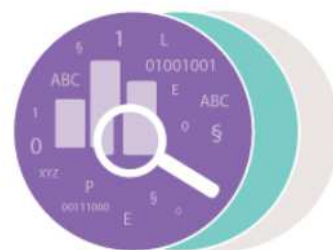




FutureTDM

Explore . Analyse . Improve



REDUCING BARRIERS AND INCREASING UPTAKE OF TEXT AND DATA MINING FOR RESEARCH ENVIRONMENTS USING A COLLABORATIVE KNOWLEDGE AND OPEN INFORMATION APPROACH

EKONOMIKA EKSPLOKACJI TEKSTU I DANYCH – ANALIZA AKTUALNYCH TRENDÓW I PRZYSZŁYCH ZASTOSOWAŃ

Jan Strycharz

Wstęp

W raporcie tym przedstawiamy ekonomiczną analizę rozwoju eksploracji tekstu i danych (*TDM* – z ang. *Text and Data Mining*). W toku narracji:

1. prezentujemy aktualne i przyszłe trendy w wykorzystywaniu *TDM* jako aktywa gospodarczego
2. opisujemy dynamikę rynkową (zarówno globalną, jak i europejską) związaną z technologiami *TDM*
3. szacujemy wartość szerszego wpływu technologii *TDM* na globalną i europejską gospodarkę

Analiza skupia się na fenomenie *TDM* rozumianym jako aktywo gospodarcze oraz praktyka biznesowa. W myśl tego podejścia podejmujemy próbę przedstawienia, jak podmioty gospodarcze (firmy) rozumieją i wykorzystują technologie *TDM* jako narzędzie tworzenia wartości rynkowej. Trzeba przy tym już na wstępie zaznaczyć, iż w zasadzie nie istnieją badania czy analizy, które mierzą się w podobny (bezpośredni) sposób z fenomenem *TDM*. Wyjaśniamy, że w ujęciu gospodarczym na *TDM* należy spoglądać jako na element szerszego fenomenu opisywanego w literaturze jako *Big Data*.

Stąd niezwłocznie trzeba zwrócić uwagę, że samego terminu *Big Data* używa się aktualnie w dwóch głównych znaczeniach – (1) jako nazwy wielkich zbiorów danych, których ze względu na ich charakterystykę nie można przeanalizować tylko i wyłącznie siłami ludzkimi, oraz (2) jako terminu ogólnego obejmującego całość procesów biznesowych, które wspierają wykorzystanie tych zbiorów danych i ich przetwarzanie na realną wartość komercyjną. W tym pierwszym znaczeniu charakterystyka wielkich zbiorów danych jest określana przez tzw. 3 V (Buchholtz 2014: 10) oznaczające z ang. rozmiar (*Volume*), różnorodność źródeł i charakteru danych (*Variety*) oraz prędkość, z jaką dane są tworzone i aktualizowane (*Velocity*). Rozwój technologii przechowywania danych z kategorii *Big Data* doprowadził nawet do wprowadzenia pojęcia „jeziora danych” zamiast „zbioru danych”, co służyć ma zaakcentowaniu nowego podejścia do zarządzania informacją w przedsiębiorstwie (patrz ramka poniżej).

Ramka 1. Czym jest jezioro danych?

„Jezioro danych to repozytorium danych, w którym znajduje się olbrzymia ilość tzw. surowych danych w formacie organicznym. Podczas gdy hierarchiczny układ danych (podejście klasyczne) używa raczej plików i folderów, jezioro danych korzysta z architektury płaskiej (podejście *Big Data*). Każdy element danych umieszczonych w jeziorze ma przypisany unikatowy identyfikator oraz zbiór metadanych. Kiedy pojawia się pytanie biznesowe, jezioro danych może być przeszukiwane pod kątem adekwatnych kategorii danych, których to zbiór można następnie analizować w celu znalezienia odpowiedzi.

Termin «jezioro danych» jest zwykle kojarzony z takimi technologiami przechowywania danych jak *Hadoop*. Podczas korzystania z tej technologii dane najpierw ładowane są na platformę *Hadoop*, a następnie wykorzystuje się narzędzia eksploracji danych i analizy biznesowej przy użyciu konkretnych grup danych z platformy”.

Źródło: Rouse M., *Data Lake*, TechTarget.com, dostęp przez stronę [www.http://searchaws.techtarget.com/definition/data-lake](http://searchaws.techtarget.com/definition/data-lake) w dn. 23 sierpnia 2016 r.

Podejście *Big Data* zakłada zatem utrzymywanie płaskiej (w przeciwieństwie do hierarchicznej) struktury przechowywania danych, gdzie każda dana ma przypisany zbiór tagów. Podejście to pozwala eksplorować wszystkie informacje adekwatne do problemu, który aktualnie jest przedmiotem zainteresowania decydentów. Jednak samo posiadanie odpowiedniej infrastruktury danych to dopiero punkt wyjścia do wykorzystania jej w celach biznesowych. Dlatego termin *Big Data* w kontekście biznesowym jest używany raczej w tym szerszym znaczeniu – jako termin ogólny, pod którym kryją się wszystkie procesy niezbędne do tego, by rozwój i utrzymanie infrastruktury *Big Data* mogły doprowadzić do wytworzenia realnej wartości komercyjnej.

Analizując aktualne dyskusje na ten temat, wyciągamy wniosek, iż *TDM* jest zbiorem technologii i technik tworzących szerszą opowieść o tym, jak firmy kapitalizują i monetyzują *Big Data*. Ten sposób rozumienia jest fundamentalny dla analiz i wniosków prezentowanych w tym raporcie. Stąd większość badań, z których korzystamy w tym opracowaniu, odnosi się do fenomenu *Big Data*. Wykorzystujemy jednak ten materiał, aby doprecyzować rozumienie *TDM* w kontekście działalności gospodarczej.

Raport składa się z trzech rozdziałów zasadniczych. Rozpoczyna się prezentacją tła teoretycznego – w sposób szczegółowy wyjaśniamy tu podstawowe aspekty rozumienia *TDM* oraz jego roli w rozwoju fenomenu *Big Data*. W tej części bronimy tezy, iż obietnica pozytywnego wpływu *Big Data* na gospodarkę opiera się głównie na dwóch elementach – (1) właściwym tworzeniu, rozwoju i utrzymaniu infrastruktury danych oraz (2) właściwej eksploracji, analizie i interpretacji danych (co rozumiemy jako *TDM*). Te dwa elementy są rzecz jasna połączone i od siebie zależne. W związku z tym w pierwszym rozdziale podkreślamy kluczową rolę *TDM* w procesie wykorzystania danych do tworzenia wartości gospodarczej i wyjaśniamy tę relację w sposób pogłębiony.

W drugim rozdziale prezentujemy zestawienia faktów i danych na temat wykorzystania *Big Data* i *TDM* przez firmy. Korzystamy przy tym z różnorodnych źródeł i badań na temat praktyk, wyzwań i sukcesów związanych z *Big Data* i *TDM* jako praktykami biznesowymi. Dzięki analizom przedstawionym w tej sekcji czytelnik zyskuje dostęp do wiedzy na temat aktualnych trendów oraz celów stosowania zautomatyzowanej analizy danych w przedsiębiorstwach, a także na temat branż najchętniej inwestujących w tego typu projekty analityczne.

Trzeci rozdział poświęciliśmy szacunkom dotyczącym wielkości rynku *Big Data* oraz rynku *TDM* – zarówno na poziomie globalnym, jak i europejskim. Co jednak ciekawsze, prezentujemy w nim nie tylko estymacje wartości rynkowej, lecz także oceniamy wartość szerszego wpływu *TDM* na gospodarkę, wykorzystując szacunki dotyczące fenomenu *Big Data* wykonane w oparciu o modele ekonomiczne przez firmę IDC oraz Warszawski Instytut Studiów Ekonomicznych.

Wierzymy, iż prezentowany raport może przyczynić się do rekomendacji wspierających osiągnięcie celów wyznaczonych w strategii rozwoju społeczno-gospodarczego Europa 2020, w szczególności w odniesieniu do celów związanych z ideą inteligentnego rozwoju rozumianego jako „stworzenie gospodarki opartej na wiedzy i innowacjach” (Europe 2020: 10). *TDM*, jak tłumaczymy w rozdziale pierwszym, można postrzegać jako część procesów badawczo-rozwojowych (B+R). Stąd wspieranie wykorzystania gospodarczego technologii *TDM* będzie jednocześnie służyć wzrostowi prywatnych wydatków na B+R, co zbliży Europę do osiągnięcia sumy wydatków na B+R w wysokości 3% PKB. Co jednak ważniejsze, będzie to oznaczać, iż produkty i usługi wykreowane przez firmy europejskie będą osadzone w metodycznym podejściu gromadzenia i analizowania wszelkich dostępnych danych. Fakt ten powinien przełożyć się na konkurencyjność europejskiej gospodarki w ujęciu globalnym.

