresie gromadzenia, przechowywania, zarządzania i analizowania tych danych [*Big Data…*, 2011]. *Big data* obejmuje transakcje biznesowe, wiadomości e-mail, zdjęcia, wideomonitoring, logi aktywności oraz inne dane generowane maszynowo, a także nieustrukturyzowany tekst umieszczany w Internecie, w tym w blogach i mediach społecznościowych [www.pcmag]. 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 (np. dane sensoryczne), dane pochodzące z Internetu i mediów społecznościowych [*Big Data…*, 2013].

Według autora niniejszego opracowania szczególnie istotne są nowe możliwości *big data* w zakresie przetwarzania w czasie rzeczywistym lub do niego zbliżonym, a także możliwości przetwarzania danych słabo ustrukturyzowanych. Natomiast sam wzrost ilości przetwarzanych danych jest ewolucyjnym trendem – efektem postępu technologicznego i coraz większych zbiorów dostępnych danych. W przeciwieństwie do typowego przetwarzania analitycznego OLAP (*Online Analytical Processing*), w którym opierano się na dobrze ustrukturyzowanych i zazwyczaj zagregowanych danych zapisanych w hurtowniach, *big data* umożliwia dokonywanie analiz w czasie rzeczywistym z wykorzystywaniem szczegółowych, często nieustrukturyzowanych danych źródłowych.

BADANIE INTERPRETACJI TERMINU *BIG DATA*

Wieloaspektowość i nowość pojęcia *big data* skłania do podjęcia badania sposobu jego zrozumienia. Z jednej strony można analizować literaturę fachową oraz próby definiowania terminu, a także dokonać przeglądu materiałów opracowanych i publikowanych przez firmy z branży technologicznej, które biznesowo wykorzystują nowe możliwości. Z drugiej strony można dokonać próby zbadania powszechnego zrozumienia tego dopiero kształtującego się pojęcia na podstawie wybranego języka. Taka próba może opierać się na badaniu interpretacji pojęcia wykorzystywanego w różnorodnych, jak najbardziej powszechnych, publikacjach typu popularnego. Zdaniem autora to drugie podejście wydaje właściwsze z punktu widzenia niniejszego opracowania.

Najdogodniejszą bazą tego typu danych wydaje się Internet. Autor postanowił wykonać próbę poszukiwania znaczenia użycia terminu *big data* dla języka polskiego. Narzucającą się możliwością jest wykorzystanie popularnych wyszukiwarek internetowych. Autor przeprowadził więc test przy zastosowaniu wyszukiwarki Google.pl po wyłączeniu, z oczywistych względów, w ustawieniach funkcji „wyniki prywatne", sprawiającej, że znajdowane pozycje powinny być lepiej dostosowane do preferencji użytkownika, a także po ograniczeniu wyszukiwania do języka polskiego. Uzyskane wyniki z punktu widzenia postawionego celu nie wydały się jednak zadowalające. Ze względu na stosowane w wyszukiwarkach algorytmy, na szczycie list wyszukiwania zazwyczaj znajdują się strony próbujące wyjaśniać hasło w sposób encyklopedyczny (np. na pierwszym miejscu Wikipedia), a także strony firm biznesowo wykorzystujących szukane hasło (prawdopodobnie ze względu na ich pozycjonowanie poparte nakładami finansowymi przeznaczanymi na ten cel). W przeprowadzonym teście (wrzesień 2014 r.) wśród pierwszych dziesięciu wyników znalazły się:

- 4 strony firm biznesowo zajmujących się omawianym pojęciem,
- 2 artykuły internetowych wydań czasopism,
- 1 artykuł na blogu,
- 1 strona o charakterze encyklopedycznym,
- 1 strona zbiorcza grupująca inne strony o zadanej tematyce,
- 1 opracowanie o charakterze naukowym.

Taki wybór stron może więc mieć niewiele wspólnego z badaniem powszechnego rozumienia sensu danego terminu. Z tego względu, zdaniem autora, lepszą metodą wydaje się ograniczenie przeszukiwania do publikacji prasowych, w miarę możliwości niezwiązanych z pismami fachowymi.

