

JĘDRZEJ WIECZORKOWSKI

Instytut Informatyki i Gospodarki Cyfrowej  
Szkoła Główna Handlowa w Warszawie

## Wykorzystanie koncepcji *big data* w administracji publicznej

### 1. Wstęp

Termin *big data* w ostatnich 2–3 latach stał się jednym z najpopularniejszych w dziedzinie technologii informatycznych. Może on być niekiedy interpretowany jako dalszy etap ewolucyjnego rozwoju przetwarzania analitycznego określonego jako *business intelligence*. Jednak ciągle kształtująca się koncepcja *big data* charakteryzuje się dodatkowymi cechami, odróżniającymi ją od dotychczasowego rozumienia *business intelligence*. Dotąd idea *business intelligence* miała w praktyce szersze zastosowanie komercyjne niż w administracji publicznej. Jednocześnie administracja publiczna zawsze w dużym stopniu wykorzystywała ogromne wolumeny danych (np. w statystyce publicznej, w rejestrach publicznych), technologia ograniczała jednak możliwości ich zastosowania w analizie danych. Celem artykułu jest przedstawienie perspektyw koncepcji *big data* w administracji publicznej i usługach publicznych oraz próba systematyzacji jej zastosowań.

### 2. Istota koncepcji *big data*

Ponieważ pojęcie *big data* jest stosunkowo nowe, trudno o jednoznaczną i powszechnie akceptowaną definicję, co więcej – w dyskusjach można spotkać się z opinią, że aktualnie pojęcia tego nie można jeszcze jednoznacznie zdefiniować. Zazwyczaj, opisując zagadnienie, przytacza się kilka „v”, czyli cech

charakteryzujących *big data* rozpoczynających się na tę literę. Zalicza się do nich co najmniej *volume*, *velocity*, *variety*, często dodając kolejne – *value*, *veracity*. Trzy pierwsze pochodzą jeszcze z raportu opublikowanego w 2001 r. przez META Group<sup>1</sup>, doradczej firmy przejętej później przez Gartner Inc. Za pomocą modelu 3V scharakteryzowano wpływ ówczesnych przemian gospodarczych, w szczególności rozwoju *e-commerce* i trendów globalizacyjnych, na wyzwania stojące przed IT. Zgrupowano je w postaci trzech wymiarów – 3V. Choć nie użyto wtedy pojęcia *big data*, raport można uznać za trafną próbę opisu zjawisk, które miały wystąpić i ulec intensyfikacji w kolejnych latach.

Do dziś jako cechy charakterystyczne *big data* powszechnie zalicza się więc elementy modelu 3V:

- Przetwarzanie dużych wolumenów danych (***volume***).
- Dużą zmienność i dynamikę przetwarzanych danych (***velocity***), w szczególności przetwarzanych w czasie zbliżonym do rzeczywistego (np. danych sensorycznych, strumieniowych, pochodzących z Internetu).
- Dużą różnorodność danych (***variety***), w szczególności danych nieustrukturyzowanych.

Pozostałe „v” dodawane do klasycznego modelu dotyczą:

- Wartości danych (***value***) na potrzeby podejmowania decyzji w organizacji.
- Wiarygodności danych (***veracity***), w tym problemów ich jakości wynikającej z przetwarzania danych pochodzących z różnorodnych źródeł.

Koncepcja *big data* jest więc pewną odpowiedzią na **problem przeciążenia informacyjnego** (*information overload*), który choć głównie istnieje w kontekście biznesu, w praktyce dotyczy też sektora publicznego. Przeciążenie informacyjne można opisać jako sytuację, w której posiada się dostęp do wystarczającej ilości istotnych dla danej organizacji informacji, ale nie ma się możliwości ich przetworzenia i przeprowadzenia procesu wnioskowania. Może więc nastąpić nadmiar informacji prawdziwej, lecz w praktyce nieprzydatnej ze względu na brak możliwości jej wykorzystania<sup>2</sup>.

W zakresie **technologii informatycznych** praktyczne wykorzystanie koncepcji *big data* wiąże się z budową komputerów o coraz większej mocy obliczeniowej, w szczególności z tzw. superkomputerami i wykorzystaniem podejścia HPC (*High-Performance Computing*), także z wykorzystaniem przetwarzania typu *in-memory*.

<sup>1</sup> D. Laney, *Application Delivery Strategies*, META Group, Stamford 2001.

