

KORNELIA BATKO¹

Możliwości wykorzystania *Big Data* w ochronie zdrowia

1. Wstęp

W ostatnich latach coraz bardziej wzrasta zapotrzebowanie na rozwiązania oferujące efektywne narzędzia analityczne. Trend ten można zaobserwować również w zakresie analizy dużych wolumenów danych (*Big Data*). Rozwiązaniami tymi interesują się niemal wszystkie sektory, chociaż niewątpliwie w wykorzystaniu analiz tego typu przodują organizacje biznesowe. Można jednak zauważyć, że również sektor publiczny, w tym w ramach ochrony zdrowia, coraz chętniej wykorzystuje te rozwiązania².

Teoria i praktyka pokazują, iż analityka *Big Data* w tym sektorze może się przyczynić m.in. do: poprawy obsługi pacjentów, wyznaczenia i wdrożenia właściwych ścieżek (sposobów) leczenia pacjentów, wsparcia leczenia klinicznego, monitorowania bezpieczeństwa opieki zdrowotnej, wytworzenia systemów zapewnienia kontroli zarządczej czy przeciwdziałania epidemiom i innym zagrożeniom³. Jednak ze zjawiskiem *Big Data* są związane też pewnego rodzaju wyzwania, w postaci złożoności, zagrożenia bezpieczeństwa i prywatności, jak również zapotrzebowania na nowe technologie i umiejętności ludzkie.

Celem artykułu jest zaprezentowanie możliwości wykorzystania analiz *Big Data* w sektorze ochrony zdrowia. Zaprezentowano w nim istotę *Big Data*, zalety tej analityki oraz potencjalne obszary jej wykorzystania. Została również omówiona architektura systemu analitycznego dla ochrony zdrowia uwzględniająca *Big Data*.

¹ Uniwersytet Ekonomiczny w Katowicach, Wydział Ekonomii.

² http://www.dlp-expert.pl/articles/id,476/big_data_w_sluzbie_ochrony_zdrowia.html (data odczytu: 20.11.2015).

³ J. Kyoungyoung, K. Gang Hoon, *Potentiality of Big Data in the Medical Sector: Focus on How to Reshape the Healthcare System*, „Healthcare Informatics Research” 2013, vol. 19(2), s. 79–85.

2. Informatyzacja sektora ochrony zdrowia w Polsce

Polska, podobnie jak wiele krajów rozwijających się, staje w obliczu problemów związanych z przekształcaniem sektora ochrony zdrowia, takich jak: reforma finansowania ochrony zdrowia, skuteczniejsza analiza informacji związanych z opieką zdrowotną czy poprawa efektywności w świadczeniu usług medycznych. W przeszłości większość systemów opieki zdrowotnej polegała na kombinacji podejścia opartego na regułach oraz na ręcznym wychwyceniu anomalii rozliczeniowych. Po wprowadzeniu do placówek medycznych rozwiązań w postaci oprogramowania wspomagającego ich pracę sytuacja uległa zmianie. Dane były przechowywane w ustrukturalizowanych bazach i hurtowniach danych, a systemy raportujące lub rozwiązania analityczne mogły na bieżąco je analizować i wykrywać wszelkie odchylenia. Warto wspomnieć, iż od wielu lat z sukcesem wdraża się w organizacjach ochrony zdrowia systemy *Business Intelligence* oraz aplikacje analityczne. Badania ukazują jednak, że nie wykorzystują one w pełni potencjału, jaki tkwi w tego typu rozwiązaniach⁴. Z tego powodu wiele zjawisk, trendów, problemów diagnostycznych czy finansowych nie zostaje zauważonych⁵. Rozwiązanie tego problemu mają dać analizy *Big Data*.

Sektor ochrony zdrowia to sektor, w którym infrastruktura IT nie jest prosta, obejmuje ona bowiem zarówno systemy dla jednostek służby zdrowia, jak i systemy wspomagania diagnostyki, przetwarzanie i analizę sygnałów medycznych (np. EKG), przetwarzanie i analizę obrazów medycznych, oprogramowanie aparatury medycznej, standardy wymiany informacji HL7 (*Health Level Seven*), DICOM (*Digital Imaging and Communications in Medicine*), medyczne bazy danych⁶. Ogromna liczba danych oraz ich różnorodność powodują, że niemal niemożliwe jest zarządzanie organizacjami działającymi w tym sektorze bez zaawansowanych systemów informatycznych, w tym specjalistycznych dla branży oraz spełniających jednocześnie wszelkie standardy w dziedzinie informacji medycznej, bezpieczeństwa i dostępu do danych oraz zgodnych z wymaganiami NFZ.

⁴ C. Olszak, K. Batko, *The Use of Business Intelligence Systems in Healthcare Organizations in Poland*, Federated Conference on Computer Science and Information Systems, 2012, <http://ieeexplore.ieee.org/> (data odczytu: 20.09.2015), s. 997–1004.

⁵ T. Słoniewski, *Od BI do „Big Data”*, <http://it-manager.pl/wp-content/uploads/Nowa-twarz-BI1.pdf> (data odczytu: 20.11.2015).

⁶ <http://www.computerworld.pl/news/396729/ERP.w.kolorze.bialym.html> (data odczytu: 20.11.2015).