W ramach kolejnego testu autor dokonał więc analizy artykułów publikowanych przez grupę Agora wykorzystując wyszukiwarkę na portalu Gazeta.pl. Wyboru tego portalu dokonano, gdyż zawiera on artykuły i notatki prasowe pochodzące z dużej grupy wydawniczej posiadającej różnorodne publikacje, zarówno tradycyjne, jak i wyłącznie internetowe. Do badania zastosowano wbudowaną wyszukiwarkę portalu. Następnie poddano analizie najwyżej sklasyfikowane pozycje dla hasła *big data*. Odrzucono kilka pozycji, w których fraza pojawiła się z punku widzenia badania przypadkowo lub artykuł dotyczył zasadniczo innego tematu, a *big data* było tematem zupełnie pobocznym (np. zapowiedzi wydarzeń lokalnych, relacja z targów, analiza rynku pracy). W ten sposób pozostawiono dziesięć najwyżej sklasyfikowanych pozycji, określono dział lub część pisma, w której pojawił się artykuł (wydzielono: ogólne informacje, biznes/finanse, technologie, reportaże) i następnie przeanalizowano treść wskazanych artykułów. Spośród nich jeden pochodził z działu bieżących informacji ogólnych, pięć informacji z biznesu lub finansów, trzy z działu technologicznego, jeden z dodatku reportaży. Treści artykułów przypisano trzem opisanym przez Jędrzeja Wieczor-

kowskiego i Przemysława Polaka podstawowym aspektom *big data* [Wieczorkowski, Polak, 2014, s. 182–196]:
- technologicznemu (technologia informatyczna i metody analityczne),
- ekonomicznemu (zastosowanie koncepcji),
- społecznemu (społeczne konsekwencje zastosowań).

Do każdego artykułu przypisano jeden lub więcej wiodących powyższych aspektów. W przypadku jednego artykułu były to wszystkie trzy aspekty, w trzech artykułach – dwa aspekty niemal równoważne – w sześciu jeden wyraźnie wyróżniający się aspekt. Wyniki analizy przedstawiono w tabeli 1.

Tabela 1. Najpopularniejsze artykuły w podziale na aspekty *big data*

Lp.	Dział	Okres	Aspekt		
			technologiczny	ekonomiczny	społeczny
1	technologie	marzec 2014 r.			×
2	biznes	grudzień 2013 r.			×
3	magazyn	kwiecień 2012 r.		×	×
4	technologie	czerwiec 2013 r.		×	
5	biznes	maj 2013 r.			×
6	biznes	maj 2013 r.			×
7	technologie	kwiecień 2013 r.	×	×	×
8	biznes	czerwiec 2014 r.	×	×	
9	informacje	maj 2013 r.			×
10	biznes	sierpień 2014 r.	×	×	

Źródło: opracowanie własne.

W trzech przypadkach do artykułu przypisano aspekt technologiczny, ale było to zawsze przypisanie na równi z co najmniej jednym innym aspektem. W pięciu przypadkach przypisano aspekt ekonomiczny, lecz w dwóch z nich w połączeniu co najmniej z aspektem społecznym. Artykuły mające wyłącznie aspekt ekonomiczny skupiały się na przykładach możliwości zastosowań *big data* – w jednym przypadku dotyczyło to szczegółowych zastosowań medycznych, pozostałe omawiały możliwe zastosowania w sposób bardziej ogólny. Najwięcej przypisań – siedem – dotyczyło aspektu społecznego. W badanej próbie szczególnie popularny był temat zbierania różnorodnych szczegółowych danych osobowych przez instytucje finansowe, który poruszono aż w trzech pozycjach. Prawdopodobnie wynikało to głównie z głośnego wywiadu udzielonego w lutym 2013 roku przez dyrektora Alior Banku, w którym wyjątkowo szczerze opowiadał o możliwości wykorzystywania danych o klientach, co odbiło się w środowisku bankowym szerokim echem. Ale co ciekawe, analizowane artykuły skupiły się nie na ekonomicznym aspekcie wykorzystania różnorodnych danych tego typu, a na aspekcie społecznym – zagrożeniu inwigilacją. Pisano też o potencjalnych zagrożeniach utraty prywatności wynikających z kontroli dużej części Internetu, w tym portali społecznościowych, przez podmioty sektora prywatnego. W jedynym ana-

lizowanym artykule, opisującym całościowo zagadnienie *big data*, znaczącą część także poświęcono zagrożeniu inwigilacją.