Ramka 2. Główne wnioski

Technologie Eksploracji Tekstu i Danych (*TDM* – z ang. *Text and Data Mining*) odgrywają kluczową rolę w zagwarantowaniu wpływu fenomenu *Big Data* na gospodarkę. Tworzenie wartości ekonomicznej na bazie nieskończonych zasobów informacyjnych stało się bowiem aktywnością konieczną w świecie charakteryzującym się ciągłym wzrostem ilości danych. Firmy prywatne, instytucje publiczne czy organizacje pozarządowe są coraz bardziej zainteresowane zrozumieniem, jak podejmować decyzje w oparciu o dane. Ponieważ przepływy informacyjne są coraz szybsze, a liczba danych – coraz większa, umiejętność „wydobycia” czy eksploracji odpowiednich danych umożliwiającą podejmowanie decyzji staje się kluczowa w odnoszeniu sukcesów rynkowych, zapewnieniu skuteczności polityk publicznych czy – w końcu – wzrostu gospodarczego i rozwoju społecznego.

Technologie *TDM* rozumiane jako zautomatyzowany sposób analizy są zatem kluczowe w tworzeniu realnej wartości w oparciu o przepastne zbiory informacji. Jednak aby tworzenie wartości przy użyciu *TDM* było w ogóle możliwe, *TDM* musi być nastawione na wywoływanie czegoś, co określa się w języku angielskim mianem *actionable intelligence*, czyli zbioru informacji, które są łatwe w interpretacji i jednocześnie umożliwiają podjęcie decyzji, za którą idzie konkretne działanie. Aby tak się stało, *TDM* musi spełniać wiele warunków – sukces użycia tych technologii zależy w dużej mierze od ludzkich umiejętności i talentów. Praktycy *TDM* – tzw. *data miners*, *data scientists*, czy *data magicians* – powinni mieć nie tylko pogłębione zrozumienie technologii oraz metod eksploracji i analizy danych. Co być może ważniejsze, powinni także dobrze rozumieć kontekst praktyczny (biznesowy, społeczny czy polityczny), w którym pracują, aby dobrze wybierać metody i zakresy eksploracji, analizy i interpretacji.

W przypadku działań biznesowych oznacza to konieczność pogłębiania wiedzy nt. danej firmy i jej procesów – im bardziej wiedza ta jest pogłębiona, tym większe szanse na odkrywanie znaczących trendów i tworzenie znaczących predykcji. Innymi słowy, proces translacji danych na wartość gospodarczą musi być mocno wspierany właśnie przez biznesową stronę organizacji – czyli osoby, które posiadają najwięcej wiedzy na temat sposobu, w jaki dana organizacja tworzy, dostarcza i sprzedaje dobra. W sytuacji odwrotnej – gdy kluczowi decydenci nie angażują się w proces dynamicznej wymiany informacyjnej z analitykami i eksploratorami danych – możemy mieć do czynienia z jedną z większych barier stojących na przeszkodzie do ustanowienia łańcucha tworzenia wartości analitycznej.

Skuteczność firm w skali mikro może przełożyć się na efekty w skali makro. Według analizy wykonanej przez Warszawski Instytut Studiów Ekonomicznych (Buchholtz *et al.* 2014) fenomen *Big and Open Data* może przyczynić się do wzrostu europejskiego PKB o dodatkowe 1,9% w roku 2020. Ten wzrost spowodują głównie trzy rodzaje efektów występujących właśnie w mikroskali: (1) lepsze wykorzystanie zasobów, (2) lepsze podejmowanie decyzji oraz (3) udoskonalanie produktów i procesów. Trzeba jednak zauważyć, iż w zasadzie wszystkie te efekty da się sprowadzić do wyzwania związanego z lepszym podejmowaniem decyzji w oparciu o dane – czyli wykorzystaniem technologii *Big Data* i *TDM*.

Jak pokazują wnioski wyciągane przez autorów różnych raportów zajmujących się tą tematyką, firmy są skuteczne w implementowaniu rozwiązań *Big Data* i *TDM* jedynie wówczas, gdy mają dostęp do odpowiednich zasobów ludzkich, posiadają zasoby finansowe na inwestycje w te technologie i mają odpowiednio przygotowane przywództwo, które angażuje się w projekty analityczne, mimo iż istnieją znaczne trudności w liczeniu stopy zwrotu z takich inwestycji.

Eksploracja i analiza wielkich zbiorów przestała już być niszą i w ostatnich latach weszła do rdzenia praktyk biznesowych. W tym czasie firmy nauczyły się tworzyć i utrzymywać infrastrukturę danych, jednak wciąż zmagają się z wyzwaniami związanymi z wykorzystaniem tej infrastruktury w procesach tworzenia, dostarczania i sprzedawania dóbr. Badania pokazują, że nie ma tu łatwych odpowiedzi, jednak jednym z kluczowych elementów tego analitycznego równania są ludzkie talenty analityczne. Bez nich wszystkie dane pozostają jedynie martwym zasobem, a infrastruktura danych przestaje być aktywem i staje się kosztem.

Firmy są zgodne, iż brak ludzkich talentów jest jednym z najpoważniejszych wyzwań stojących na drodze translacji analizy na działania i wartość biznesową. Nawet najwięksi gracze przyznają, że niedobór wysoko wykwalifikowanych analityków nie pozwala w pełni wykorzystać potencjału zautomatyzowanej analityki danych. Zasadną wydaje się zatem predykcja, iż w najbliższym czasie firmy będą inwestować w pozyskiwanie i rozwój własnych talentów. Dodatkowym ważnym elementem w tym procesie jest konieczność wypracowania „informacyjnie zmyślnej” (z ang. *data savvy*) kultury organizacyjnej. Tylko taka kultura umożliwi szybki „transport” i implementację efektów pracy eksploratorów danych wewnątrz całej organizacji.

Coraz więcej firm eksperymentuje z projektami analitycznymi wykorzystującymi *Big Data* i technologie *TDM* – trend ten widoczny jest szczególnie w sektorach handlu detalicznego, w komunikacji oraz w sektorze sprzedaży prądu i gazu. Niektóre badania wskazują, iż 48% firm używających *Big Data* korzysta nie tylko z danych liczbowych, lecz także analizuje dane tekstowe.

Wartość finansowa technologii *TDM* rośnie relatywnie szybko. Należy przy tym pamiętać, iż firmy ciągle najwięcej inwestują w stworzenie i utrzymywanie infrastruktury *Big Data*, a dopiero w następnym kroku myślą o mocach związanych z eksploracją i analizą. Duża część firm znajduje się już jednak na tym drugim etapie rozwoju zautomatyzowanej analityki i odczuwa silną motywację do inwestowania w pozyskiwanie odpowiednich technologii analitycznych i – co najważniejsze – ludzi o odpowiednich umiejętnościach. Estymacje pokazują, iż w roku 2016 globalny rynek *Big Data* był wart 23,8 miliarda dolarów, z czego 6,4 miliarda dotyczyło technologii *TDM* (oprogramowanie, obsługa i szkolenia). Z kolei na rynku europejskim wartości te wyniosły odpowiednio 9,4 miliarda dolarów dla *Big Data* i 2,5 miliarda dla *TDM*. Europejski rynek *TDM* może urosnąć do 10,3 miliarda dolarów w roku 2021. Dla porównania można przytoczyć, iż wpływy ze sprzedaży smartphonów w roku 2015 wyniosły ok. 14 miliardów dolarów dla Europy Środkowej i Wschodniej.

Tworzenie wartości rynkowej w oparciu o dane będzie miało również przełożenie na efekty w makroskali. Efekty te będą oczywiście zależne od charakterystyki danego kraju czy regionu – w szczególności od zdolności absorpcji technologii IT. Łącząc analizy takich firm doradczych jak IDC i WISE, można powiedzieć, że 1 euro zainwestowane w *Big Data* przełoży się na efekty warte dla gospodarki 10,7 euro. Oznacza to, że makroekonomiczny wpływ *Big Data* jest dwukrotnie silniejszy, niż wpływ analityki opartej na danych tradycyjnych, gdzie 1 euro tłumaczone było na 5 euro wartości. Ponieważ *TDM* dotyczy głównie danych typu *Big Data*, można wyciągnąć wniosek, iż wpływ ekonomiczny tych technologii wyniósł w roku 2016 blisko 26,7 miliarda dolarów (scenariusz optymalny). Wpływ ten może wzrosnąć do 110,1 miliarda dolarów w roku 2020.

Spis treści

Wstęp	2
Główne wnioski	4
1. Eksploracja tekstu i danych jako aktywo gospodarcze.....	7
1.1 Od danych do wartości – podstawowe definicje <i>Big Data</i> i <i>TDM</i>	7
1.2 Pogłębiona charakterystyka <i>Text</i> i <i>Data Miningu</i> i konotacja biznesowa	12
2. Perspektywa i doświadczenie firm	22
2.1 Praktyki i bariery.....	22
2.2 Zróżnicowanie branżowe i źródła danych	31
2.3 Konkluzja – ludzkie talenty analityczne niezbędnym składnikiem konwersji <i>Big Data</i> na wartość gospodarczą.....	36
3. Rynkowa i gospodarcza wartość <i>Big Data</i> i <i>TDM</i>	38
3.1. Szacunki globalne i europejskie.....	38
3.2 Szerszy wpływ danych na gospodarkę	45
Bibliografia.....	47

1. Eksploracja tekstu i danych jako aktywo gospodarcze

1.1 Od danych do wartości – podstawowe definicje *Big Data* i *TDM*

Eksploracja tekstu i danych (*TDM*) może być rozumiana w ujęciu ekonomicznym jako aktywność nakierowana na zautomatyzowane pozyskanie wartości z przepastnych zasobów informacyjnych. W tym sensie aktywność tę należy postrzegać jako proces, który może stanowić konkretny zasób wspierający rozwój i konkurencyjność gospodarczą zarówno w skali mikro, jak i makro. Aby jednak *TDM* stał się takim zasobem, jego wyniki jako procesu muszą prowadzić do praktycznych wniosków, dzięki którym działanie podejmowane po jego zastosowaniu jest jakościowo lepsze, niż gdyby nie zostało ono zasilone wynikami *TDM*.