<sup>2</sup> Por. J. Wieczorkowski, M. Dałek, *Problem przeciążenia informacyjnego a integracja systemów informatycznych*, w: *Europejska przestrzeń komunikacji elektronicznej*, t. 1, „Zeszyty Naukowe”, nr 762, „Ekonomiczne Problemy Usług”, nr 104, Uniwersytet Szczeciński, Szczecin 2013, s. 439–448.

Jednocześnie koniecznością staje się rozproszenie przetwarzania i wykorzystanie rozproszonej infrastruktury, w tym koncepcji przetwarzania w chmurze (*cloud computing*) oraz paradygmatu MapReduce wraz z rozwiązaniem Apache Hadoop. Typowe jest również inne niż przy przetwarzaniu transakcyjnym podejście do baz danych, w szczególności stosowanie rozwiązań typu NoSQL i baz kolumnowych.

W przypadku prób zastosowania niektórych z powyższych technologii związanych z koncepcją *big data* w administracji publicznej można spodziewać się licznych barier o charakterze mentalnym i prawnym. Problematyczne jest przede wszystkim rozproszenie przetwarzania w chmurze. W badaniach dotyczących tego typu barier można napotkać m.in. kwestie braku zaufania do bezpieczeństwa danych oraz nieokreślonej lokalizacji danych (w szczególności prawnie chronionych danych osobowych)<sup>3</sup>. Prawdopodobna jest więc podczas wdrażania rozwiązań *big data* w administracji publicznej konieczność pokonywania analogicznych problemów.

Próbując wskazać, co jest faktycznie nowego w koncepcji *big data* w stosunku do wcześniej stosowanych rozwiązań informatycznych (przykładowo *business intelligence, data mining*), należy – zdaniem autora – zwrócić szczególną uwagę na przetwarzanie w czasie rzeczywistym lub do niego zbliżonym, a także na przetwarzanie danych słabo ustrukturyzowanych. Sam wzrost ilości przetwarzanych danych można uznać jedynie za ewolucyjny efekt postępu technologicznego i dostępu do coraz większych zbiorów, niemożliwe jest więc określenie granicy dla *big*. Bardzo duże wolumeny danych były dotąd także wykorzystywane zarówno w przetwarzaniu transakcyjnym OLTP (*On-Line Transaction Processing*), jak i w przetwarzaniu analitycznym OLAP (*On-Line Analytical Processing*). Należy jednak zaznaczyć, że w analitycznym przetwarzaniu dużych wolumenów podstawowym, powszechnie dotąd stosowanym podejściem była budowa dobrze ustrukturyzowanych, wydzielonych od przetwarzania transakcyjnego hurtowni danych, aktualizowanych okresowo w procesie ETL (*extraction, transformation, loading*).

Obok narzędzi wspomagających budowę hurtowni danych rozwijane były narzędzia i **metody analizy danych** oparte przede wszystkim na koncepcji *data mining*. Wywodzi się ona w znacznym stopniu z uczenia maszynowego i jest

---

<sup>3</sup> J. Cypryjański, *Rozwój zastosowań chmury obliczeniowej w administracji publicznej – prognozy, bariery, korzyści*, „Roczniki” KAE, z. 29, Oficyna Wydawnicza SGH, Warszawa 2013, s. 79–90; M. Kamiński, *Przetwarzanie w chmurze a ochrona danych osobowych*, „IT w Administracji” 2013, nr 6.

powszechnie stosowana w biznesie do rozwiązywania takich problemów, jak modelowanie ryzyka, wykrywanie oszustw (*fraud detection*), przewidywanie odejść klientów (*churn*) i prognozowanie obciążeń systemów elektroenergetycznych. Głównym zadaniem metod *data mining* jest odkrywanie niejawnych zależności oraz wzorców w dużych zbiorach danych<sup>4</sup>. Jedną z cech *data mining*, podobnie jak *big data*, jest operowanie na dużych wolumenach danych. Często jednak takie analizy opierają się na ograniczonej reprezentatywnej próbie, co w wielu zastosowaniach nie jest wystarczające. Zdaniem autora, termin *big data* wskazuje więc na nowe tendencje będące istotną zmianą jakościową w stosunku do dotychczas stosowanych metod przetwarzania, w szczególności podkreślając możliwość prowadzenia analiz w czasie rzeczywistym bezpośrednio na transakcyjnych danych źródłowych.