W Polsce proces informatyzacji sektora ochrony zdrowia jest opóźniony w stosunku do pozostałych krajów członkowskich Unii Europejskiej. Powstaje wiele rozwiązań zarówno ogólnokrajowych, jak i regionalnych, jednakże terminy wdrożenia docelowych rozwiązań mających przenieść sektor ochrony zdrowia w realia cyfrowe są ciągle przekładane. Zgodnie z ustawą o systemie informacji w ochronie zdrowia w najbliższym czasie do obiegu zostaną wprowadzone m.in.: elektroniczne recepty, e-skierowania, e-zlecenia i aplikacja – Internetowe Konto Pacjenta, dzięki któremu pacjent będzie miał dostęp do swojej historii leczenia on-line. Obecnie jednostki medyczne ochrony zdrowia mają czas do końca lipca 2017 r. na wdrożenie systemu Elektronicznej Dokumentacji Medycznej (EDM). Wiąże się to również z tym, że podmioty świadczące usługi lecznicze będą musiały przygotować swoje systemy informatyczne do wymiany danych w ramach platform teleinformatycznych, które jeszcze nie powstały⁷.

Wprowadzenie tych rozwiązań przybliży sektor ochrony zdrowia w Polsce do możliwości monitorowania stanu pacjenta on-line, tym bardziej że technologia rozwija się i powstają urządzenia, które mogą wspierać sektor ochrony zdrowia w osiągnięciu tego celu⁸. Zdaniem ekspertów opieka zdrowotna stanie się już wkrótce o wiele bardziej spersonalizowana⁹. Dzięki gromadzeniu danych o aktywności fizycznej, sposobie odżywiania, stylu życia danej osoby oraz przeprowadzaniu w czasie rzeczywistym badania ciśnienia krwi, cholesterolu i innych parametrów będzie również możliwe wczesne wykrycie potencjalnych zagrożeń zdrowia i reagowanie na nie¹⁰. Odpowiednie rezultaty w postaci optymalnej ścieżki postępowania terapeutycznego będzie można jednak uzyskać dopiero po przeanalizowaniu ogromnej liczby danych z urządzeń monitorujących.

3. *Big Data* w ochronie zdrowia – istota i wyzwania

W literaturze przedmiotu można znaleźć wiele różnych definicji *Big Data*. Pojęcie to ewoluowało na przestrzeni ostatnich lat i nie jest jednoznacznie

⁷ F. Smólczyński, *Cloud computing dla e-Zdrowia*, <http://di.com.pl/cloud-computing-dla-e-zdrowia-48858> (data odczytu: 21.05.2014).

⁸ P. Jankowski, *Elektroniczna Dokumentacja Medyczna. Co zmieni cyfrowa rewolucja służby zdrowia?*, <http://www.wprost.pl/ar/520172/E-selowolucja-w-sluzbie-zdrowia-Co-nas-czeka-pocyfryzacji-placowek-medycznych/> (data odczytu: 20.10.2015).

⁹ <http://manager.money.pl/ludzie/twarz-w-twarz/artykul/sluzba-zdrowia-przed-rewolucja-dane-medyczne,180,0,1717172.html> (data odczytu: 20.11.2015).

¹⁰ http://www.web.gov.pl/wiedza/587_4420.html (data odczytu: 20.11.2015).

rozumiane. Przytaczając definicję D. Laneya, za *Big Data* można uznać ogromną liczbę danych generowanych w bardzo szybkim tempie i zawierającą wiele treści. Dane te pochodzą ze źródeł nieustrukturyzowanych, takich jak: strumień kliknięć na stronach WWW, sieci społecznościowe (Twitter, blogi, Facebook), nagrania wideo ze sklepów, nagrania rozmów w call center itp., informacje w czasie rzeczywistym z różnego rodzaju czujników, sensorów, RFID, urządzeń GPS, telefonów komórkowych i innych urządzeń identyfikujących i monitorujących¹¹. *Big Data* stanowią potężne silosy cyfrowych danych, surowych, gromadzonych z różnego rodzaju źródeł, niestrukturalnych i trudnych lub wręcz niemożliwych do analizy przy użyciu konwencjonalnych technik stosowanych dotychczas do relacyjnych baz danych¹².

Opisując *Big Data*, nie można pominąć faktu, że termin ten odnosi się raczej do pewnego zjawiska niż do konkretnych technologii¹³. Dlatego zamiast definiować to zjawisko, próbuje się je opisywać, podając jego charakterystyki ujęte w zbiór 4V (ang. *volume, variety, velocity, value*), czyli objętość (duża liczba danych, wolumen danych), różnorodność (zróżnicowanie danych, dowolne typy danych), prędkość (prędkość, z jaką dane napływają do organizacji, duża zmienność, dynamiczność danych) i wartość (wartość danych, korzyści z danych, ocena wyrażana poprzez weryfikację)¹⁴. Szczególnie jest zaakcentowane to w definicji Gartnera z 2012 r., według której *Big Data* to zbiory informacji o dużej objętości, dużej zmienności lub dużej różnorodności, wymagające nowych form przetwarzania w celu wspomaganie podejmowania decyzji, odkrywania nowych zjawisk oraz optymalizacji procesów¹⁵.

Bez wątplenia zbiory *Big Data* różnią się od dotychczas wykorzystywanych przez organizacje źródeł danych, takich jak bazy danych czy hurtownie danych. Najważniejsze cechy związane z nowym rodzajem danych można podsumować następująco¹⁶:

¹¹ <http://www.chip.pl/news/wydarzenia/statystyka/2013/02/big-bang-theory-czyli-jak-eksplozja-danych-zmienia-swiat-1> (data odczytu: 20.11.2015).