Ciekawym spostrzeżeniem przy analizie artykułów jest fakt, że jako przetwarzanie typu *big data* rozumie się masowe wykorzystanie danych osobowych na potrzeby biznesu, natomiast, mimo że pisze się również o problemie prywatności w kontekście masowego przetwarzania takich danych przez urzędy administracji publicznej, zazwyczaj nie używa się wówczas terminu *big data*. W 2014 roku w prasie pisano dużo na temat konieczności dostosowania różnorodnych ustaw o służbach państwa do orzeczeń Trybunału Konstytucyjnego. W artykułach poruszano wówczas temat zagrożenia inwigilacją wynikającą z automatycznego masowego przetwarzania danych osobowych, ale mimo że spełnia ono rozumienie przetwarzania typu *big data*, zazwyczaj ten termin nie padał.

Należy wziąć oczywiście pod uwagę fakt, że portal ma charakter uniwersalny i zagadnienia naukowe oraz techniczne stanowią stosunkowo niewielką część informacji, choć jednocześnie dział biznesowy, czyli potencjalnych zastosowań koncepcji, jest dość obszerny. Tym niemniej wynik testu wskazuje na fakt, że „fachowe" zdefiniowanie *big data* znacznie rozbiega się z potocznym rozumieniem pojęcia, a także na to, że tematy nurtujące dziennikarzy, prawdopodobnie wynikające z zainteresowań społeczeństwa, dotyczą przeważnie zagadnień społecznych. Można także przypuszczać, że powyższa prawidłowość wynika z faktu, że wielu dziennikarzy, a nawet naukowców, stara się podążać za aktualnymi i modnymi tematami, a jednocześnie nie jest kompetentna do zajmowania się zagadnieniami technologicznymi. Tym samym rozwijają oni „miękki" temat wpływu przemian technologicznych i ekonomicznych na społeczeństwo. Niemniej jednak, niezależnie od przyczyny zainteresowań zagadnieniami społecznymi, to one odgrywają bardzo istotną rolę w obecnym rozumieniu pojęcia *big data*.

SPOŁECZNE ZAGADNIENIA *BIG DATA*

Jak wskazuje opisany powyżej test, społeczny aspekt *big data* dotyczy przede wszystkim przetwarzania i wykorzystywania danych osobowych, problemów naruszania prywatności oraz zagrożeń inwigilacją.

Najistotniejszym źródłem niemal niewyczerpanych danych osobowych jest obecnie Internet, w tym w szczególności serwisy społecznościowe. Ich użytkownicy traktowani są jako potencjalni konsumenci i dane na ich temat mogą mieć znaczącą wartość. Jeszcze przed współczesnymi możliwościami analizy danych w czasie rzeczywistym zwracano uwagę na korzyści z eksploracji danych pochodzących z Internetu, wówczas z wykorzystaniem technologii hurtowni danych [Pawełoszek-Korek, 2008, s. 553–559]. Obecnie podstawowy model biznesowy serwisów społecznościowych zakłada udostępnienie społeczności platformy użytkowej w zamian za dostęp do spersonalizowanych strumieni informacji współ-

tworzonych i współdzielonych przez społeczność [Polańska, Wassilew, 2013]. Reklamy oglądane przez użytkowników Internetu mogą opierać się na danych profilowanych w czasie rzeczywistym. Banki i firmy pożyczkowe mogą analizować portale społecznościowe w celu profilowania klientów i lepszej oceny ich wiarygodności kredytowej. Firmy ubezpieczeniowe mogą dzięki danym pozyskiwanym z Internetu lepiej szacować indywidualne ryzyko klienta.

Wartość biznesową mogą mieć także masowe dane osobowe pochodzące z innych źródeł, przykładowo 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. Wykorzystanie wszystkich tych danych wiąże się z koniecznością przetwarzania bardzo dużych wolumenów i zastosowania metod typowych dla *big data*.

W zakresie wykorzystania koncepcji *big data* na potrzeby funkcjonowania państwa jako całości, w szczególności zapewnienia bezpieczeństwa publicznego, ważną rolę odgrywa przetwarzania danych osobowych przez takie organy jak policja, agencje wywiadowcze i inne służby specjalne. Przetwarzanie może dotyczyć infiltracji całych grup mogących stanowić potencjalne zagrożenie oraz wykrywania naruszeń prawa przez poszczególne jednostki. W obszarze działań prewencyjnych szczególnie istotne jest obecnie śledzenie Internetu. Takie działania, jak pokazało ujawnienie programu PRISM, są od kilku lat prowadzone w Stanach Zjednoczonych w szerokim zakresie. Brak jest podobnych informacji w przypadku Polski, tym niemniej podejmuje się próby dotarcia do nich. Przykładem jest opracowany przez Fundację Panoptykon raport analizujący skalę oficjalnych zapytań o dane użytkowników otrzymywanych od różnych organów państwa przez dostawców usług internetowych [Szymielewicz, Szumańska, 2013].