Jeśli przyjąć, że koncepcja *big data* jest rozwinięciem koncepcji *business intelligence* i *data mining*, naturalne wydają się liczne jej zastosowania w sektorze komercyjnym, przykładowo: w telekomunikacji, bankowości, ubezpieczeniach, handlu detalicznym, reklamie. W dalszej części artykułu zostaną jednak przeanalizowane możliwości jej zastosowań w administracji i sektorze publicznym. Administracja publiczna w zasadzie od zawsze była związana z wykorzystywaniem ponadprzeciętnie dużych zbiorów danych, w szczególności dotyczy to statystyki publicznej oraz publicznych rejestrów.

### 3. Statystyka publiczna i rejestry publiczne

**Statystyka publiczna** traktowana jest od dawna jako niezbędny element systemu informacyjnego państwa. Historia polskiego Głównego Urzędu Statystycznego sięga 1918 r.<sup>5</sup> Obecnie jego działanie jest regulowane przez ustawę z dnia 29 czerwca 1995 r. o statystyce publicznej (Dz. U. z 1995 r. Nr 88, poz. 439). Zadania statystyki publicznej w oczywisty sposób wiążą się z przetwarzaniem bardzo dużej ilości danych wykorzystywanych m.in. do celów wspomagania decyzji na poziomie zarządzania państwem. Trudno aktualnie określić sposób funkcjonowania statystyki publicznej mianem *big data*, przede wszystkim z uwagi na duże opóźnienie dostępności danych wynikowych w stosunku do okresów,

---

<sup>4</sup> R. Szupiluk, *Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli data mining*, Oficyna Wydawnicza SGH, Warszawa 2013.

<sup>5</sup> <http://www.stat.gov.pl>.

których dotyczą, a także ze względu na dobrze ustrukturyzowany charakter danych. Można jednak zaobserwować elementy ewolucji statystyki publicznej w kierunku przetwarzania w czasie bardziej zbliżonym do rzeczywistego, stąd *big data* wydaje się dość naturalnym kierunkiem dalszej jej ewolucji.

Ze statystyką publiczną wiążą się **rejestry publiczne**, za których prowadzenie w Polsce odpowiadają różne jednostki organizacyjne administracji, w tym GUS. Pojęcie rejestru nie jest w pełni jednoznaczne, temat ten został przedstawiony przez T. Staweckiego<sup>6</sup>, który dokonał z prawnego, historycznego i językowego punktu widzenia porównania takich pojęć, jak: rejestr, ewidencja, wykaz, lista, księga. Zaproponował również definicję, w której rejestr publiczny to zbiór informacji o osobach, rzeczach lub prawach, spełniający następujące warunki:

- Został utworzony zgodnie z przepisami prawa obowiązującego.
- Jest prowadzony przez organ rejestrowy o charakterze publicznym.
- Przyjęcie, utrwalenie, a następnie ujawnienie określonych w nim informacji następuje co do zasady w wyniku podjęcia przez organ rejestrowy decyzji.
- Prowadzenie rejestru i ujawnianie w nim określonych informacji rodzi skutki prawne zarówno dla osoby, której wpis dotyczy, jak i dla organów władzy publicznej.
- Jest jawny, czyli dostęp do niego oprócz organu rejestrowego mają przynajmniej osoby, których rejestr dotyczy, oraz inne organy władzy publicznej, a co do zasady szeroka kategoria publiczności. Jawność może być pełna lub ograniczona.

Opierając się z kolei na definicji pojęcia sformułowanej na podstawie ustawy z dnia 17 lutego 2005 r. o informatyzacji podmiotów realizujących zadania publiczne (Dz. U. z 2005 r. Nr 64, poz. 565 z późn. zm.), można powiedzieć, że rejestr publiczny to rejestr, ewidencja, wykaz, lista, spis albo inna forma ewidencji służąca do realizacji zadań publicznych, prowadzona przez podmiot publiczny na podstawie odrębnych przepisów ustawowych.

W kontekście niniejszej pracy wygodnym źródłem wiedzy o rejestrach publicznych (oraz innych systemach informacyjnych, nie wchodząc w rozważania definicyjne rejestru publicznego) jest prowadzona przez GUS ewidencja systemów informacyjnych administracji publicznej będących źródłem danych w trakcie badań statystyki publicznej<sup>7</sup>. Zawiera ona ponad 600 pozycji, dla każdego sys-

---

<sup>6</sup> T. Stawecki, *Rejestry publiczne. Funkcje instytucji*, LexisNexis, Warszawa 2005, s. 19–31.