¹² J. Kyoungyoung, K. Gang Hoon, op.cit.

¹³ M. Mach-Król, *Analiza i strategia Big Data w organizacjach*, „Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą” 2015, t. 74, s. 43–55.

¹⁴ N. Williams, N.P. Ferdinand, R. Croft, *Project management maturity in the age of Big Data*, „International Journal of Managing Projects in Business” 2014, vol. 7(2), s. 311–317.

¹⁵ http://www.web.gov.pl/wiedza/587_4420.html (data odczytu: 20.11.2015).

¹⁶ M. Mach-Król, *A survey and assessment of maturity models for Big Data adoption*, w: *Proc. 11th International Conference on Strategic Management and its Support by Information Systems*, Uherske Hradiste, Czech Republic, 21–22 May 2015.

- a) organizacje muszą zacząć postrzegać dane jako strumienie (*flows*), a nie zbiory (*stocks*), co pociąga za sobą konieczność wdrożenia tzw. analityki strumieniowej (*streaming analytics*);
- b) główną rolę zaczynają odgrywać tzw. *data scientists* w miejsce analityków danych, czyli osoby, które rozumieją procesy analityczne, ale również są doskonale obeznane z rozwiązaniami IT, a często mają także umiejętności w zakresie fizyki czy biologii obliczeniowej lub socjologii zorientowanej na media społecznościowe;
- c) analityka musi przenieść się z poziomu działu IT na poziom biznesu i jego funkcji operacyjnych, ponieważ *Big Data* wymaga odmiennego podejścia do technologii i procesów.

Wymienione cechy powodują konieczność zastosowania nowych narzędzi informatycznych, które pozwalają na jak najpełniejsze wykorzystanie nowych danych¹⁷. Zjawisko *Big Data*, nierozzerwalnie związane z ogromnym wzrostem liczby danych, którymi dysponują różne organizacje lub indywidualne osoby, stwarza szanse dostępu do wartościowych analiz, wniosków oraz podejmowania bardziej trafnych decyzji¹⁸. Koncepcja *Big Data* cały czas ewoluuje i w chwili obecnej nie skupia się już uwagi na ogromnej liczbie danych, a raczej na procesie tworzenia z nich wartości.

W sektorze ochrony zdrowia na strumienie *Big Data* składają się różne rodzaje danych, a w szczególności dane¹⁹:

- a) kliniczne, czyli pozyskiwane z elektronicznej dokumentacji medycznej, ze szpitalnych systemów informatycznych, centrów obrazowych, laboratoriów, aptek oraz innych organizacji świadczących usługi zdrowotne;
- b) biometryczne dostarczane z różnego rodzaju urządzeń, które monitorują np. wagę, ciśnienie, poziom glukozy itp.;
- c) finansowe, stanowiące pełny zapis operacji gospodarczych odzwierciedlających prowadzoną działalność;
- d) z działalności naukowo-badawczej, wyniki prowadzonych badań, w tym badań nad lekami, projektowanie urządzeń medycznych, nowe metody leczenia itp.;

¹⁷ A. McAfee, E. Brynjolfsson, *Big Data, czyli przełom w zarządzaniu firmą*, „Harvard Business Review Polska” 2013, nr 122, kwiecień.

¹⁸ D. Bollier, *The promise and Peril of Big Data*, Raport of the Aspen Institute, Communications and Society Program, Washington 2010.

¹⁹ K. Marconi, M. Dobra, C. Thompson, *The use of Big Data in Healthcare*, w: *Big Data and Business Analytics*, red. J. Liebowitz, CRC Press, Boca Raton 2013, s. 229–248.

- e) dostarczane przez pacjentów, zawierające opis preferencji, poziom satysfakcji, informacje z systemów samodzielnego monitorowania swojej aktywności, np. ćwiczeń, snu, spożywanych posiłków itp.;
- f) z mediów społecznościowych.

Zarówno charakterystyka *Big Data*, jak i specyfika procesu analizy takich danych sprawiają, że przed organizacjami stają nowe wyzwania technologiczne i organizacyjne²⁰. Sektor ochrony zdrowia od zawsze generował ogromne liczby danych, co jest związane m.in. z koniecznością przechowywania dokumentacji medycznej pacjentów. Ponadto *Big Data* w ochronie zdrowia charakteryzują się jeszcze większą różnorodnością niż dane biznesowe. Stąd wniosek, że problem analiz *Big Data* w ochronie zdrowia sprowadza się nie tylko do przytłaczającej objętości, ale również niespotykanej różnorodności pod względem rodzajów, formatów danych oraz szybkości, z jaką powinny być analizowane w celu dostarczenia na bieżąco niezbędnych informacji. Zarządzanie nimi przy pomocy tradycyjnego oprogramowania, m.in. szpitalnych systemów informatycznych, oprogramowania *Business Intelligence* czy klasycznych systemów analitycznych, staje się coraz trudniejsze i bardziej złożone. Trudno również zastosować do nich tradycyjne narzędzia i metody służące do zarządzania danymi²¹. Dlatego też potencjał widzi się w analizach *Big Data*, zwłaszcza w aspekcie poprawy jakości opieki medycznej, ratowania życia czy obniżenia kosztów. Wyodrębnienie z tego gąszczy danych reguł asocjacyjnych, wzorców, trendów pozwoli dostawcom usług zdrowotnych, a także innym interesariuszom sektora ochrony zdrowia oferować dokładniejsze i bardziej wnikliwe diagnozy pacjentów, spersonalizowane leczenie, lepszą jakość usług medycznych i opieki nad pacjentem, a jednocześnie możliwość obniżenia kosztów.