Ramka 3. Porównywarki cenowe – praktyczna ilustracja *TDM*

Porównywarki cenowe takie jak ceneo.pl czy skapiec.pl są dobrym przykładem praktycznego zastosowania technik *TDM*. Konsumenci po uruchomieniu tej usługi webowej wpisują nazwę konkretnego towaru i otrzymują listę cen proponowanych za ten produkt przez wielu różnych sprzedawców. Silnik takiej porównywarki jest oparty o algorytm, który w sposób zautomatyzowany eksploruje sieć www pod kątem wydobycia informacji (tekstowych, numerycznych i graficznych) o cenie i sprzedawcy. Pozwala to konsumentowi na tańszy zakup. Ta decyzja zakupowa (doinformowana procesem *TDM*) oznacza, że środki pieniężne są alokowane bardziej efektywnie. Efekt społeczny jest taki, że poprawia się finansowy dobrobyt danego gospodarstwa domowego. Na poziomie makroekonomicznym można mówić, iż zaoszczędzone środki napędzają wzrost gospodarczy – poprzez dodatkową konsumpcję czy inwestycje (pośrednie lub bezpośrednie).

Wydaje się, że technologie *TDM* stają się w zasadzie konieczne w świecie, który charakteryzuje zalew informacyjny. Możliwości technologiczne sprawiają, iż duża część ludzkiej aktywności jest rejestrowana cyfrowo, co tworzy konkretny zapis, który można analizować w dowolnym czasie. Jednak ilość i prędkość powstawania tych danych jest na tyle duża, że człowiek nie ma możliwości analizowania ich w sposób tradycyjny – stąd wymogiem stają się zautomatyzowane algorytmy. Warto jednocześnie podkreślić, iż gdy dostęp do danych przestaje być problemem, wyzwaniem staje się krystalizowanie wartości z chaotycznego świata informacyjnego; zwracano na to uwagę już w połowie lat 90. XX wieku:

„Wiedza i informacja są dzisiaj obecne w wystarczającej obfitości – to, czego brakuje, to możliwości ich użycia w sposób znaczący”

(OECD 1996: 11)

TDM nabiera zatem kluczowego znaczenia jako warunek konieczny wyławiania znaczących treści informacyjnych z przepastnych zbiorów (czy jezior) danych. Znaczenie *TDM* jest tym bardziej szczególne, gdy mamy do czynienia właśnie ze zbiorami płaskimi, nieustrukturyzowanymi

hierarchicznie, czyli ze zbiorami *Big Data*¹. W tym sensie *TDM* i *Big Data* to dwa elementy tej samej układanki, co wyjaśniamy dokładniej w ramce poniżej i w dalszych akapitach tego rozdziału.

Ramka 4. *Big Data* a *Data Mining*

„*Big Data* i *Data Mining* to dwa różne fenomeny. Obydwa związane są z gromadzeniem wielkich zbiorów danych i raportowaniem wyników analizy, która ma służyć firmom lub innym odbiorcom. Jednakże każdy z tych fenomenów odnosi się do odmiennych elementów tego procesu.

Big Data to termin określający wielkie zbiory danych. Są to zbiory, które swoim rozmiarem przerastają tradycyjne technologie gromadzenia danych. Przykładowo, zbiory danych, które są zbyt wielkie, aby nimi zarządzać w takim programie jak Excel, można zakwalifikować jako zbiory *Big Data*.

Data Mining odnosi się do aktywności związanej z przeszukiwaniem zbiorów *Big Data* pod kątem znalezienia adekwatnych informacji. Ten rodzaj działania jest dobrym zobrazowaniem starego przysłowia „szukać igły w stogu siana”. Chodzi o to, że firmy gromadzą obszerną ilość danych w sposób automatyczny. Decydenci potrzebują jednak dostępu do precyzyjnych informacji, które są jedynie małą częścią tych obszernych zbiorów. Używają wówczas technik *Data Miningu* (czyli eksploracji danych), aby wydobyć dane, które pomogą w podejmowaniu decyzji wpływających na kierunek prowadzenia biznesu.

Data Mining może być wykonywany przy użyciu różnorodnych programów czy narzędzi analitycznych. Proces ten może zostać całkowicie zautomatyzowany lub wymagać intensywnej pracy ludzkiej, kiedy to pojedynczy pracownicy wysyłają serię konkretnych zapytań o dane znajdujące się w archiwum czy bazie danych. Ogólnie rzecz biorąc, *Data Mining* odnosi się do złożonych i skomplikowanych operacji, które doprowadzają do bardzo precyzyjnych wyników. Przykładowo (w swojej najprostszej formie) narzędzie do *Data Miningu* może przeszukiwać dziesiątki lat informacji księgowych, aby znaleźć konkretną kolumnę wydatków odnoszącą się do konkretnej kategorii wpływów w danym roku.

W skrócie, *Big Data* jest aktywem, a *Data Mining* jest przewodnikiem i doradcą używanym do dostarczenia wartościowych wyników”.

Źródło: Techopedia, What is the difference between big data and data mining?, dostęp przez stronę www.techopedia.com/7/29678/technology-trends/what-is-the-difference-between-big-data-and-data-mining w dn. 18 lipca 2016 r.

Wyjaśniając relacje między *Big Data* i operacjami spod znaku *Data Mining*, warto również nadmienić, iż wyrażenie *Big Data* ma także kolokwialne znaczenie, które – do pewnego stopnia – wskazuje na sposób, w jaki przedsiębiorstwa podchodzą do wykorzystania tego fenomenu na swoje potrzeby.

¹ Warto przy tym pamiętać, iż w kontekście biznesowym (czy korporacyjnym) *Big Data* jest również używane (jak opisujemy to we wprowadzeniu do tego raportu) jako termin ogólny zawierający w sobie wszystkie elementy i procesy biznesowe mające na celu translację danych typu *Big Data* w wartość biznesową czy komercyjną. W tym znaczeniu termin ten obejmuje takie aktywności jak *TDM*, rozwój strategii, inwestowanie czy kulturę organizacyjną. Tutaj używamy jednak sformułowania *Big Data* w jego węższym aspekcie – jako typ danych o konkretnej charakterystyce.

Termin ten bowiem używany jest często przez praktyków biznesowych na określenie danych o klientach zbieranych w toku wielu interakcji firmy z odbiorcą końcowym. Równie często odnosi się on do danych zbieranych w złożonych procesach produkcyjnych, których analiza może doprowadzić do optymalizacji produkcji.

Biorąc pod uwagę zmieniające się praktyki wokół fenomenu *Big Data*, własną definicję proponuje również jedna z kluczowych firm doradczych McKinsey (2013: 1):

„Termin «Big Data» odnosi się do zbiorów danych, których rozmiar przekracza możliwości typowych programów do obsługi baz danych [...]. Przy czym nie definiujemy wielkości, jaką dany zbiór musi osiągnąć w sensie konkretnej liczby terabajtów. Zakładamy, że w miarę rozwoju technologii rozmiar zbiorów, które będą się kwalifikować jako Big Data, również będzie się zwiększał. Warto także zauważyć, że definicja może się różnić w zależności od sektora, rodzaju szeroko dostępnego oprogramowania i rozmiaru baz danych partykularnego dla danej branży”.

Warto również przypomnieć (co zaznaczyliśmy już we wstępie), iż zbiory danych typu „*Big*” definiowane są również przez tzw. 3 V (Buchholtz 2014: 10) oznaczające z ang. rozmiar (*Volume*), różnorodność źródeł (*Variety*) oraz prędkość, z jaką dane są tworzone i aktualizowane (*Velocity*).

Tłumacząc znaczenie fenomenu *Big Data* i technologii *TDM*, warto jednak nie skupiać się na definicjach, które z racji wczesnego rozwoju tematyki pozostają wciąż zmienne. Warto pamiętać o sensie, który jest bezpośrednio związany z faktem, iż współczesne gospodarki rozwinięte funkcjonują według modelu „gospodarki opartej na wiedzy”. Oznacza to, że „produkcja i usługi są oparte na działaniach nasyconych wiedzą” (Powell i Snellman 2014: 201). Przekłada się to na „większą zależność [sukcesu] od możliwości intelektualnych, aniżeli od wkładu fizycznego czy posiadania zasobów naturalnych” (*Ibidem*). Aby zatem osiągnąć sukces w gospodarce opartej na wiedzy, trzeba podejmować decyzje zasilane adekwatną analizą wszelkich dostępnych informacji. A więc umiejętność wytypowania adekwatnych informacji przeznaczonych pod dalszą analizę (ważny aspekt praktyki *TDM*) również staje się zdolnością o kluczowym znaczeniu. Tłumacząc te relacje, warto odwołać się do anglojęzycznego terminu *actionable intelligence*, który trudno przetłumaczyć na język polski. Chodzi tu o zdobywanie informacji, które umożliwiają podjęcie ważnych decyzji w sposób definitywny (patrz ramka poniżej).