Zakładając, że państwo nie stosuje pozaprawnych metod sięgania po takie dane, można na podstawie raportu przypuszczać, że zapytania tego typu nie mają na razie w Polsce charakteru masowej inwigilacji związanej z koncepcją *big data*. Innym źródłem danych zbieranych na potrzeby bezpieczeństwa publicznego jest monitoring miejski i drogowy. Dostarcza on sensorycznych danych o charakterze nieustrukturyzowanym możliwych do wykorzystania dzięki metodom *big data*. Przeprowadzona przez Cezarego Stępniaka analiza [Stępniak, 2013, s. 295–307] podkreśla, że obok bezpośredniej obserwacji ludzi, w szczególności w miejscach i terminach ponadprzeciętnie zagrożonych, duże znaczenie ma również identyfikacja pojazdów na podstawie ich tablic rejestracyjnych i dalsze przetwarzanie tych danych.

Kolejnym źródłem danych wykorzystywanych na potrzeby funkcjonowania państwa są różnorodne, zazwyczaj wielkie bazy danych, zarówno te tworzone przez administrację publiczną, jak i przedsiębiorstwa. Przykładowo analiza danych pochodzących z administracji podatkowej, celnej i rejestrów publicznych, wsparta śledzeniem Internetu znajduje zastosowanie w wykrywaniu nadużyć fi-

nansowych. Analiza danych pochodzących od operatorów telekomunikacyjnych wykorzystywana jest w celu wykrywania sprawców przestępstw [Wieczorkowski, Polak, 2014, s. 567–579].

PRAWNE ZAGADNIENIA *BIG DATA*

Głównym zadaniem prawnego systemu państwa w kontekście *big data* jest zapewnienie niezbędnego minimum ochrony prywatności jednostki. Problem ten został dość obszernie potraktowany na poziomie prawa europejskiego, w którym istnieje dyrektywa 95/46/WE Parlamentu Europejskiego i Rady z 24 października 1995 r. w sprawie ochrony osób fizycznych w zakresie przetwarzania danych osobowych i swobodnego przepływu tych danych [Dyrektywa…].

W części wstępnej zauważono w niej, że coraz częściej korzysta się z przetwarzania danych osobowych w różnych sferach życia gospodarczego i społecznego, a postęp w dziedzinie technologii informatycznych sprawia, że przetwarzanie i wymiana takich danych stają się coraz łatwiejsze. Odnotowano tam, że jeżeli w ramach społeczeństwa informacyjnego ma znaczenie rozwój technik gromadzenia, przekazywania, kompilowania, rejestrowania, przechowywania i przesyłania danych dźwiękowych i obrazowych osób fizycznych, dyrektywa powinna mieć zastosowanie do przetwarzania takich danych, a ochrona osób fizycznych musi odnosić się zarówno do automatycznego, jak i ręcznego przetwarzania danych, zaś zakres tej ochrony nie może w swym skutku być zależny od zastosowanych technik.

W dyrektywie, jako dane osobowe zdefiniowano wszelkie informacje dotyczące zidentyfikowanej lub możliwej do zidentyfikowania osoby fizycznej. Osoba możliwa do zidentyfikowania to osoba, której tożsamość można ustalić bezpośrednio lub pośrednio, szczególnie przez powołanie się na numer identyfikacyjny lub jeden bądź kilka szczególnych czynników określających jej fizyczną, fizjologiczną, umysłową, ekonomiczną, kulturową lub społeczną tożsamość. Natomiast przetwarzanie danych osobowych oznacza każdą operację lub zestaw operacji dokonywanych na danych osobowych przy pomocy środków zautomatyzowanych lub innych, jak np. gromadzenie, rejestracja, porządkowanie, przechowywanie, adaptacja lub modyfikacja, odzyskiwanie, konsultowanie, wykorzystywanie, ujawnianie poprzez transmisję, rozpowszechnianie lub udostępnianie w inny sposób, układanie lub kompilowanie, blokowanie, usuwanie lub niszczenie [Dyrektywa…]. Szczególną rolę odgrywają dane zazwyczaj określane jako wrażliwe. W dyrektywie potraktowano je jako szczególne kategorie danych i stwierdzono że państwa członkowskie z wyjątkiem wyszczególnionych przypadków zabraniają przetwarzania danych osobowych ujawniających pochodzenie rasowe lub etniczne, opinie polityczne, przekonania religijne lub filozoficzne, przynależność do związków zawodowych, jak również przetwarzanie danych dotyczących zdro-

wia i życia seksualnego. Natomiast przetwarzanie danych dotyczących przestępstw, wyroków skazujących lub środków bezpieczeństwa może być dokonywane jedynie pod kontrolą władz publicznych.