Analizując zjawisko *Big Data* w sektorze ochrony zdrowia, należy zauważyć, że może być ono rozpatrywane z punktu widzenia trzech obszarów: epidemiologicznego, klinicznego oraz biznesowego.

Z klinicznego punktu widzenia analiza *Big Data* ma na celu poprawę zdrowia i kondycji pacjentów, umożliwienie długofalowych predykcji odnośnie do stanu ich zdrowia oraz wdrażanie odpowiednich procedur terapeutycznych. Docelowo zastosowanie analizy danych w medycynie ma pozwalać na dostosowanie terapii do konkretnego pacjenta, czyli spersonalizowanie medycyny (*precision, personalized medicine*). Z punktu widzenia epidemiologicznego jest pożądane

²⁰ B. Schmarzo, *Big Data. Understanding How Data Powers Big Business*, John Wiley & Sons Inc., Indianapolis 2013.

²¹ Ibidem.

uzyskanie dokładnej prognozy zachorowalności w celu wdrażania programów prewencyjnych z odpowiednim wyprzedzeniem. W kontekście biznesowym analiza *Big Data* może umożliwić oferowanie spersonalizowanych pakietów usług komercyjnych czy określenie prawdopodobieństwa wystąpienia jednostkowego chorób i zakażeń.

Warto podkreślić fakt, że *Big Data* to nie tylko gromadzenie i przetwarzanie danych, ale przede wszystkim wnioskowanie i wizualizacja danych, niezbędne do uzyskania określonych korzyści biznesowych²². W sektorze ochrony zdrowia dane stanowią jeden z kluczowych zasobów organizacji. Dostarczają je nie tylko pacjenci, ale również organizacje i instytucje, a także są przesyłane przez różnego rodzaju urzędnika monitorujące, czujniki czy instrumenty²³. Dane, które zostały dotychczas wygenerowane w sektorze ochrony zdrowia, są przechowywane w formie zarówno tradycyjnej, papierowej, jak i cyfrowej. Chcąc wprowadzić nowe sposoby zarządzania, rozwiązania w zakresie efektywności, transparentności, musimy uczynić te dane bardziej dostępnymi, cyfrowymi, możliwymi do przeszukania, przeanalizowania i wizualizowania.

4. Potencjalne obszary wykorzystania *Big Data* w ochronie zdrowia

Od wieków leczenie pacjentów było oparte na osądzie lekarzy, którzy podejmowali decyzje dotyczące leczenia, ale w ciągu ostatnich lat coraz bardziej na znaczeniu zyskuje medycyna oparta na dowodach (*evidence-based medicine*), która jest związana z systematyczną analizą danych klinicznych i podejmowaniem decyzji odnośnie do leczenia na podstawie najlepszych dostępnych informacji²⁴.

Organizacje z sektora ochrony zdrowia dostrzegają szansę rozwoju poprzez inwestycje w analitykę *Big Data*. W ostatnich latach, poprzez gromadzenie danych medycznych pacjentów i przekształcanie ich w *Big Data* oraz zastosowanie do nich odpowiednich algorytmów, wytworzono wiarygodne informacje, które pomagają pacjentom, lekarzom i innym interesariuszom sektora ochrony

²² http://www.dlp-expert.pl/articles/id,476/big_data_w_sluzbie_ochrony_zdrowia.html (data odczytu: 20.11.2015).

²³ H. Chen, R.H. Chiang, V.C. Storey, *Business Intelligence and Analytics: From Big Data to Big Impact*, „MIS Quarterly” 2012, vol. 36, no. 4, s. 1165–1188.

²⁴ J. Kyoungyoung, K. Gang Hoon, op.cit.

zdrowia identyfikować wartości oraz szanse²⁵. Warto zauważyć, iż w samej strukturze sektora ochrony zdrowia również zachodzi wiele zmian i pojawiają się nowe wyzwania. Digitalizacja, łączenie oraz efektywne wykorzystanie *Big Data* w ochronie zdrowia może przynieść korzyści praktycznie każdemu interesariuszowi tego sektora – zarówno pojedynczemu lekarzowi, jak i całemu systemowi opieki zdrowotnej. Potencjalne możliwości osiągnięcia korzyści i efektów z tego typu analiz w ochronie zdrowia prezentuje tabela 1.