Ramka 5. Wyjaśnienie terminu „*actionable intelligence*”

Uchwycenie różnicy między informacją „po prostu” a *actionable intelligence* jest kluczowe dla zrozumienia roli, jaką odgrywa TDM w szerszej historii związanej z fenomenem *Big Data*. „Informacja” jest bowiem terminem generycznym, który oznacza w zasadzie – jak definiuje słownik języka polskiego – każdą „wiadomość o czymś”.

Z drugiej strony *actionable intelligence* odnosi się bezpośrednio do świata praktyki, działania i decyzji związanych z konkretnymi celami, rezultatami i osiągnięciami. Wyróżnienie wyrażenia „*actionable intelligence*” wskazuje jednocześnie na to, że nie każda informacja ma praktyczne zastosowanie niezależnie od kontekstu. Aby zakwalifikować informację jako *actionable*, musi ona być dostarczona w odpowiedniej formie i w odpowiednim czasie do konkretnej osoby, której praca polega na realizowaniu konkretnego procesu nastawionego na osiągnięcie jakiegoś celu. Farmer i Carter (2014) definiują ten termin podając następujący przykład:

„Charakter i znaczenie *actionable intelligence* dobrze oddaje poniższa wypowiedź dyrektora jednego z większych przedsiębiorstw działających w branży dóbr konsumenckich:

Dokładnie monitorowaliśmy wystartowanie z produktem, którego wartość oszacowano na 55,6 miliona dolarów w ciągu następujących 12 miesięcy. Narzędzie analityczne pokazało nam, że przewidywania na temat globalnego popytu na ten produkt nagle wzrosły o 160 tysięcy jednostek na następne pół roku [...]. Ta konkretna informacja pozwoliła nam proaktywnie zareagować na wzrost popytu, dzięki czemu zmniejszyliśmy koszty produkcji, ale przede wszystkim uniknęliśmy sytuacji, w której brakuje nam towaru”.

(Farmer i Carter 2014: xv)

Wyzwanie translacji zasobów informacyjnych na *actionable intelligence* związane jest głównie z koniecznością dopasowania jej do kontekstu działania oraz z faktem, że nie ma ona żadnej konkretnej charakterystyki. Innymi słowy, nigdy nie wiadomo, jakiego rodzaju obserwacja pozwoli konkretnemu decydentowi na dostrzeżenie możliwości, szansy czy zagrożenia, które uruchomi serię wymaganych działań.

Aby zatem uczynić informację użyteczną, każdorazowo musi ona zostać przetransformowana w *actionable intelligence*. W przeciwnym wypadku jej moc wpływania na procesy organizacyjne pozostaje czysto potencjalna.

Obraz komplikuje dodatkowo fakt, iż – jak się okazuje – termin „«wydobycie» (od ang. *mining*) nie oddaje dobrze czynności, które osoby zatrudnione w tej branży rzeczywiście wykonują jako część swoich obowiązków. Wydobycie implikuje bowiem, iż chodzi o znalezienie pojedynczych danych, które z jakiegoś powodu okażą się cenne. Oznaczałoby to, że praktycy angażują się w odkrywanie nowych faktów w archiwach. Jednak w praktyce sprawy mają się inaczej. Aplikacje do *Data Miningu* okazują się bowiem narzędziami do odkrywania trendów i wzorców ukazujących się w analizie bardzo dużych zbiorów informacji [...]” (Fayyad i Uthurusamy, 1999). Chodzi tu zatem bardziej o syntezę niż znalezienie konkretnych faktów.

Oznacza to, iż wypracowywanie wartości z *Big Data* przy użyciu technologii *TDM* polega raczej na wyszukiwaniu adekwatnych ciągów informacyjnych, które można przeanalizować i niejako zsyntetyzować do konkretnych wniosków prowadzących do podjęcia decyzji. Być może to dlatego można się często spotkać z używaniem terminu „analityka” (z ang. *analytics*) wymiennie z terminem „wydobycie” (z ang. *mining*) w odniesieniu do praktyk *TDM*. Wielu praktyków twierdzi wręcz, że nie można nakreślić jasnej granicy między *wydobyciem* a *analizą* w przypadku pracy ze zbiorami *Big Data*.

Powyższe wyjaśnienia bezpośrednio nawiązują do sformułowanej już tezy, iż informacja nie jest „łatwo przekształcalna w obiekt standardowych transakcji ekonomicznych” (OECD). Innymi słowy, aby przetransformować dane na aktywne ekonomiczne, musi zajść wiele różnorodnych procesów. Jak już powiedziano, działania i technologie związane z terminem „*Text i Data Mining*” są kluczowe, kiedy mamy do czynienia ze źródłami informacji zbyt obszernymi, aby można było nimi zarządzać bez użycia zaawansowanych technologii cyfrowych.

„[Współcześnie] skala [...] danych przerasta jakiegokolwiek możliwości człowieka, jeśli chodzi o nadawanie sensu faktom. Przy takiej skali trendy są zwykle zbyt subtelne, a powiązania między danymi – zbyt skomplikowane, aby można je było zaobserwować, po prostu patrząc na zbiory. Data Mining jest narzędziem automatyzacji tego procesu służącym odkrywaniu interpretowalnych wzorców czy trendów; pomaga widzieć las, nie gubiąc się między drzewami”.

(Furnas 2012)

Znaczenie technologii i działań *TDM* jest jeszcze bardziej widoczne, kiedy weźmiemy pod uwagę, w jaki sposób praktyki związane z *Big Data* (jak i samo rozumienie fenomenu) zmieniają się z biegiem czasu. Nawiązując do definicji cytowanej w poprzedniej części tekstu (opartej o trzy „V”), wydaje się, „że ze wzrostem użycia *Big Data* nie mniej ważną rolę zaczyna odgrywać dodatkowe «V» – chodzi o *Veracity* (czyli wiarygodność). W przeciwieństwie do trzech oryginalnych V [*Volume* – prędkość, *Variety* – różnorodność, *Velocity* – prędkość] «wiarygodność» nie odnosi się do swoistej charakterystyki danych, a raczej do ich jakości, która jest uzależniona od użyteczności w zastosowaniach praktycznych” (*Ibidem*). Stąd kryterium wiarygodności zwraca uwagę na (oczywisty zdawałoby się) fakt, iż możliwość wykorzystania *Big Data* jako zasobu ekonomicznego zależy nie tylko od skali i różnorodności, lecz także od umiejętności wyselekcjonowania danych, którym można ufać. W przeciwnym razie zwiększa się ryzyko, że odkrywane trendy czy wzorce, na których bazuje się późniejsze analizy, predykcje i decyzje, nie oddają rzeczywistości. Fakt ten wzmacnia rolę praktyk *TDM* – to bowiem od *data minerów* zależy określenie odpowiednich zakresów eksploracji danych. W tym sensie znaczenie *TDM* zaczyna do pewnego stopnia przewyższać istotność umiejętności budowy infrastruktury danych *Big Data*.

„W roku 2016 firmy będą chciały uciec od irrelevantnego hałasu kreowanego przez niektóre dane. Firmy uznają, iż różnorodność i prędkość mogą zniechęcać i podejmą bardziej przemyślane działania nastawione na analizę danych rzeczywiście użytecznych, aby osiągać znaczące wglądy w sposób

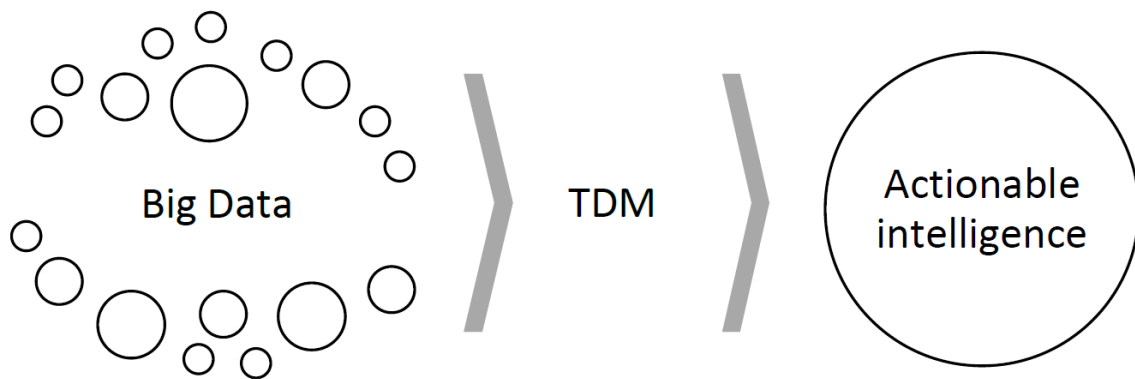
szybki i holistyczny. Zamiast inwestować czas i pieniądze w tworzenie infrastruktury IT zarządzającej dużą ilością danych, wyzwanie będzie polegało raczej na zarządzaniu różnorodnością i prędkością sptywania informacji, co umożliwi osiągnięcie wartościowych wglądów [...]”.

(Gutierrez 2015)

Podsumowując, biorąc pod uwagę charakterystykę danych *Big Data*, *TDM* umożliwia wydobywanie i translację informacji na *actionable intelligence*, którą następnie można wykorzystać do usprawniania wszelkich procesów biznesowych – ale także procesów związanych z projektowaniem i rozwojem polityk publicznych czy procesów organizacji trzeciego sektora.

Diagram 1.