W polskim systemie w omawianym kontekście najważniejszą rolę odgrywa ustawa z 29 sierpnia 1997 r. o ochronie danych osobowych [Ustawa…]. Określenie danych osobowych w pełni odpowiada w niej definicji z powyższej dyrektywy. We współczesnym świecie, w którym gospodarka jest silnie zglobalizowana, problemem jest zasięg obowiązywania prawa krajowego.

W szczególności problem ten jest widoczny w przypadku norm dotyczących przetwarzania danych. Z uwagi na wykorzystywanie Internetu będącego siecią o zasięgu globalnym, przetwarzanie rozproszone, w tym stosowanie chmur obliczeniowych, a także rozproszenie geograficzne organizacji gromadzących i przetwarzających dane, niejednokrotnie trudno jest egzekwować na danym terenie prawo krajowe. Choć przepisy dotyczące przetwarzania niektórych danych mogą przykładowo regulować zasady przechowywania ich na serwerach poza granicami danego kraju, dotychczasowy model stanowienia prawa krajowego, wspomagany umowami o międzynarodowym charakterze, coraz wyraźniej nie zdaje egzaminu.

Prawne uregulowanie problemu przetwarzania danych osobowych w praktyce dotyczy przede wszystkim dwóch podstawowych aspektów:

- przetwarzania danych na potrzeby biznesu,
- przetwarzania danych na potrzeby funkcjonowania państwa, w szczególności bezpieczeństwa publicznego i zarządzania finansami publicznymi.

REGULACJE DOTYCZĄCE *BIG DATA* W BIZNESIE

Wymienione wcześniej akty prawne skupiają się na pierwszym z wymienionych aspektów. We wspomnianej dyrektywie zaznaczono, że nie ma ona zastosowania do przetwarzania danych osobowych na rzecz bezpieczeństwa publicznego, obronności, bezpieczeństwa państwa, a przykładowo przetwarzanie danych dźwiękowych i obrazowych, w przypadku nadzoru kamer wideo, nie wchodzi w zakres stosowania niniejszej dyrektywy, jeśli dokonywane jest dla potrzeb bezpieczeństwa publicznego, obronności, bezpieczeństwa narodowego.

Interesującym prawnie zagadnieniem jest umożliwienie użytkownikowi w oparciu o orzeczenie Trybunału Sprawiedliwości Unii Europejskiej [Trybunał…] bycia zapomnianym w Internecie, a konkretnie w wynikach haseł wyszukiwarki internetowej. Warunkiem jest wykazanie priorytetu prawa do prywatności nad prawem dostępu do wyników wyszukiwania zawierających imię i nazwisko. Trybunał Sprawiedliwości UE stwierdził więc, że operator wyszukiwarki internetowej jest odpowiedzialny za dokonywane przezeń przetwarzanie danych osobowych, które pojawiają się na stronach internetowych publikowanych przez

osoby trzecie. Mimo że użytkownicy Internetu mogą mieć uzasadniony interes w uzyskaniu dostępu do informacji, prawo do prywatności osoby, której dotyczą dane, jest co do zasady nadrzędne wobec interesu internautów. Należy jednak dążyć do znalezienia punktu równowagi pomiędzy interesem użytkowników a prawami podstawowymi osoby, której dotyczą dane.

Ciągle pojawiają się nowe prawne dylematy związane z zapewnieniem prywatności. Przykładem takich problemów w Polsce są aktualnie prowadzone prace nad przepisami umożliwiającymi wprowadzenie inteligentnych liczników poboru energii elektrycznej. W sprawę zaangażowany został Generalny Inspektor Ochrony Danych Osobowych, gdyż zachodzi obawa zbierania zbyt szczegółowych danych o zachowaniach użytkowników energii. Podobne wątpliwości pojawiły się odnośnie do przepisów dotyczących zbierania odpadów i ewentualnym znakowaniu worków z odpadami pochodzącymi z gospodarstw domowych w celu identyfikacji ich źródła pochodzenia.