Tabela 1. Potencjalne obszary wykorzystania analiz *Big Data* w ochronie zdrowia

Obszar zastosowania	Możliwości wykorzystania analiz <i>Big Data</i>
Poprawa jakości usług zdrowotnych	<ul style="list-style-type: none"> • ocena przez lekarza postawionej diagnozy i wskazanego przez siebie sposobu leczenia choroby przy wykorzystaniu systemu wspomaganie decyzji pracującego na zbiorach <i>Big Data</i>²⁶; • wykrywanie skuteczniejszych z medycznego punktu widzenia oraz bardziej opłacalnych sposobów diagnozowania i leczenia pacjentów; • analiza dużych wolumenów danych w celu dotarcia do praktycznych informacji przydatnych do identyfikacji potrzeb, wprowadzania nowych usług zdrowotnych, zapobiegania i przeciwdziałania kryzysom; • predykcja zachorowalności na dane schorzenia; • wykrywanie trendów, które prowadzą do polepszenia zdrowia i stylu życia społeczeństwa; • analiza ludzkiego genomu w celu wprowadzenia leczenia spersonalizowanego
Wspomaganie pracy personelu medycznego	<ul style="list-style-type: none"> • porównanie przez lekarza bieżącego przypadku medycznego do przypadków z przeszłości w celu lepszej diagnostyki oraz lepszego dopasowania leczenia; • wykrywanie chorób na wcześniejszych etapach, gdy mogą być one łatwiej i szybciej wyleczone; • wykrywanie zagrożeń epidemiologicznych i poprawa kontroli ognisk chorobotwórczych i szybkości reakcji; • identyfikacja pacjentów, u których przewiduje się największe ryzyko wystąpienia konkretnych, zagrażających życiu chorób, dzięki zestawieniu danych dotyczące historii najczęściej występujących chorób u leczących się osób ze zgłoszeniami wpływającymi do towarzystw ubezpieczeniowych; • zarządzanie zdrowiem każdego pacjenta z osobna (spersonalizowana medycyna) oraz zarządzanie zdrowiem całego społeczeństwa;

²⁵ N. Kamal, S. Wiebe, J. Engbers, M.D. Hill, *Big Data and Visual Analytics in Health and Medicine: From Pipe Dream to Reality*, „Health & Medical Informatics” 2014, vol. 5(5).

²⁶ <http://serwerb.com/big-data-w-sluzbie-zdrowia> (data odczytu: 20.11.2015).

Obszar zastosowania	Możliwości wykorzystania analiz <i>Big Data</i>
	<ul style="list-style-type: none"> • przechwytywanie i analizowanie w czasie rzeczywistym dużych liczb danych ze szpitalnych i domowych urządzeń monitorujących funkcje życiowe w celu monitorowania bezpieczeństwa oraz przewidywania zdarzeń niepożądanych; • analiza profili pacjentów w celu identyfikacji osób, w stosunku do których powinna zostać zastosowana profilaktyka, zmiana stylu życia, czyli opieka prewencyjna; • możliwość przewidywania wystąpienia konkretnych schorzeń lub pogorszenia wyników badań pacjenta; • prognozowanie progresji choroby i czynników ją determinujących oraz szacowanie ryzyka powikłań; • wykrywanie interakcji leków, ich działań ubocznych
Wspomaganie działalności naukowo-badawczej	<ul style="list-style-type: none"> • wspomaganie pracy nad nowymi lekami i badaniami klinicznymi dzięki możliwości analize „wszystkich danych” zamiast doboru próby do badań; • możliwość identyfikacji pacjentów o określonych, szczególnych cechach biologicznych, którzy wezmą udział w specjalistycznych badaniach klinicznych²⁷; • wyselekcjonowanie takiej grupy pacjentów, w przypadku której badany lek ze znacznym prawdopodobieństwem odniesie pożądany skutek i nie wystąpią efekty uboczne; • wykorzystanie modelowania i analiz predykcyjnych do projektowania lepszych leków oraz urządzeń
Obszar biznesowy (ekonomiczny)	<ul style="list-style-type: none"> • obniżenie kosztów i przeciwdziałanie nadużyciom i praktykom korupcyjnym²⁸; • szybsza i skuteczniejsza identyfikacja nieprawidłowych lub nieuprawnionych operacji finansowych w celu zapobiegania nadużyciom i eliminacji błędów; • wzrost rentowności poprzez wykrywanie pacjentów generujących wysokie koszty lub identyfikację lekarzy, których praca, procedury, sposoby leczenia kosztują najwięcej, w celu zaproponowania im rozwiązań mających wpływ na zmniejszenie wydawanych środków; • identyfikacja zbytecznych działań i procedur medycznych, np. duplikowanych testów

Źródło: opracowanie własne.

Powyższa specyfikacja nie stanowi pełnej listy potencjalnych obszarów wykorzystania analiz *Big Data* w ochronie zdrowia, gdyż możliwości zastosowania tychże analiz są praktycznie nieograniczone. Dodatkowo zaawansowane

²⁷ http://www.dlp-expert.pl/articles/id,476/big_data_w_sluzbie_ochrony_zdrowia.html (data odczytu: 20.11.2015).

²⁸ https://www.uthsc.edu/prevmed/docs/BigData_Healthcare_Shankar2014.pdf (data odczytu: 20.11.2015).

narzędzia analityczne pozwalają na analizę danych ze wszystkich możliwych źródeł i przeprowadzenie krzyżowych analiz w celu zapewnienia lepszego wglądu w dane²⁹. Przykładowo, analiza krzyżowa może dotyczyć zestawienia cech pacjenta i kosztów oraz wyników opieki, co może pomóc zidentyfikować najlepsze pod względem medycznym oraz najbardziej opłacalne leczenie czy zabiegi, co – być może – pozwoli na lepsze dopasowanie oferty usługodawcy³⁰. Z kolei analiza profili pacjentów (np. segmentacja i modelowanie predykcyjne) pozwala na identyfikację osób, które powinny zostać objęte profilaktyką, prewencją lub powinny zmienić styl życia³¹.

Zaawansowane techniki i technologie umożliwiają wyszukiwanie, gromadzenie, dystrybucję, zarządzanie oraz analizę danych w sektorze ochrony zdrowia³². Istniejące techniki analityczne mogą zostać wykorzystane do ogromnej liczby istniejących (ale dotychczas nieanalizowanych) danych dotyczących zdrowia pacjenta i powiązanych z nimi danych medycznych w celu lepszego rozumienia uzyskanych informacji i wyników oraz zaprojektowania optymalnych ścieżek klinicznych³³.