<sup>7</sup> *Systemy informacyjne administracji publicznej – źródła danych dla badań statystyki publicznej*, GUS, 2013.

temu przypisano gestora, instytucję administrującą, akty prawne stanowiące podstawę funkcjonowania systemu. Opisany został cel, sposób prowadzenia, zakres informacyjny, źródła danych wraz z częstotliwością aktualizacji, termin i formy udostępniania danych oraz stan wdrożenia. Opisane systemy informacyjne są bardzo różnej wielkości, począwszy od wielkich systemów zaliczanych do ogólnokrajowych referencyjnych systemów identyfikacji, na niewielkich, specjalistycznych rejestrach prowadzonych przykładowo przez jednostki samorządowe skończywszy. Zdecydowana większość systemów jest prowadzona w formie elektronicznej lub jako połączenie formy elektronicznej i papierowej. Jedynie kilkadziesiąt niewielkich systemów informacyjnych jest prowadzonych wyłącznie w formie papierowej.

Przykładem największych elektronicznych ogólnokrajowych systemów identyfikacji jest rejestr ludności, czyli liczący kilkadziesiąt milionów podstawowych rekordów Powszechny Elektroniczny System Ewidencji Ludności (PESEL). Jego celem jest wspomaganie informatyczne procesów zachodzących w sferze ewidencji ludności, w tym automatyzacja obsługi spraw osobowych w urzędach gminnych oraz zasilania rejestrów ewidencji ludności i udostępniania danych osobowych na szczeblu centralnym, wojewódzkim i gminnym. Zawiera dane (aktualne i archiwalne informacje osobowe i adresowe) obywateli polskich i cudzoziemców, którym numer PESEL został nadany na wniosek organu gminy, płatnika składek ubezpieczeniowych, polskiego konsula. Źródłem zasilania PESEL są urzędy gminne i wojewódzkie przekazujące dane osobowe na bieżąco. Dane osobowe udostępnia się osobom lub podmiotom uprawnionym w formie pisemnej lub za pomocą środków komunikacji elektronicznej oraz za pomocą urządzeń teletransmisji danych lub w drodze weryfikacji. Dane są przekazywane w formie odpowiadającej złożonemu wnioskowi<sup>8</sup>. Do innych rejestrów publicznych o charakterze identyfikacyjnym należą przykładowo: Krajowy Rejestr Urzędowy Podmiotów Gospodarki Narodowej (REGON), Ewidencja Gruntów i Budynków, Krajowy Rejestr Sądowy (KRS), Centralny Rejestr Podmiotów Krajowej Ewidencji Podatników (CRP KEP).

Można postawić pytanie, czy dotychczasowe wykorzystywanie takich referencyjnych systemów identyfikacji można uznać za przetwarzanie typu *big data*. Przemawia, być może, za tym wielkość przetwarzanych zbiorów, lecz ich struktura i sposób przetwarzania nie spełniają wymogów zaliczenia ich do *big data*. Typowe dotychczasowe zastosowanie takich systemów można w większości przypadków uznać za tradycyjne przetwarzanie transakcyjne, spełniające

---

<sup>8</sup> Ibidem.

wymogi przetwarzania wsadowego lub OLTP. Rejestry takie są zazwyczaj dobrze ustrukturyzowane i przechowywane w bazach danych w postaci tabelarycznej. Ich konstrukcja i sposób wykorzystania są charakterystyczne dla systemów ewidencyjnych. Nie są one zazwyczaj podstawą do zaawansowanych analiz.

Jednocześnie jednak należy wskazać kierunek, w którym w przyszłości zapewne będzie podążać zastosowanie tego typu baz danych i systemów, tj. wykorzystanie ich jako źródła danych dla różnego typu analiz wykonywanych przez systemy informatyczne instytucji publicznych, a być może także innych zewnętrznych organizacji. Możliwości wykorzystania rejestrów publicznych do celów analitycznych wiążą się z ich strukturą oraz wzajemnymi zależnościami. Problem ten został zauważony także w polskim prawie, w którym w ustawie o informatyzacji podmiotów realizujących zadania publiczne zdefiniowano pojęcie interoperacyjności, rozumianej jako zdolność różnych podmiotów oraz używanych przez nie systemów teleinformatycznych i rejestrów publicznych do współdziałania na rzecz osiągnięcia wzajemnie korzystnych i uzgodnionych celów, z uwzględnieniem współdzielenia informacji i wiedzy przez wspierane przez nie procesy biznesowe realizowane za pomocą wymiany danych za pośrednictwem wykorzystywanych przez te podmioty systemów teleinformatycznych.