TDM jako interfejs między Big Data a działaniem operacyjnym



Źródło: opracowanie własne

1.2 Pogłębiona charakterystyka *Text i Data Miningu* i konotacja biznesowa

Praca eksploratorów danych jest w swojej charakterystyce zbliżona do każdej innej pracy o silnym komponencie kreatywnym, stąd trudno jest przedstawić jej ustandaryzowany opis. Można jednak zidentyfikować dwa kluczowe cele, do których należy dążyć w pracy z gęszcem informacyjnym – opis i predykcję (Fayyad *et al.* 1996).

„Wyciąganie wniosków z danych przybiera dwie zasadnicze formy: opisu i predykcji. W tak dużej skali bardzo trudno jest wskazać, co rzeczywiście pokazują dane. Techniki i działania Data Miningu wykorzystuje się po to, by uprościć dane w taki sposób, aby umożliwić ich zrozumienie oraz wyciąganie wniosków na temat partykularnych przypadków w oparciu o obserwowane wzorce”. (Furnas op. cit.)

Zwykle oznacza to stosowanie różnych metod statystycznych – poniżej prezentujemy główne sposoby wykrywania wzorców, łącząc je ze światem praktyki biznesowej.

Tabela 1. Metody TDM i ich biznesowe odniesienie (Furnas 2012)

<i>Sposób wykrywania wzorców</i>	<i>Opis</i>
Wykrywanie anomalii	<p>W przypadku dużych zestawów danych możliwe jest odkrycie pewnej tendencji, typowego sposobu, w jaki układają się poszczególne dane. Metody statystyczne można wówczas wykorzystać do określenia elementów odróżniających się od wzorca. Na przykład, administracja podatkowa tworzy modele typowych deklaracji podatkowych, a wykorzystując wykrywanie anomalii, może zidentyfikować konkretne zyski różniące się od tego przeglądu. Łatwiej wówczas wytypować przypadki wymagające kontroli pod względem legalności.</p>
Uczenie się powiązań	<p>Jest to rodzaj eksploracji danych napędzający system rekomendacji Amazon. Może okazać się, że klienci, którzy zdecydowali się na zakup shakera i książki z przepisami na koktajle, będą również zainteresowani kupnem kieliszków do martini. Tego rodzaju wyników często używa się do targetowania kuponów, ofert lub reklam. Taki sposób wykorzystywania danych (choć w bardziej skomplikowanej formie) stoi także za rekomendacjami obejrzenia konkretnych filmów na platformie Netflix na podstawie wcześniejszej historii oglądania.</p>
Wykrywanie klastrów	<p>Jeden ze sposobów wykrywania wzorców szczególnie przydatny przy rozpoznawaniu odrębnych grup lub podkategorii danych. Bez <i>Data Miningu</i> analityk musiałby spojrzeć na dane i samodzielnie zdecydować, który zestaw kategorii oddaje istotne różnice między grupami w widocznych danych. W tym przypadku pojawia się jednak ryzyko utraty ważnych kategorii. Dzięki technikom <i>Data Miningu</i> możliwe jest, aby dane samoistnie określiły swoją grupę przynależności. Jest to jeden z rodzajów trudnej do zrozumienia „czarnej skrzynki” algorytmów. Używając jednak prostego przykładu, da się wskazać, o co tutaj chodzi – możemy wyobrazić sobie, że zwyczaje zakupowe różnych grup hobbistów będą wyglądać zupełnie inaczej: różnice pomiędzy ogrodnikami, rybakami i entuzjastami modeli samolotów byłyby dość wyraźne. Algorytmy maszynowego myślenia są w stanie wykryć wszystkie podgrupy w zbiorze danych znacząco różniące się od pozostałych.</p>

Klasyfikacja	Jeżeli istniejąca struktura jest już znana, eksplorację danych można zastosować do klasyfikowania nowych przypadków do wcześniej ustalonych kategorii. Ucząc się z dużego zbioru wstępnie zaklasyfikowanych przykładów, algorytmy mogą wykryć utrzymujące się różnice systemowe między pozycjami w każdej z grup i zastosować je do nowych problemów klasyfikacyjnych. Filtry spamowe są tego doskonałym przykładem – duże zbiory wiadomości e-mail, które zidentyfikowano jako spam, pozwoliły filtrom wykrywać różnice w wykorzystywanym słownictwie pomiędzy istotnymi a spamowymi wiadomościami i z dużym stopniem dokładności klasyfikować wszystkie przychodzące wiadomości według tych reguł.
Regresja	Eksplorację danych można wykorzystać do konstruowania modeli predykcyjnych bazujących na wielu zmiennych. Facebook, na przykład, może być zainteresowany przewidywaniem przyszłego zaangażowania użytkownika na podstawie jego dotychczasowych zachowań. Czynniki takie jak ilość udostępnionych informacji osobistych, liczba oznaczonych zdjęć, zaproszonych i przyjętych znajomych, komentarzy czy polubień itd. mogą zawrzeć się w takim modelu. Z biegiem czasu na podstawie tego, jak Facebook porównuje prognozy z obserwowanym zachowaniem, może on być udoskonalany, tak aby dobierać i ważyć informacje w inny sposób. Obserwacje te mogą być wykorzystywane do projektowania usługi w taki sposób, aby zwiększyć liczbę zachowań prowadzących do zwiększonego zaangażowania.
<p>Wzory i struktury wykryte dzięki opisowej eksploracji danych często wykorzystywane są do przewidywania innych aspektów danych. Przykład firmy Amazon pokazuje, jak odpowiednio wykorzystać ustalenia opisowe do prognozowania. Ten (hipotetyczny) związek pomiędzy shakerem a zamówieniem kieliszków do Martini można wykorzystać do wykrycia wielu innych podobnych związków w ramach modelu prognozowania prawdopodobieństwa, że dany użytkownik dokona określonego zakupu. Można go użyć do wszystkich skojarzeń z historii zakupów użytkownika i przewidzieć, jakie produkty najprawdopodobniej będzie chciał jeszcze kupić. Amazon może wówczas przygotować dla niego odpowiednią reklamę w oparciu o te przewidywania.</p>	

Powyższa tabela prezentuje techniki statystyczne używane w procesie *TDM*, dzięki którym możliwe jest nadawanie sensu przepastnym zbiorom informacji. Jednak samo zastosowanie tych technik nie prowadzi w sposób automatyczny do zamiany danych na *actionable intelligence* – a ten warunek musi zostać spełniony, aby danych można było użyć w kontekście biznesowym. Kwestia ta została poruszona już w roku 1996 przez Fayyada (: 39), który stwierdził, iż „ślepe stosowanie metod *Data Miningu* (słusznie krytykowane w literaturze dotyczącej stosowania metod statystycznych w ogóle) może być działaniem niebezpiecznym i łatwo prowadzić do odkrywania wzorców, które są wręcz błędne i nieprawdziwe”. Autorzy tych słów ukuli nawet termin „odkrywanie wiedzy w bazach danych” (*Knowledge Discovery in Databases – KDD*), który tłumaczą w następujący sposób:

„Naszym zdaniem «odkrywanie wiedzy» odnosi się do ogólnego procesu odkrywania użytecznej wiedzy z danych, podczas gdy *Data Mining* odnosi się

do pojedynczego kroku w tym procesie. Data Mining jest zastosowaniem konkretnych algorytmów, w celu wyławiania wzorców z danych [...]. Dodatkowe kroki, które należą do procesu odkrywania wiedzy, obejmują takie aspekty, jak przygotowanie danych, selekcja danych, czyszczenie danych, włączanie innej adekwatnej wiedzy oraz interpretację wyników Miningu. Te wszystkie elementy stanowią kluczowe kroki gwarantujące, że na bazie analizowanych danych wypracowuje się użyteczną wiedzę”. (Ibidem)

W kontekście biznesowym (czy organizacyjnym w ogóle) stawiamy jako autorzy tego raportu znak równości między terminami „wiedza użyteczna” i „*actionable intelligence*”. Innymi słowy, tylko użyteczna wiedza może być wykorzystana do realizacji konkretnych działań. Zatem podobnie jak konkretne warunki muszą zostać spełnione, aby wytwarzać wiedzę użyteczną, tak te same wymogi stosują się do procesu TDM rozumianego jako narzędzie wypracowywania informacji napędzających działania (*actionable intelligence*). W pierwszej kolejności eksploratorzy danych muszą rozumieć nie tylko dane, z którymi obcuja, lecz także kontekst, w którym działają. W kontekście działań biznesowych oznacza to zwykle konieczność pogłębionej znajomości danego przedsiębiorstwa i jego operacji – im bardziej wiedza ta jest pogłębiona, tym większe szanse na odkrywanie użytecznych wzorców i kreślenie użytecznych predykcji. Brak wzajemnego zrozumienia między eksploratorami danych a kluczowymi decydentami (liderami) danego przedsiębiorstwa jest jedną z większych barier dla rozwijania projektów z kategorii *Big Data* (co opisujemy bardziej szczegółowo w następnym rozdziale).

Zrozumienie kontekstu jest kluczowe i dopiero wówczas eksploratorzy danych mogą zacząć podejmować decyzje dotyczące zakresu danych do analizy – a następnie podejść do czyszczenia tych danych i przygotowania ich do analizy zgodnie z zasadą „wiarygodności”. Dopiero w następnym kroku eksploratorzy decydują się na konkretną metodę statystyczną, zaczynają analizę eksploracyjną, poszukują wzorców i zaczynają je interpretować. Ostateczny wynik ich działań powinien być na tyle operacyjny, żeby pozwalał na podjęcie decyzji prowadzącej do rozwiązania konkretnego problemu biznesowego. Cały proces prezentujemy w poniższej tabeli, jednak należy nadmienić, iż współcześnie rzadko jest on linearny – o wiele częściej ma charakter iteracyjny i wymaga płynnego przemieszczania się po poszczególnych etapach, także w kierunku wstecznym (*Ibidem*: 42).