5. Schemat architektury systemu analitycznego *Big Data* dla organizacji z sektora ochrony zdrowia

W literaturze przedmiotu można spotkać wiele propozycji systemów analitycznych dla ochrony zdrowia. Jednakże wyzwaniem projektowym jest stworzenie systemu analitycznego, który da dostęp do czytelnych informacji szerokiej grupie użytkowników³⁴. System taki zapewniłby możliwość analizy eksploracyjnej danych historycznych, ale również mógłby być wykorzystywany w bieżącej

²⁹ P. Groves, B. Kayyali, D. Knott, S. Van Kuiken, *The 'Big Data' revolution in healthcare. Accelerating value and innovation*, http://www.pharmatalents.es/assets/files/Big_Data_Revolution.pdf (data odczytu: 20.11.2015).

³⁰ W. Raghupathi, V. Raghupathi, op.cit.

³¹ Ibidem.

³² J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Hung Byers, *Big Data: The next frontier for innovation, competition and productivity*, McKinsey Global Institute Raport, May 2011, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation (data odczytu: 20.09.2015).

³³ W. Raghupathi, V. Raghupathi, op.cit.

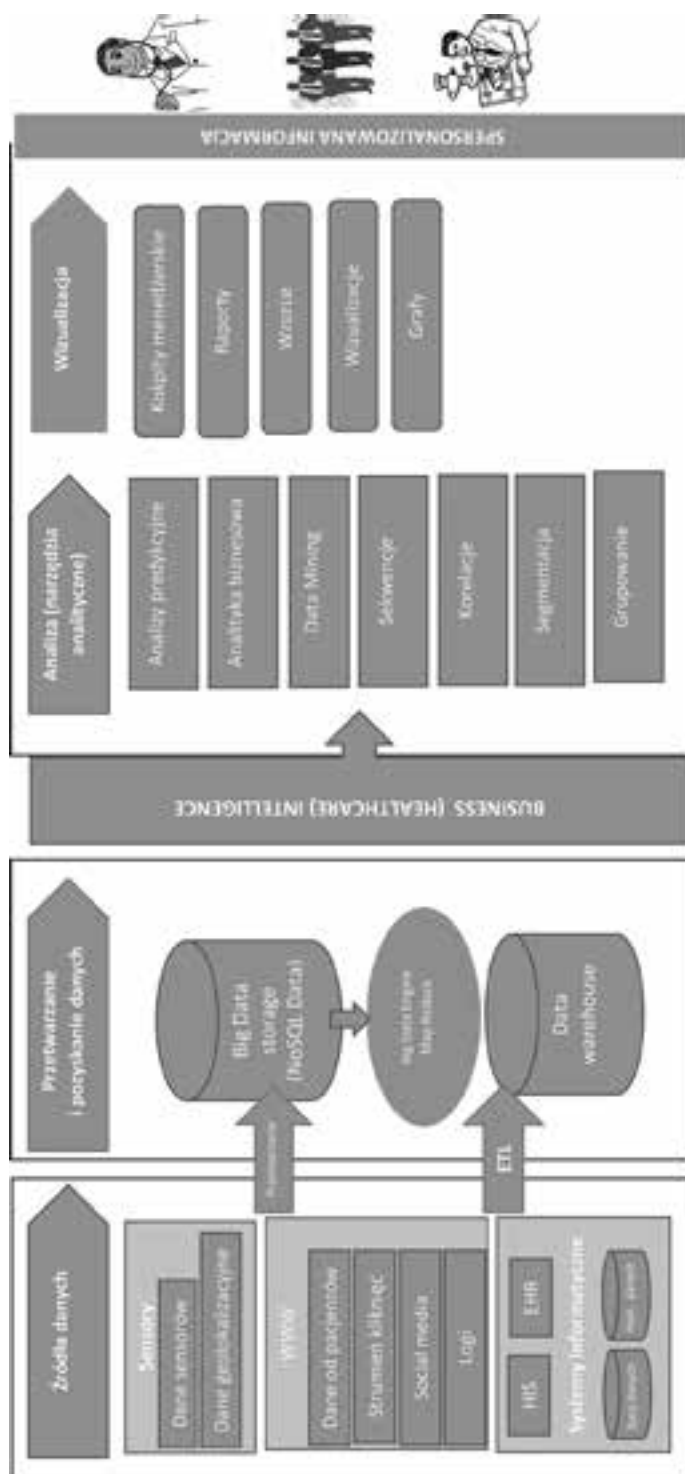
³⁴ M. Mazurek, *Architektura systemu wspomaganie decyzji medycznych wykorzystująca technologię przetwarzania danych Big Data*, http://rocznikikae.sgh.waw.pl/p/roczniki_kae_z35_16.pdf (data odczytu: 20.11.2015).

pracy lekarzy do wsparcia diagnostyki. Proponowanym rozwiązaniem tego problemu jest wykorzystanie możliwości, jakie stwarza technologia przetwarzania dużych, nieustrukturalizowanych danych, czyli *Big Data*, z zaletami systemów *Business Intelligence* oraz narzędzi do wizualizacji danych. Reasumując: w schemacie tym połączono możliwości, jakie dają rozwiązania *Business Intelligence* (BI), analizy biznesowe i eksploracja danych (rysunek 1).