Przykładowo, w zakresie usług medycznych w polskim systemie prawnym funkcjonują grupy rejestrów podmiotowych (z danymi dotyczącymi takich obiektów, jak: zakłady opieki zdrowotnej, apteki, lekarze, produkty lecznicze, ubezpieczeni) oraz przedmiotowych (dane statystyczne o wystąpieniach konkretnych chorób). Rejestry te muszą komunikować się wzajemnie oraz z rejestrami referencyjnymi, takimi jak PESEL, TERYT (Krajowy Rejestr Urzędowy Podziału Terytorialnego Kraju), CRP KEP<sup>9</sup>. Interoperacyjność jest więc w obszarze usług medycznych podstawą do zapewnienia możliwości przeprowadzania statystycznych analiz w zakresie medycyny. W przypadku nadmiernej złożoności systemów informacyjnych sektora publicznego oraz braku interoperacyjności (wymuszonych przez niskiej jakości system prawny lub przez błędy projektowe przy budowie rejestrów) może być znacząco utrudniona funkcja analityczna tych systemów. Nadmierna autonomizacja i dezintegracja systemów oraz ograniczanie wymiany danych pomiędzy nimi prowadzi do niepotrzebnej redundancji, zagraża spójności danych i w konsekwencji obniża ich jakość. W takim stanie wykorzystanie systemów informatycznych do różnych zastosowań analitycznych typu *big data* jest wyraźnie utrudnione.

---

<sup>9</sup> T. Górski, *Symulacyjne środowisko badania wydajności platformy integracyjnej rejestrów medycznych*, „Roczniki” KAE, z. 29, Oficyna Wydawnicza SGH, Warszawa 2013, s. 595–610.

Problem wykorzystania rejestrów publicznych do celów analitycznych wiąże się także z fundamentalnym pytaniem, na ile informacje gromadzone przez instytucje publiczne powinny być powszechnie dostępne, czyli – czy i w jakim stopniu prawo powinno zezwalać na ponowne wykorzystywanie tego, co powstało za pieniądze publiczne. Obecnie w USA toczy się dyskusja na temat traktowania informacji, które zostały zgromadzone przez instytucje publiczne, jako powszechne i ogólnie dostępne dobro, pod warunkiem że informacje te nie naruszają prywatności i bezpieczeństwa<sup>10</sup>. Jest tutaj ewentualnie miejsce na działalność prywatnych firm doradczych zajmujących się komercyjnym wyszukiwaniem i przetwarzaniem publicznie dostępnych danych. Z drugiej strony pozostaje kluczowe pytanie o granice pożądanego poziomu ochrony jawności informacji obejmujących przykładowo dane osobowe i dane podmiotów gospodarczych. Powyższe kwestie prawne stanowią niewątpliwie potencjalne istotne bariery dla rozwoju *big data* w sektorze publicznym.

#### 4. Kierunki zastosowań koncepcji *big data* w administracji publicznej i usługach publicznych

Zakładając dostępność potrzebnych danych, należy zastanowić się nad pytaniem o potencjalne zastosowania koncepcji *big data* w administracji publicznej i usługach publicznych. Temat ten na przykładzie możliwości wykorzystania *big data* przez urzędy administracji publicznej w USA został poruszony w przygotowanym przez TechAmerica Foundation raporcie, w którym wymieniono następujące zastosowania *big data*<sup>11</sup>:

- poprawa jakości i efektywności funkcjonowania służby zdrowia,
- monitorowanie istotnych danych w celu wczesnego wykrywania zagrożeń epidemiologicznych i sanitarnych,
- wspomaganie zarządzania transportem,
- ocena i poprawa jakości systemu edukacji,
- wykrywanie nieprawidłowości w finansowaniu służby zdrowia,
- zapewnienie bezpieczeństwa w cyberprzestrzeni,

---

<sup>10</sup> <http://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innovation-new-goldmine>.

<sup>11</sup> *Demystifying Big Data: A Practical Guide To Transforming The Business of Government*, TechAmerica Foundation, Washington 2012.