Dyskusje dotyczące regulacji dopuszczalnego poziomu przetwarzania masowych danych osobowych dotyczą praktycznie wszystkich rozwiniętych państw. Przykładowo w Stanach Zjednoczonych trwa debata, czy informacje, które nie naruszają prywatności i bezpieczeństwa, zgromadzone przez publiczne instytucje za publiczne środki powinny być traktowane jako ogólne dobro i w konsekwencji być powszechnie dostępne m.in. dla biznesu [Economist 2013]. W Polsce, w 2013 roku, ówczesny minister administracji i cyfryzacji Michał Boni mówił o zachowaniu równowagi między możliwościami biznesowymi a sposobami wyrażania zgody na przetwarzanie danych, twierdząc, że ministerstwo nie jest przeciwne profilowaniu danych osobowych przy zachowaniu określonych warunków [https://mac].

W zakresie regulacji przetwarzania danych osobowych na potrzeby biznesu, prawo z jednej strony powinno zapewniać ochronę prywatności, jednocześnie z drugiej strony nie blokując poprzez nadmierną ochronę danych rozwoju wykorzystania możliwości współczesnej technologii i przez to nie wpływając negatywnie na rozwój gospodarczy. Stosowane zapisy muszą być wystarczająco ogólne, aby nie ograniczać się do znanych na daną chwilę rozwiązań i wyprzedzić powstanie nowych możliwości oferowanych przez technologię, a jednocześnie muszą być na tyle precyzyjne, aby utrudniać obchodzenie prawa.

REGULACJE DOTYCZĄCE *BIG DATA* W FUNKCJONOWANIU PAŃSTWA

Istotnym zagadnieniem prawnym jest konieczność zapewnienia odpowiednim organom dostępu do danych i możliwości ich przetwarzania na potrzeby sprawnego funkcjonowania państwa i społeczeństwa. Natomiast pytaniem pozostaje granica dozwolonej inwigilacji poszczególnych osób.

W 2013 r. Najwyższa Izba Kontroli opublikowała raport dotyczący pozyskiwania i przetwarzania w latach 2011–2012 przez uprawnione ograny państwowe danych telekomunikacyjnych, m.in. o bilingach i danych lokalizacyjnych[2]. Zwrócono w nim uwagę, że 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. Problemem jest niejednolitość i ogólnikowość odpowiednich przepisów. Może ona nasuwać wątpliwości, co do współmierności stosowanych ograniczeń w sferze wolności komunikacji z zasadami określonymi w Konstytucji RP.

Problem ten na bardziej ogólnym poziomie był rozpatrywany przez Trybunał Konstytucyjny. Sprawa została wniesiona niezależnie przez rzecznika praw obywatelskich (na wniosek Helsińskiej Fundacji Praw Człowieka) oraz prokuratora generalnego. Kwestia rozpatrywana była przez Trybunał w pełnym składzie na dwóch posiedzeniach w kwietniu i lipcu 2014 r. i dotyczyła określenia katalogu zbieranych informacji o jednostce za pomocą środków technicznych w działaniach operacyjnych oraz zasad niszczenia pozyskanych danych. Wnioski podkreślały, że działania służb specjalnych, które ingerują w życie prywatne jednostki muszą opierać się na dostatecznych podstawach prawnych w randze ustawy. Zarzutem więc jest przede wszystkim brak wystarczającej precyzji w dość ogólnych zapisach ustawowych, dotyczących przykładowo rodzaju danych o jednostce pozyskiwanych przez służby w trybie kontroli operacyjnej oraz środków które mogą być wykorzystywane przez służby. Problem dotyczył zgodności z Konstytucją RP i Konwencją o ochronie praw człowieka i podstawowych wolności m.in. ustaw o Policji, Straży Granicznej, kontroli skarbowej, Żandarmerii Wojskowej, Agencji Bezpieczeństwa Wewnętrznego, Agencji Wywiadu, Służbie Wywiadu Wojskowego, Służbie Kontrwywiadu Wojskowego oraz Centralnym Biurze Antykorupcyjnym. Stanowisko Trybunału tylko częściowo było zgodne z wnioskami rzecznika praw obywatelskich i prokuratura generalnego. Zakwestionowano część przepisów, jednocześnie uznając część z nich za zgodną z Konstytucją. Trybunał uznał m.in., że konstytucyjną ochroną objęte są wszelkie sposoby przekazywania wiadomości, w każdej formie komunikowania się, bez względu na ich fizyczny nośnik.