Konceptualny schemat systemu analitycznego dla ochrony zdrowia nawiązuje do tradycyjnych systemów analitycznych. Kluczowa różnica to sposób przetwarzania danych. W tradycyjnych systemach analitycznych przeznaczonych dla sektora ochrony zdrowia przetwarzanie danych może być przeprowadzone przez narzędzie w postaci systemu *Business Intelligence* zainstalowanego w komputerze³⁵. W przypadku systemu obsługującego *Big Data* przetwarzanie musi być rozproszone, czyli podzielone i realizowane pomiędzy wieloma węzłami.

W skład *Big Data* w sektorze medycznym wchodzi m.in.: wszelkie dane związane z opieką nad pacjentem oraz jego zdrowym stylem życia wraz z danymi klinicznymi z systemów szpitalnych oraz systemów wspomagania decyzji klinicznych (odręczne i cyfrowe zapiski lekarzy, recepty, obrazy medyczne, dane z laboratoriów, aptek, ubezpieczalni oraz pozostałe dane administracyjne), dane pacjentów w elektronicznych rekordach pacjenta (EHR), dane generowane przez urządzenia monitorujące, sensory i czujniki, dane z mediów społecznościowych (posty z Tweetera, statusy z Facebooka, posty na blogach itp.), a także mniej specyficzne dla pacjenta dane w postaci danych w sytuacjach alarmowych, kanały wiadomości czy artykuły w czasopismach medycznych. W celu analizy *Big Data* dane te muszą być połączone, a następnie przetworzone lub przekształcone poprzez: 1) etapy ekstrakcji, transformacji i ładowania danych (ETL) i umieszczane w hurtowni danych lub 2) umieszczenie w repozytorium nierelacyjnym (*Big Data storage*), a następnie zaciągnięte do silnika Hadoop. Dlatego też najważniejszym założeniem architektury jest współwystępowanie repozytorium relacyjnego oraz repozytorium nierelacyjnego (bazy NoSQL). Następnie narzędzia analityczne umożliwiają m.in.: przeprowadzanie analiz predykcyjnych, analizę skupień, asocjacji, sekwencji, klasyfikację, segmentację, grupowanie, rekomendacje.

³⁵ K. Bartuś, *Koncepcja wykorzystania systemów Competitive Intelligence w działalności organizacji*, w: *Informatyka przyszłości 2*, red. W. Chmielarz, J. Kisielnicki, T. Parys, Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego, Warszawa 2015, s. 153–163.



Rysunek 1. Model integracji systemów opieki zdrowotnej z BI oraz *Big Data*

Źródło: opracowanie własne.

Prezentowany schemat ukazuje system analityczny, który może sprostać wyzwaniu w postaci *Big Data* poprzez włączenie do architektury platformy IBM Hadoop, która pozwala rozwiązać takie problemy, jak przechowywanie, przetwarzanie i dostęp do nieustrukturalizowanych danych (*Big Data*).

Chcąc skutecznie wykorzystać potencjał analiz *Big Data*, czyniąc je jednocześnie przystępnymi dla użytkowników z sektora ochrony zdrowia, należy wykorzystać metody wizualizacji i tworzenia spersonalizowanych widoków danych dla różnych użytkowników systemu (lekarzy, biznesu, naukowców oraz pacjentów). Uzyskane poprzez analizy wyniki mogłyby być również przetwarzane przez systemy *Business Intelligence* i prezentowane w narzędziach wizualizacyjnych, np. kokpitach menedżerskich³⁶. Rozwiązanie takie pozwala na wykonanie skomplikowanych analiz, a jednocześnie prezentuje uzyskane wyniki w formie pozwalającej na natychmiastowe wykorzystanie, co jest bardzo istotne szczególnie w obszarze analiz klinicznych.

6. Podsumowanie i kierunki dalszych badań

W dzisiejszych czasach nowe technologie umożliwiają organizacjom komercyjnym czerpanie korzyści z analizy ogromnej liczby danych. Podobnie mogą funkcjonować organizacje z sektora ochrony zdrowia. Jednakże, zarówno na świecie, jak i w Polsce organizacje te dopiero zaczynają się uczyć, jak wykorzystać tę szansę. Ich sukces będzie zależał również od technicznych zdolności do efektywnego integrowania i analizowania informacji za pomocą nowych technologii, opracowania systemów wspomagania decyzji, a także wsparcia skutecznego podejmowania decyzji przez analityków³⁷. Po pokonaniu wielu wyzwań będzie możliwa identyfikacja ukrytych wzorców w bardzo rozbudowanych zbiorach danych i przełożenie ich na użyteczne informacje, na podstawie których będzie można podejmować bardziej trafne decyzje. Zanim jednak to nastąpi, organizacje te czeka niezwykle dużo pracy zarówno nad integracją wszystkich istotnych wewnętrznych oraz zewnętrznych źródeł danych, jak i nad odpowiednim interpretowaniem wyników analiz.

³⁶ C. Olszak, K. Batko, *Business Intelligence systems – new chances and possibilities for healthcare organizations*, „Informatyka Ekonomiczna” 2012, nr 3(25), s. 123–138.

³⁷ A. McAfee, E. Brynjolfsson, op.cit.

W kolejnej pracy związanej z wykorzystaniem *Big Data* w ochronie zdrowia jest planowany przegląd technologii *data mining* do analizy *Big Data* oraz przykładów wykorzystania *Big Data* w obszarze ochrony zdrowia.