- wykrywanie nieprawidłowości w rozliczeniach podatkowych,
- lepsze prognozowanie pogody i przewidywanie klęsk żywiołowych,
- analiza rynku pracy i przeciwdziałanie bezrobociu.

W innym raporcie McKinsey Global Institute bada, w celu oszacowania potencjalnej wartości zastosowania podejścia *big data* w sektorze administracji publicznej w Unii Europejskiej, trzy główne ogólne kategorie korzyści<sup>12</sup>:

- oszczędności wynikające z poprawy efektywności działań operacyjnych,
- redukcję kosztów wynikających z oszustw i błędów,
- wzrost poboru dochodów podatkowych.

Powyższe listy ogólnych idei lub szczegółowych zastosowań podejścia *big data* mogą być traktowane jako zestaw pomysłów, które należy spróbować usystematyzować. Zdaniem autora, pierwszą grupą zastosowań, w której można już dziś zaobserwować faktyczne próby wykorzystania koncepcji *big data*, jest **wykrywanie różnego typu nadużyć**, w szczególności finansowych. W najbliższej przyszłości można prognozować dalszy rozwój zastosowań takich rozwiązań w zakresie poprawy efektywności funkcjonowania służb podatkowych i celnych (np. badanie szarej strefy w gospodarce – w szczególności przez analizę handlu internetowego, odnajdowanie majątku niemającego odzwierciedlenia w wykazywanych dochodach), ograniczenia wyłudzeń zasiłków (np. chorobowych, dla bezrobotnych), wykrywania prób niedozwolonego wpływania na rynek papierów wartościowych.

Drugą intensywnie rozwijającą się grupą zastosowań podejścia *big data* jest **bezpieczeństwo publiczne**. Dotyczy to zwłaszcza nowych możliwości związanych ze śledzeniem Internetu, które otrzymuje policja i służby specjalne w zakresie infiltracji środowisk mogących stanowić zagrożenie dla państwa i porządku publicznego. Wykorzystywane są także coraz częściej sensoryczne dane o charakterze nieustrukturyzowanym, jak obraz z systemów monitoringu<sup>13</sup>. Współczesne możliwości technologiczne dają też szersze perspektywy przetwarzania różnorodnych danych historycznych w celu wyszukiwania miejsc oraz terminów ponadprzeciętnie zagrożonych i w konsekwencji przeciwdziałania przestępczości. Stosunkowo nową koniecznością w zakresie bezpieczeństwa publicznego jest zapewnienie bezpieczeństwa cyberprzestrzeni, w szczególności

---

<sup>12</sup> J. Manyika et al., *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, 2011, s. 129–130.

<sup>13</sup> Por. C. Stępniaak, *Kierunki wykorzystania systemów monitoringu miejskiego w zarządzaniu rozwojem miast*, „Roczniki” KAE, z. 29, Oficyna Wydawnicza SGH, Warszawa 2013, s. 295–307.

funkcjonowania Internetu. Wymaga ono realizowanej w czasie rzeczywistym ciągłej obserwacji sieci.

Zastosowania *big data* w dziedzinie wykrywania nadużyć i ochrony bezpieczeństwa publicznego wiążą się ściśle z zagadnieniem dopuszczalnego poziomu **inwigilacji społeczeństwa**. Powszechnie wiadomo, że na całym świecie podejmowane są realne próby wykorzystania tej koncepcji, lecz ze zrozumiałych względów informacje na ten temat są mocno ograniczone. W odpowiedzi na pytanie, jak wygląda ta kwestia w Polsce, pomaga raport opracowany przez Fundację Panoptykon<sup>14</sup>, która na podstawie działalności kilku dużych polskich firm świadczących usługi drogą elektroniczną szacuje skalę i strukturę otrzymywanych od różnych organów państwa oficjalnych zapytań o dane użytkowników. Choć w badanym okresie (od stycznia 2012 r. do czerwca 2013 r.) liczba zapytań stale rosła, nie przekraczała jednak kilku tysięcy rocznie. Ponieważ większość zapytań pochodziła od prokuratury (63%) i policji (33%), a znacząco mniejszy był udział różnych służb specjalnych (łącznie 2%), można przypuszczać, że zapytania głównie dotyczyły konkretnych postępowań karnych. Zakładając, że państwo nie stosuje metod niejawnego i pozaprawnego sięgania po dane, można stwierdzić, że – przynajmniej w zakresie zapytań kierowanych do dostawców usług internetowych przez organy państwa – nie można na razie mówić w Polsce o masowej inwigilacji opartej na koncepcji *big data*.