Ochrona konstytucyjnych wolności i praw w związku z korzystaniem z Internetu i innych elektronicznych sposobów porozumiewania się na odległość nie różni się od ochrony dotyczącej tradycyjnych form komunikowania się, czy też innej aktywności. Z punktu widzenia zasady określoności prawa i ustawowej formy ograniczeń konstytucyjnych wolności i praw nie jest bezwzględnie konieczne unormowanie w każdej ustawie zamkniętego katalogu środków technicznych kontroli operacyjnej.

[2] Informacja o wynikach kontroli „Uzyskiwanie i przetwarzanie przez uprawnione podmioty danych z bilingów, informacji o lokalizacji oraz innych danych, o których mowa w art. 180 c i d ustawy Prawo telekomunikacyjne", NIK, Warszawa 2013.

Najwyższa Izba Kontroli wypowiedziała się w 2014 r. także na temat funkcjonowania w Polsce miejskiego monitoringu wizyjnego. W raporcie NIK [NIK, 2014] zauważono, że obserwacja oraz rejestrowanie obrazu zdarzeń w otwartej przestrzeni publicznej nie może naruszać podstawowych praw obywatelskich, gdyż każdy ma prawo do poszanowania i ochrony życia prywatnego. Ingerencja władzy publicznej w korzystanie z tych praw jest dopuszczona wyłącznie w przypadkach przewidzianych przez ustawę i koniecznych w demokratycznym społeczeństwie z uwagi na bezpieczeństwo państwowe, bezpieczeństwo publiczne lub dobrobyt gospodarczy kraju, ochronę porządku i zapobieganie przestępstwom, ochronę zdrowia i moralności lub ochronę praw i wolności osób. W wyniku kontroli przeprowadzonej w latach 2010–2013 stwierdzono, że w Polsce brak jest kompleksowych unormowań zasad instalowania i prowadzenia systemów monitoringu przez instytucje państwowe, samorządowe oraz podmioty prywatne. Stan ten, wobec praw i wolności obywatelskich zagwarantowanych w Konstytucji RP, w tym prawa do ochrony prywatności, ochrony danych osobowych, wymaga wprowadzenia rozwiązań rangi ustawowej, dotyczących budowy i funkcjonowania systemów monitoringu wizyjnego.

PODSUMOWANIE

Opisany w artykule przeprowadzony test pokazuje, że używając pojęcia *big data* powszechnie najczęściej rozumie się pod nim społeczne konsekwencje przetwarzania masowych danych osobowych, w szczególności problem naruszenia prywatności oraz inwigilacji. Nowy potencjał w zakresie technologii informatycznych oraz metod analitycznych z jednej strony pozytywnie wpływa na możliwości biznesowe oraz narzędzia służące zapewnieniu bezpieczeństwa funkcjonowania państwa, z drugiej jednak strony niesie wspomniane zagrożenia.

W celu minimalizacji tych zagrożeń niezbędne jest odpowiednie prawo precyzujące dopuszczalne granice ingerencji w prywatność. Dotyczy to zarówno ograniczenia prywatności wynikającego z prowadzenia działalności gospodarczej, jak i z zapewnienia funkcjonowania państwa jako całości i jego poszczególnych organów. Nie można zawsze stawiać ochrony prywatności jako priorytetu przy stanowieniu prawa, niezbędna jest natomiast jednoznaczność ustalonych zasad i zapewnienie ich przestrzegania.

BIBLIOGRAFIA

Big Data for the Enterprise, *An Oracle White Paper*, June 2013.

Big Data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute 2011.

Dyrektywa 95/46/WE Parlamentu Europejskiego i Rady z dnia 24.10.1995 w sprawie ochrony osób fizycznych w zakresie przetwarzania danych osobowych i swobodnego przepływu tych danych.

Economist 2013, *A new goldmine. Making official data public could spur lots of innovation*, http://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innovation-new-goldmine (dostęp: 30.10.2014).

https://mac.gov.pl/dzialania/michal-boni-o-big-data-na-warsaw-international-media-summit (dostęp: 30.10.2014).

Laney D., 2001, *Application delivery strategies*, META Group, Stamford.

McKinsey, 2011, *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute.

Ministerstwo Administracji i Cyfryzacji 2013, https://mac.gov.pl/dzialania/michal-boni-o-big-data-na-warsaw-international-media-summit (dostęp: 30.10.2014).