Tabela 2. Proces translacji danych na działanie z wykorzystaniem technik TDM (Fayyad et al. 1996)

Krok	Wytłumaczenie procesu
Zrozumienie	Po pierwsze należy zrozumieć domenę aplikacji i istniejącą już wiedzę na jej temat oraz zidentyfikować cele procesu wytwarzania wiedzy użytecznej z praktycznego punktu widzenia.
Wybór	Po drugie, powinno się stworzyć docelowy zbiór danych: wybrać zestaw danych lub skoncentrować się na podzbiorze zmiennych czy próbek danych, w których wykorzystana ma być eksploracja.

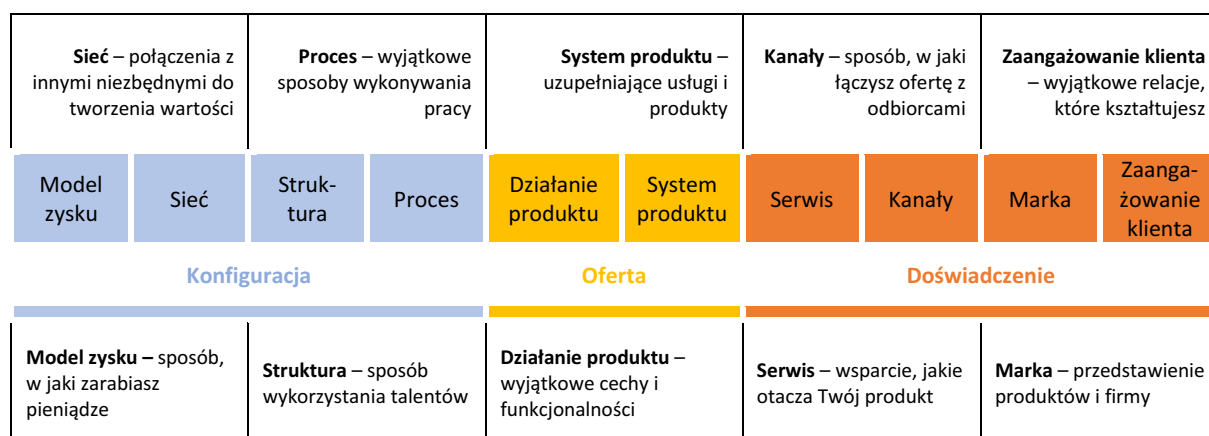
Obróbka	Krok trzeci to czyszczenie i obróbka danych. Podstawowe operacje powinny obejmować usunięcie szumów informacyjnych, jeśli się pojawiły, zebranie najistotniejszych informacji do modelowania, podjęcie decyzji na temat strategii dotyczącej obsługi brakujących pól danych oraz zarządzania sekwencjami czasu i znanymi zmianami.
Transformacja	Po czwarte, potrzebna jest redukcja danych i projekcja: znalezienie funkcji matematycznych przydatnych do reprezentowania danych w zależności od celu wykonywania zadania. Dzięki redukcji wymiarowości lub metodzie transformacji liczbę zmiennych w ramach danego rozpatrzenia można zmniejszyć lub znaleźć zamienniki dla nieistniejących danych.
Wybór metody	Krok piąty to dopasowanie celów procesu wytwarzania wiedzy użytecznej do konkretnej metody eksplorowania danych. Na przykład – podsumowania, klasyfikacja, regresji, klastrów itp. (patrz Tabela 1).
Eksploracja	Etap szósty stanowi analiza rozpoznawcza, model i wybór hipotezy: wybór algorytmu(-ów) eksploracji i sposobu(-ów) selekcjonowania może być stosowany do poszukiwania wzorców danych. Proces ten obejmuje decydowanie o tym, jakie modele i parametry będą odpowiednie (na przykład, modele skategoryzowanych danych różnią się od modeli wektorów ciągu liczb rzeczywistych), oraz dopasowanie konkretnego sposobu eksploracji danych do ogólnych kryteriów procesu <i>KDD</i> (na przykład użytkownik końcowy może być bardziej zainteresowany zrozumieniem modelu niż jego możliwościami predykcyjnymi).
Odkrywanie schematów	Po siódme: poszukiwanie wzorców w konkretnym formularzu reprezentacyjnym lub zestawień takich reprezentacji, w tym klasyfikacji reguł, regresji i klastrów. Użytkownik może znacząco wspomóc eksplorację, prawidłowo wykonując powyższe kroki.
Interpretacja	Po ósme należy zinterpretować wykryte wzory, prawdopodobnie wracając do każdego z etapów 1–7 w celu wykonania dalszych iteracji. Ten krok może również obejmować wizualizację wyciągniętych wzorów i danych uzyskanych dzięki wyodrębnionym modelom.
Działanie	Ostatni, dziewiąty etap to działania na zyskanej wiedzy: jej bezpośrednie wykorzystanie, włączenie do innego systemu do dalszych działań lub po prostu udokumentowanie i raportowanie zainteresowanym stronom. Proces ten obejmuje także sprawdzanie i rozwiązywanie potencjalnych konfliktów z dotychczas posiadaną wiedzą.

Wszystkie kroki przedstawione powyżej należy odczytywać jako warunki konieczne do spełnienia, aby zapewnić skuteczność technik *TDM*. Odkrywanie wzorców i czynienie predykcji musi być osadzone w pogłębionym zrozumieniu złożoności, w której operuje eksplorator danych. Oznacza to, że wiedzę

biznesową specjalistów od analizy danych należy stale aktualizować w związku ze zmianami wprowadzanymi ze strategicznego poziomu przedsiębiorstwa.

To, na co konkretnie mogą wpływać wyniki *TDM*, prowadząc do wzrostu przewagi konkurencyjnej danej firmy, dobrze opisuje model innowacyjności firmy opracowany przez Keely'ego (*et al.* 2013). Dzieli on elementy biznesu stale wymagające innowacyjności na trzy zasadnicze grupy: (1) konfigurację biznesu, (2) ofertę oraz (3) relacje z klientami. Poniższy diagram prezentuje 10 elementów podzielonych według tych trzech grup.

Diagram 2. 10 elementów biznesu wymagających innowacji (Keely 2013)



Uważamy, iż powyższy model jest niezwykle pomocny, jeśli chodzi o wytłumaczenie, jak konkretnie *Big Data* i techniki *TDM* mogą pozytywnie wpływać na innowacyjność przedsiębiorstw. Wiele bowiem opracowań wskazuje, że pozytywne efekty *Big Data* będą głównie odczuwalne na poziomie podejmowania decyzji. Innymi słowy podejmowane decyzje będą prowadzić do umocnienia się danego przedsiębiorstwa o wiele szybciej, jeżeli będą zapadały z wykorzystaniem *Big Data*. Jednak podejmowanie decyzji jest aktywnością w zasadzie wszechobecną w każdej organizacji. Model Keely'ego wskazuje, na usprawnienie których konkretnie elementów prowadzenia biznesu mogą one mieć przełożenie. Poniżej umieściliśmy tabelę, która w sposób szczegółowy opisuje każdy z dziesięciu elementów, którego rozwój można zasilić procesami *TDM*.

Tabela 3. 10 elementów biznesu wymagających innowacji (Keely 2013)

Kategoria innowacji	Element innowacji	Opis
Konfiguracja	Model zysku	Innowacyjne modele zysku pomagają znaleźć nową drogę do konwersji oferty firmy w gotówkę. Najlepsze z nich odzwierciedlają głębokie zrozumienie tego, co dla klientów i użytkowników najważniejsze i tego, gdzie znajdują się nowe źródła przychodów. Innowacyjne modele zysku często stanowią wyzwanie dla przemysłu zmęczonego starymi założeniami dotyczącymi ofert i zdobywania przychodu. Jest to duża część ich siły: w większości branż dominuje model zysku niekwestionowany często przez dziesięciolecia.