Kolejną grupą są zastosowania *big data* w zakresie dostarczania informacji wspomagających **zarządzanie państwem**, w skali zarówno całej gospodarki, jak i rozwiązywania konkretnych, szczegółowych, często lokalnych problemów. Od wielu lat, jak to wcześniej zauważono, podstawą do zarządzania gospodarką są dane pochodzące ze statystyki publicznej. Koncepcja *big data* powinna w przyszłości zapewnić dostępność użytecznych danych pochodzących z danych statystyki publicznej w czasie zbliżonym do rzeczywistego. Dotyczy to nie tylko wspomagania zarządzania państwem na szczeblu centralnym, lecz także samorządowym. Do tej grupy zastosowań można zaliczyć wspomnianą aktywizację zawodową i przeciwdziałanie bezrobociu na podstawie aktualnych danych statystyki publicznej. W skali świata można odnaleźć próby szerokiego zastosowania podejścia *big data* na poziomie samorządów. Przykładowo, w Nowym Jorku powstało stanowisko Chief Analytics Officer. Jednostka ta ma

---

<sup>14</sup> K. Szymielewicz, M. Szumańska, *Dostęp państwa do danych użytkowników usług internetowych. Siedem problemów i kilka hipotez*, Fundacja Panoptykon, Warszawa 2013.

dbać o to, aby dane pochodzące ze wszystkich departamentów i biur mogły być wykorzystane w efektywny sposób, tak by urzędnicy mogli podejmować na ich podstawie decyzje. Na dalszym etapie jej zadaniem będzie zapewnienie mieszkańcom dostępu do potencjalnie przydatnych danych publicznych<sup>15</sup>.

Koncepcja *big data* daje szerokie możliwości w **zarządzaniu usługami publicznymi**. Ta grupa zastosowań obejmuje przede wszystkim służbę zdrowia. Jest ona chętnie opisywana w różnorodnych publikacjach jako potencjalne miejsce zastosowania podejścia *big data*<sup>16</sup>. W medycynie podejmuje się coraz więcej prób diagnozowania z wykorzystaniem szerokiej bazy danych historycznych. Na podstawie danych historycznych oraz bieżących można prowadzić monitorowanie zagrożeń epidemiologicznych. Tego typu analizy pomagają zarządzać całym systemem służby zdrowia w celu efektywnego wykorzystania środków publicznych. Usługi publiczne mogą być finansowane ze środków publicznych, lecz wykonywane przez podmioty prywatne. Często z tą sytuacją można się spotkać właśnie w przypadku służby zdrowia. Rozwiązania typu *big data* znajdują więc w praktyce także zastosowanie w stosunkowo niewielkich podmiotach prywatnych działających w systemie usług publicznych.

Innym przykładem zastosowania koncepcji *big data* do usprawniania usług publicznych jest zarządzanie transportem. Analiza danych historycznych dotyczących potoków komunikacyjnych (obejmujących dane z monitoringu, a także o charakterze geolokalizacyjnym) wspomaga przykładowo planowanie sieci dróg, a także zarządzanie transportem publicznym. Bieżąca analiza natężenia ruchu i wykrywanie w czasie rzeczywistym pojawiających się incydentów (np. wypadków drogowych) pomaga sterować ruchem.

Do grupy zastosowań *big data* we wspomaganiu usług publicznych można także zaliczyć planowanie i zarządzanie edukacją, opieką społeczną i aktywizacją zawodową, z wykorzystaniem przede wszystkim danych demograficznych pochodzących m.in. z publicznych rejestrów ewidencji ludności. Inną aktywnością realizowaną przez sektor publiczny i będącą miejscem na potencjalne szerokie wykorzystanie koncepcji *big data* jest działalność różnych archiwów i bibliotek.

---

<sup>15</sup> <http://300polityka.pl/300news/polityka-oparta-na-big-data-jak-burmistrz-bloomberg-chce-zmienic-sposob-rzadzenia-nowym-jork>.

<sup>16</sup> Por.: J. Manyika et al., op.cit., s. 39–53; P. Simon, *Too big to ignore – The business case for big data*, Wiley, Hoboken 2013, s. 147–151; V. Mayer-Schonberger, K. Cukier, *Big data – A revolution that will transform how we live, work, and think*, An Eamon Dolan Book Houghton Mifflin Harcourt, Boston 2013, s. 1–3.