NIK 2013, Informacja o wynikach kontroli „Uzyskiwanie i przetwarzanie przez uprawnione podmioty danych z bilingów, informacji o lokalizacji oraz innych danych, o których mowa w art. 180 c i d ustawy Prawo telekomunikacyjne", Najwyższa Izba Kontroli, Warszawa.

NIK 2014, Informacja o wynikach kontroli „Funkcjonowanie miejskiego monitoringu wizyjnego", Najwyższa Izba Kontroli, Warszawa.

Oracle, *Big Data for the Enterprise*, An Oracle White Paper June 2013 http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf (dostęp: 30.10.2014).

Pawełoszek-Korek I., 2008, *Zastosowanie eksploracji danych w celu pozyskiwania wiedzy z Internetu. Fenomen Internetu*, t. II, Wydawnictwo Hogben, Szczecin.

PcMag, http://www.pcmag.com/encyclopedia/term/62849/big-data (dostęp: 30.10.2014).

Polańska K., Wassilew A., *Analizy Big Data w serwisach społecznościowych* (w druku).

Stępniak C., 2013, *Kierunki wykorzystania systemów monitoringu miejskiego w zarządzaniu rozwojem miast*, Roczniki Kolegium Analiz Ekonomicznych SGH, z. 29, Warszawa.

Szymielewicz K., Szumańska M., 2013, *Dostęp państwa do danych użytkowników usług internetowych, Siedem problemów i kilka hipotez*, Fundacja Panoptykon, Warszawa.

Trybunał Sprawiedliwości Unii Europejskiej, Komunikat prasowy nr 70/14, Luksemburg 2014, http://curia.europa.eu/jcms/upload/docs/application/pdf/2014-05/cp140070pl.pdf (dostęp: 30.10.2014).

Ustawa z dnia 29.08.1997 r. o ochronie danych osobowych (Dz.U. z 1997 r., nr 133, poz. 883).

Wieczorkowski J., Polak P., 2014, *Big data: Three-aspect approach*, "Online Journal of Applied Knowledge Management", Vol. 2.

Wieczorkowski J., 2014, *Wykorzystanie koncepcji big data w administracji publicznej*, Roczniki Kolegium Analiz Ekonomicznych SGH, z. 33, Warszawa.

www.pcmag.com/encyclopedia/term/62849/big-data (dostęp: 30.10.2014).

Streszczenie

Popularny w ostatnim czasie termin *big data* dopiero się kształtuje i na chwilę obecną obejmuje dość szeroki zakres pojęciowy. Autor opierając się na tematyce publikowanych w ostatnim okresie w polskich gazetach artykułów, które wykorzystywały omawiane pojęcie, dokonał próby charakterystyki powszechnego zrozumienia terminu *big data*. Wyniki tego badania pozwalają na stwierdzenie, że powszechnie pod tym terminem rozumie się w szczególności problematykę prywatności i zagrożenia permanentną inwigilacją, jako konsekwencję możliwości masowego przetwarzania danych. Tak więc, mimo że ogólnie przyjęte wyjaśnienie terminu *big data* opiera się na aspektach technologicznych metod przetwarzania danych masowych, odbiega ono znacząco od powszechnego jego odbioru.

Zagadnienia związane z problemami przetwarzania przede wszystkim różnego typu danych osobowych określone zostały w artykule jako społeczny aspekt *big data*. Jest on ściśle związany z zagadnieniami prawnymi, gdyż prawo, usiłując nadążyć za postępem technologicznym i potrzebami biznesu, próbuje określić dopuszczalne granice przetwarzania danych. Zagadnienia społeczne i prawne w koncepcji *big data* są podstawowym tematem artykułu.

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

Big Data – Social and Legal Issues

Summary

The term *big data* is very popular recently. The concept is new and is shaping up. The author conducted a study into the usage of the term big data in popular media. Generally, the analyzed texts focused mainly on the danger of surveillance and the threats to privacy resulting from the analysis of web content, including social networks, by the private sector. This is a social aspect of *big data*.

The paper proposes a three-faceted explanation of the term, by distinguishing three basic aspects of big data: technological (including the opportunities offered by IT and modern analytical methods), business (including a variety of applications of the concept) and social (focusing on the consequences of its implementation). Nonetheless the paper is focused on the social aspect – the risks associated with the mass processing of personal data. It is related to an additional legal sub-aspect. The social aspect with the legal sub-aspect are dependent in relation to the other aspects: technological and business.

Keywords: *big data*, data analysis, personal data, privacy

JEL: C81, Y10, C23, D82, D12