Bibliografia

- Bartuś K., *Koncepcja wykorzystania systemów Competitive Intelligence w działalności organizacji*, w: *Informatyka przyszłości 2*, red. W. Chmielarz, J. Kisielnicki, T. Parys, Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego, Warszawa 2015, s. 153–163.
- Bollier D., *The promise and Peril of Big Data*, Raport of the Aspen Institute, Communications and Society Program, Washington 2010.
- Chen H., Chiang R.H., Storey V.C., *Business Intelligence and Analytics: From Big Data to Big Impact*, „MIS Quarterly” 2012, vol. 36, no. 4, s. 1165–1188.
- Kamal N., Wiebe S., Engbers J., Hill M.D., *Big Data and Visual Analytics in Health and Medicine: From Pipe Dream to Reality*, „Health & Medical Informatics” 2014, vol. 5(5).
- Kyoungyoung J., Gang Hoon K., *Potentiality of Big Data in the Medical Sector: Focus on How to Reshape the Healthcare System*, „Healthcare Informatics Research” 2013, vol. 19(2), s. 79–85.
- Mach-Król M., *A survey and assessment of maturity models for Big Data adoption*, w: *Proc. 11th International Conference on Strategic Management and its Support by Information Systems*, Uherske Hradiste, Czech Republic, 21–22 May 2015.
- Mach-Król M., *Analiza i strategia Big Data w organizacjach*, „Studia i Materiały” Polskiego Stowarzyszenia Zarządzania Wiedzą 2015, t. 74, s. 43–55.
- Marconi K., Dobra M., Thompson C., *The use of Big Data in Healthcare*, w: *Big Data and Business Analytics*, red. J. Liebowitz, CRC Press, Boca Raton 2013, s. 229–248.
- Mazurek M., *Architektura systemu wspomagania decyzji medycznych wykorzystująca technologię przetwarzania danych Big Data*, http://rocznikikae.sgh.waw.pl/p/roczniki_kae_z35_16.pdf (data odczytu: 20.11.2015).
- McAfee A., Brynjolfsson E., *Big Data, czyli przełom w zarządzaniu firmą*, „Harvard Business Review Polska” 2013, nr 122, kwiecień.
- Olszak C., Batko K., *Business Intelligence systems – new chances and possibilities for healthcare organizations*, „Informatyka Ekonomiczna” 2012, nr 3(25), s. 123–138.
- Raghupathi W., Raghupathi V., *Big Data analytics in healthcare: promise and potential*, „Health Information Science and Systems” 2013, vol. 2(3).
- Schmarzo B., *Big Data. Understanding How Data Powers Big Business*, John Wiley & Sons Inc., Indianapolis 2013.

Williams N., Ferdinand N.P., Croft R., *Project management maturity in the age of Big Data*, „International Journal of Managing Projects in Business” 2014, vol. 7(2), s. 311–317.

Źródła sieciowe

Groves P., Kayyali B., Knott D., Van Kuiken S., *The ‘Big Data’ revolution in healthcare. Accelerating value and innovation*, http://www.pharmatalents.es/assets/files/Big_Data_Revolution.pdf (data odczytu: 20.11.2015).

<http://manager.money.pl/ludzie/twarza-w-twarz/artykul/sluzba-zdrowia-przed-rewolucja-dane-medyczne,180,0,1717172.html> (data odczytu: 20.11.2015).

<http://serwerblog.com/big-data-w-sluzbie-zdrowia/> (data odczytu: 20.11.2015).

<http://www.chip.pl/news/wydarzenia/statystyka/2013/02/big-bang-theory-czyli-jak-eksplozja-danych-zmienia-swiat-1> (data odczytu: 20.11.2015).

<http://www.computerworld.pl/news/392036/Big.Data.w.sluzbie.zdrowia.html> (data odczytu: 20.11.2015).

<http://www.computerworld.pl/news/396729/ERP.w.kolorze.bialym.html> (data odczytu: 20.11.2015).

http://www.dlp-expert.pl/articles/id,476/big_data_w_sluzbie_ochrony_zdrowia.html (data odczytu: 20.11.2015).

http://www.hbrp.pl/nie_przeocz.php?id=1172&t=raport-praktyczne-zastosowanie-metod-analitycznych-big-data (data odczytu: 20.11.2015).

http://www.web.gov.pl/wiedza/587_4420.html (data odczytu: 20.11.2015).

https://www.uthsc.edu/prevmmed/docs/BigData_Healthcare_Shankar2014.pdf (data odczytu: 20.11.2015).

Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Hung Byers A., *Big Data: The next frontier for innovation, competition and productivity*, McKinsey Global Institute Raport, May 2011, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation (data odczytu: 20.09.2015).

Olszak C., Batko K., *The Use of Business Intelligence Systems in Healthcare Organizations in Poland*, Federated Conference on Computer Science and Information Systems, 2012, <http://ieeexplore.ieee.org/> (data odczytu: 20.09.2015), s. 997–1004.

Słoniewski T., *Od BI do „Big Data”*, <http://it-manager.pl/wp-content/uploads/Nowa-twarz-BI1.pdf> (data odczytu: 20.11.2015).

* * *

The possibilities of using big data in healthcare

Summary

This article provides an overview of big data analytics in healthcare. Firstly, the various advantages and characteristics of big data analytics used in healthcare are defined and discussed. Then, the challenges are identified. Thirdly, the areas for the potential use of big data analytics are identified and characterized. Fourthly, the architectural framework of big data analytics in healthcare is described. Lastly, the conclusions are presented.

Keywords: big data, big data analytics, healthcare