## 5. Podsumowanie

Administracja publiczna ma potencjalny dostęp do bardzo dużej ilości danych zgromadzonych w postaci cyfrowej. Wymienione w artykule grupy możliwych zastosowań koncepcji *big data* należy traktować jako przykładowe, dające się przewidzieć w najbliższych latach. Widoczny jest ich silny związek z charakterystycznymi cechami przetwarzania typu *big data*, tj. bardzo dużym wolumenem przetwarzanych danych, szerokim wykorzystaniem danych nieustrukturyzowanych oraz przetwarzaniem w czasie rzeczywistym lub zbliżonym do rzeczywistego, w tym danych strumieniowych i sensorycznych. Postępujący rozwój technologii i metod numerycznych będzie zapewne wpływał na rozszerzanie możliwości zastosowań potencjalnie dostępnych różnego typu danych we wspomaganiu działalności administracji publicznej, w tym wspomaganiu usług publicznych.

## Bibliografia

1. Cypryański J., *Rozwój zastosowań chmury obliczeniowej w administracji publicznej – prognozy, bariery, korzyści*, „Roczniki” KAE, z. 29, Oficyna Wydawnicza SGH, Warszawa 2013.
2. *Demystifying big data: A practical guide to transforming the business of government*, TechAmerica Foundation, Washington 2012.
3. Górski T., *Symulacyjne środowisko badania wydajności platformy integracyjnej rejestrów medycznych*, „Roczniki” KAE, z. 29, Oficyna Wydawnicza SGH, Warszawa 2013.
4. Kamiński M., *Przetwarzanie w chmurze a ochrona danych osobowych*, „IT w Administracji” 2013, nr 6.
5. Laney D., *Application delivery strategies*, META Group, Stamford 2001.
6. Manyika J. et al., *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, 2011.
7. Mayer-Schonberger V., Cukier K., *Big data – A revolution that will transform how we live, work, and think*, An Eamon Dolan Book Houghton Mifflin Harcourt, Boston 2013.
8. Simon P., *Too big to ignore – The business case for big data*, Wiley, Hoboken 2013.
9. Stawecki T., *Rejestry publiczne. Funkcje instytucji*, LexisNexis, Warszawa 2005.

10. Stępniaak C., *Kierunki wykorzystania systemów monitoringu miejskiego w zarządzaniu rozwojem miast*, „Roczniki” KAE, z. 29, Oficyna Wydawnicza SGH, Warszawa 2013.
11. *Systemy informacyjne administracji publicznej – źródła danych dla badań statystyki publicznej*, GUS, 2013.
12. Szupiluk R., *Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli data mining*, Oficyna Wydawnicza SGH, Warszawa 2013.
13. Szymielewicz K., Szumańska M., *Dostęp państwa do danych użytkowników usług internetowych. Siedem problemów i kilka hipotez*, Fundacja Panoptykon, Warszawa 2013.
14. Ustawa z dnia 29 czerwca 1995 r. o statystyce publicznej (Dz. U. z 1995 r. Nr 88, poz. 439).
15. Ustawa z dnia 17 lutego 2005 r. o informatyzacji podmiotów realizujących zadania publiczne (Dz. U. z 2005 r. Nr 64, poz. 565 z późn. zm.).
16. Wiczorkowski J., Dałek M., *Problem przeciążenia informacyjnego a integracja systemów informatycznych*, w: *Europejska przestrzeń komunikacji elektronicznej*, t. 1, „Zeszyty Naukowe”, nr 762, „Ekonomiczne Problemy Usług”, nr 104, Uniwersytet Szczeciński, Szczecin 2013.

## Źródła sieciowe

1. <http://300polityka.pl/300news/polityka-oparta-na-big-data-jak-burmistrz-bloomberg-chce-zmienic-sposob-rzadzenia-nowym-jork>.
2. <http://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innovation-new-goldmine>.
3. <http://www.stat.gov.pl>.

\* \* \*

## Big data in public administration

### Summary

The article presents the concept of big data and differences between business intelligence and big data. The problem of using big data in public administration is discussed. The paper contains some proposals of big data applications in the public sector, i.e. in government agencies and public services.

**Keywords:** big data, business intelligence, public administration