	Sieć	<p>W dzisiejszym przyłączonym do sieci świecie żadna firma nie może – lub nie powinna – robić wszystkiego sama. Innowacje sieciowe zapewniają firmom drogę do korzystania z procesów, technologii, kanałów i marek firm – a zatem prawie wszystkich elementów ich biznesów. Innowacje te oznaczają, że firma może bazować na własnych siłach przy jednoczesnym wykorzystaniu możliwości i aktywów innych. Innowacje sieciowe pomagają również kierownictwu w podziale ryzyka podczas opracowywania nowych ofert i przedsięwzięć. Współpraca ta może być krótka lub trwała i może być zawierana tak między bliskimi sojusznikami, jak i nawet zagorzałymi konkurentami.</p>
	Struktura	<p>Innowacje dotyczące struktury koncentrują się na organizowaniu aktywów firmy – twardych, ludzkich, materialnych – w unikalny, tworzący wartość sposób. Mogą one obejmować wszystko: od zarządzania talentami po pomysłowe systemy konfiguracji sprzętu kapitałochłonnego. Koszty stałe przedsiębiorstwa oraz funkcje korporacyjne również mogą poprawić się dzięki innowacjom struktury, w tym działów takich jak HR, R&D oraz IT. Ostatecznym celem jest to, by innowacje przyciągały do organizacji talenty, tworząc niezwykle produktywne środowisko pracy i wspierały poziom wydajności, z którym konkurencja nie może się równać.</p>
	Proces	<p>Innowacje procesowe odnoszą się do działań, w wyniku których powstaje pierwotna oferta przedsiębiorstwa. Innowacyjność wymaga tutaj diametralnej zmiany tzw. „<i>business as usual</i>”, która pozwoli firmie wykorzystać unikalne możliwości, działać efektywnie, szybko dostosowywać się do zmieniających się okoliczności oraz uzyskiwać duże marże. Innowacje procesowe często tworzą rdzeń kompetencji przedsiębiorstwa i mogą obejmować opatentowane i własnościowe podejścia, które dają przewagę na lata lub dziesięciolecia. Najlepiej byłoby, gdyby zawierały one „magiczny składnik”, którego konkurenci nie będą w stanie replikować.</p>

Oferta	Wydajność produktu	Innowacyjna wydajność produktu odnosi się do wartości, cech i jakości oferty firmy. Ten rodzaj innowacji często oznacza zarówno zupełnie nowe produkty, jak również aktualizacje i rozszerzenie linii, co dodaje wartość. Zbyt często błędnie podchodzi się do kwestii wydajności jako do wyznacznika jakości wszystkich pozostałych innowacji. Z pewnością wydajność jest ważna, ale należy pamiętać, że jest to tylko jeden z dziesięciu rodzajów innowacji, często najprostszy do skopiowania przez konkurencję. Warto pomyśleć o dowolnej wojnie produktowej, której było się świadkiem – przyczepność i twardość opon, łatwiejsze do używania i trzymania szczoteczki do zębów czy nawet jakość wózków dziecięcych. Zbyt szybko wszystko to przeobraża się w swoistą gorączkę. Innowacje wydajności produktu, które zapewniają długoterminową konkurencyjność, są raczej wyjątkiem niż regułą.
	System produktu	Innowacje systemu produktu zakorzenione są w sposobie, w jaki poszczególne produkty i usługi łączone są w celu stworzenia solidnych i skalowalnych systemów. Działania te wspierane są przez współdziałanie, modułowość, integrację i inne sposoby tworzenia wartościowego połączenia między odrębnymi i zróżnicowanymi ofertami. Innowacje systemowe wyrobów pomagają budować ekosystemy urzekające i zachwycające klientów oraz bronić się przed konkurencją.
Doświadczenie	Serwis	Innowacje w odniesieniu do serwisu zapewniają i zwiększają użyteczność, wydajność i odczuwalną wartość oferty. Tworzą produkt łatwiejszy w użytkowaniu i cieszeniu się nim. Ujawniają te jego cechy i funkcje, które klienci mogą przeoczyć. Ponadto rozwiązują problemy, które klient napotyka podczas użytkowania produktu. Jeśli są dobrze sporządzone, mogą przekształcić nawet średniej jakości produkty w doświadczenie tak przekonujące, że gwarantuje ono powrót danego klienta.
	Kanały	Innowacje kanałów obejmują wszystkie sposoby nastawione na połączenie oferty firmy z klientami i użytkownikami. Choć <i>e-commerce</i> stał się dominującą w ostatnich latach siłą, tradycyjne kanały, takie jak fizyczne sklepy, nadal są ważne – szczególnie w zakresie tworzenia wciągających doświadczeń. Wykwalifikowani innowatorzy kreują zwykle wiele komplementarnych sposobów, aby przekazać swoje produkty i usługi klientom. Ich celem jest zapewnienie użytkownikom możliwości zakupu tego, czego chcą, kiedy chcą i jak chcą przy minimalnych problemach i kosztach, a przy maksimum zadowolenia.

	Marka	Innowacyjność marki przyczyni się do zapewnienia klientom i użytkownikom oferty, którą rozpoznają i zapamiętują, a także wybierają ją chętniej niż ofertę konkurentów czy substytuty danej firmy. Największe marki mają odrębną tożsamość, która przyciąga kupujących. Taka tożsamość jest zazwyczaj wynikiem starannie przygotowanych strategii, które realizowane są w wielu punktach styku między firmą a klientami, w tym w publikacjach, reklamach, usługach, interakcjach oraz środowiskach pracowników, biznesu i partnerów. Innowacje marki mogą przekształcać towary w cenne produkty, oferty i przedsiębiorstwa, nadając im sens, intencję i wartość.
	Zaangażowanie klienta	Innowacje kreujące zaangażowanie klientów dotyczą zrozumienia głęboko zakorzenionych aspiracji klientów i użytkowników oraz wykorzystania tych informacji do opracowania sensownych połączeń między nimi a firmą. Dobre innowacje tego typu powinny ukazać szerokie ścieżki dla poszukiwania i pomocy w znalezieniu sposobu, jak pewną część ich życia uczynić bardziej godną zapamiętania, wspaniałą, a nawet magiczną.

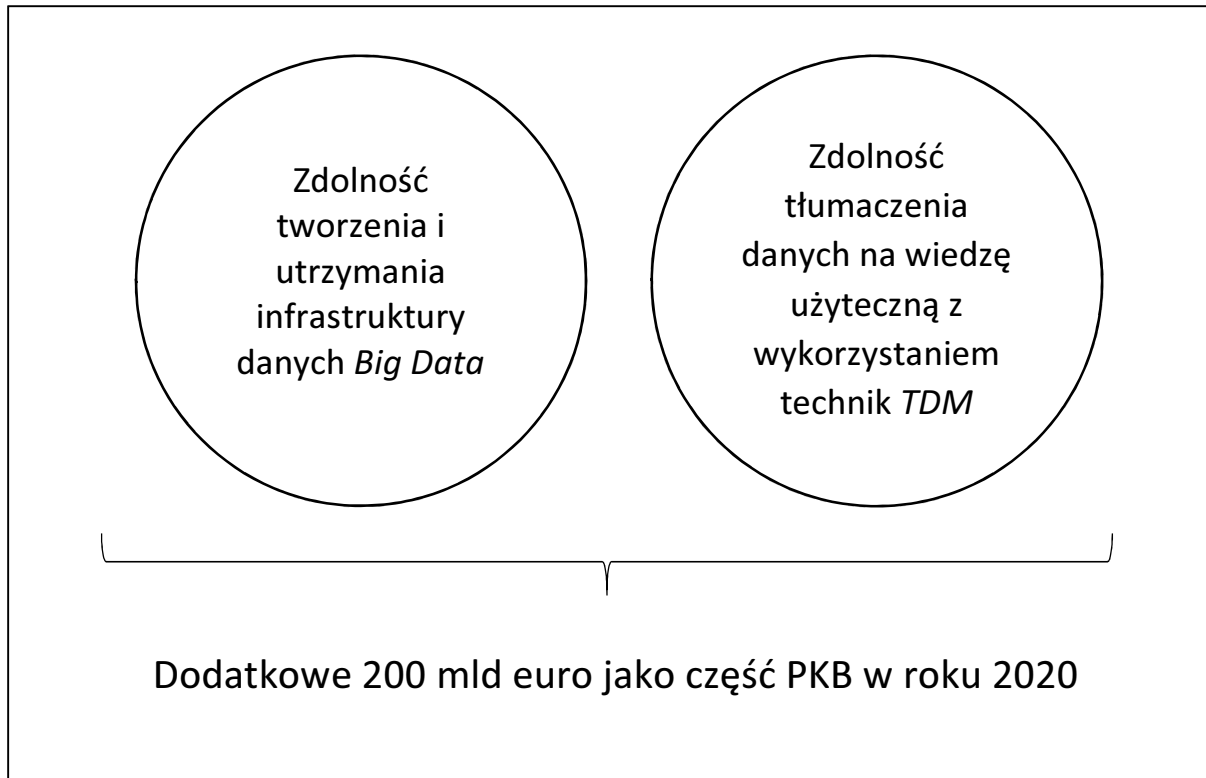
Warto zwrócić uwagę, iż efekty w skali mikro spowodowane wprowadzaniem innowacji osiągnięte przez odpowiednią liczbę przedsiębiorstw mogą doprowadzić do konkretnych efektów gospodarczych w makroskali. Obietnica dobrego wykorzystania *Big Data* oraz technik *TDM* została wyceniona dla gospodarki europejskiej na ok. 200 miliardów euro w roku 2020 – czyli ok. 1,9% europejskiego PKB (Buchholtz *et al.* 2014). Wynik ten zostanie osiągnięty, jeżeli firmy dobrze wykorzystają potencjał ukryty w analizie obszernych zbiorów danych do wprowadzania:

1. usprawnień nastawionych na osiągnięcie wyższej efektywności w odniesieniu do zarządzania produkcją, dystrybucją czy działaniem marketingowym,
2. innowacji produktowych i procesowych poprzez wykorzystanie danych w procesach badawczo-rozwojowych,
3. usprawnień w zarządzaniu dzięki decyzjom podejmowanym w oparciu o twarde dane (*Ibidem*: 11).

Warto pamiętać, iż efekty te będą wywołane tylko wówczas, gdy firmy będą w stanie skutecznie tworzyć własne infrastruktury danych typu *Big* oraz – co być może ważniejsze – rozwiną umiejętności transformacji tych danych w wiedzę użyteczną, czyli *actionable intelligence*, adekwatnie używając technik *TDM*. Każdy z tych komponentów jest równie ważny w umożliwieniu mikro- i makrozmiany uznanej za istotną przez ekspertów gospodarczych. Aktualnie wydaje się – co rozwijamy w następnym rozdziale tego raportu – iż pierwszy element tego równania skupia na sobie większość uwagi liderów biznesowych i politycznych, podczas gdy drugi (w którym znaczną rolę odgrywa *TDM*) pozostaje zaniedbany. Oznacza to, że *TDM* jest aktualnie słabym ogniwem dotychczasowej historii kapitalizacji zbiorów *Big Data*.

Schemat 1.

Logika konwersji Big Data na wartość gospodarczą z wykorzystaniem technik TDM



2. Perspektywa i doświadczenie firm

2.1 Praktyki i bariery

W tym rozdziale raportu prezentujemy różnorodne informacje na temat przedsiębiorstw, które pokazują, w jaki sposób podchodzą one do wdrażania *Big Data* i *TDM* na swoje potrzeby. Jak wyjaśniliśmy w poprzednim rozdziale, firmy, wdrażając rozwiązania zautomatyzowanej analityki danych, coraz aktywniej zabiegają o tłumaczenie gromadzonych przez nie danych na wartość biznesową.

„Sukces projektów Big Data leży w materializacji wartości biznesowej” (Capgemini & Informatica 2016)

Warto przy tym pamiętać, iż świat *Big Data* zmienia się bardzo dynamicznie, co stanowi równocześnie szansę i wyzwanie dla projektów analitycznych.

„Nowe i coraz bardziej zróżnicowane źródła danych – łącznie z logami webowymi, różnego rodzaju czujnikami czy danymi płynącymi z mediów społecznościowych – stają się pełnoprawnymi elementami analitycznego portfolio [...]” (Forbes Insights 2015: 2)

Można zatem powiedzieć, że ekosystem danych zmienia się w szybkim tempie. Jednak zadanie, które coraz bardziej świadomie stawiają przed sobą przedsiębiorstwa, można opisać jako wyzwanie polegające na zamianie konkretnego zasobu, jakim jest *Big Data*, w tzw. *actionable intelligence* przy użyciu metod i technik zautomatyzowanej eksploracji i analizy danych (*TDM*) w celu osiągnięcia wyników na poziomie większej innowacyjności i efektywności zarządzania firmą (patrz schemat poniżej).

Schemat 2.

Łańcuch tworzenia wartości TDM

Zasób	Proces	Aktywo	Efekt
Big Data >>	<< TDM >>	<< Actionable intelligence >>	>> Lepsza efektywność zarządzania zasobami
			>> Innowacje produktowe
			>> Ulepszenia w zarządzaniu przedsiębiorstwem

Warto nadmienić, iż w literaturze spotyka się również określenie „*Big Data and Analytics*” (*BDA*) służące podkreśleniu procesowego aspektu wyzwania związanego z wykorzystaniem danych w

praktykach biznesowych, w których *TDM* odgrywa zasadniczą rolę. Używając tego sformułowania, Buytendijk (2014) proponuje pewną klasyfikację ułatwiającą zrozumienie, co firmy rzeczywiście robią w tym zakresie (:9). I tak:

1. „Analityka *Big Data* wykorzystywana do lepszej obsługi klienta, analityka danych sprzedażowych, analityka *Big Data* w e-handlu oraz systemy obsługi klienta typu *intent-driven* stały się częścią **Analityki *Big Data* służącej do Zdobywania Informacji o Kliencie (*Customer Intelligence*)**.
2. Rozpoznawanie mowy oraz analiza tekstu, obrazu i video stały się elementem **Analityki Treści**.
3. Monitorowanie mediów społecznościowych stało się częścią **Analityki Społecznościowej (*Social Analytics*)**.
4. *Grid computing* oparte o rozwiązania chmurowe jest częścią **Analityki Chmurowej (*Cloud Computing*)**.
5. Tradycyjna analityka łańcucha dostaw została zastąpiona przez **Analitykę Łańcucha Dostaw *Big Data***”.

Łatwo zatem zrozumieć, że na aktualnym etapie rozwoju *Big Data* podstawowymi źródłami informacji dla przedsiębiorstw stają się: (1) dane zbierane w toku interakcji z klientem (na poziomie marketingu, dystrybucji czy sprzedaży), (2) mniej ustrukturyzowane dane tworzone przez firmy (jak np. wewnętrzna komunikacja czy dokumentacja), (3) wszelkie dane wytwarzane za pośrednictwem mediów społecznościowych oraz (4) dane gromadzone przez sensory zaangażowane w proces produkcji czy te związane z fenomenem internetu rzeczy. Warto jednak przy tym pamiętać, iż obraz ten podlega ciągłym i dynamicznym przekształceniom.

Firmy napotykają różnego rodzaju przeszkody, usiłując skorzystać z technologii *Big Data* i *TDM*. Klasyfikacji tych barier dokonała firma Gartner, która od 2013 roku prowadzi badania wyzwań związanych z praktykami *Big Data* wśród firm, które prowadzą (lub rozważają prowadzenie) tego typu projektów. Poniższa tabela przedstawia tę klasyfikację wraz ze szczegółowym wyjaśnieniem każdej kategorii.

Tabela 4. Bariery utrudniające prowadzenie i wykorzystanie zautomatyzowanej analityki w firmach (na podstawie: Gartner 2013)	
Nazwa	Opis
Ustalenie, jak uzyskać wartość z <i>Big Data</i>	Pokonanie tej bariery wymaga znajomości danych gromadzonych przez organizację oraz głębokiego zrozumienia modelu biznesowego firmy. Wyzwaniem jest powiązanie danych z obecnym modelem tworzenia i sprzedawania wartości. Wymaga to także wizjonerstwa i umiejętności teoretyzowania na temat tego, jak firma może zyskać, rozwijając strategię <i>Big Data</i> i działając na ich podstawie.

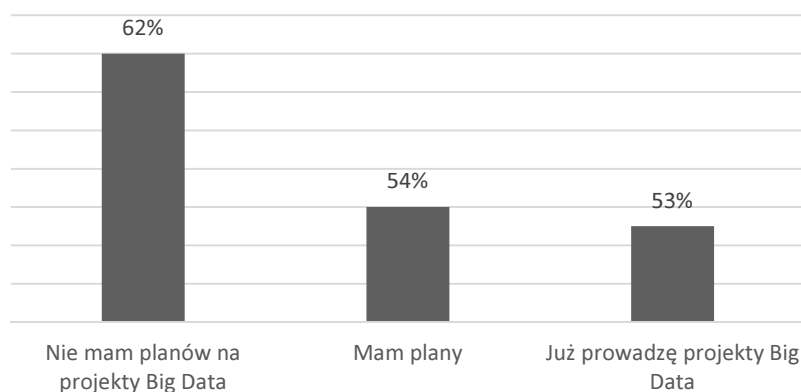
Posiadanie umiejętności i zdolności	Rozwój odpowiednich umiejętności we wszystkich aspektach związanych z wdrażaniem strategii rozwoju <i>Big Data</i> w danej firmie. Jednak przede wszystkim odnosi się to do trzech zasadniczych aspektów: (1) umiejętności informatycznych (aby stworzyć i utrzymać infrastrukturę danych) (2), analityki danych (aby przetłumaczyć dane na wdrażalne wnioski), (3) przywództwa (aby zarządzać zmianami wywołwanymi przez wdrażanie projektów <i>Big Data</i>).
Kwestie ryzyka i zarządzania	Odnoszą się do bezpieczeństwa, prywatności i kwestii prawnych, ale dotyczą również jakości danych. Jeśli decyzje powinny być podejmowane pod wpływem analityki danych, dane muszą być solidne. Wyzwaniem staje się zatem określenie wiarygodnych źródeł i dbanie o jakość danych.
Finansowanie inicjatyw związanych z <i>Big Data</i>	<i>Big Data</i> będzie wymagało finansowania wewnętrznego lub zewnętrznego. To dlatego duże firmy postrzegane są jako przodownicy rewolucji <i>Big Data</i> – mają one bowiem przeważnie większy dostęp do niezbędnych zasobów.
Określanie strategii	Opracowanie właściwej strategii w celu wdrożenia zautomatyzowanej analityki danych w procesy danego przedsiębiorstwa.
Integracja wielu źródeł danych	Wyzwanie to łączy się z koniecznością integracji wielu źródeł danych, występujących często w różnorodnych formatach. Wiąże się ono z podejmowaniem decyzji, które źródła danych stanowią cenne źródła informacyjne, aby ostatecznie skupiać się na kluczowych źródłach.
Integracja technologii <i>Big Data</i> z istniejącą infrastrukturą	Firmy przez lata rozwijają struktury IT. Bardzo często są one podzielone między różne działy firmy. Wdrażanie strategii <i>Big Data</i> oznacza, że musi zostać zbudowana infrastruktura danych poświęcona <i>Big Data</i> . Należy ją także zintegrować z istniejącymi już strukturami. Integracja jest z kolei problemem zarówno inżynierskim, jak i analitycznym.
Infrastruktura lub architektura	Opracowanie właściwej architektury IT <i>Big Data</i> . Wybór technologii, która ma być stosowana.
Kwestie przywódcze i organizacyjne	Jak w przypadku każdego procesu biznesowego, konieczne jest dobre przywództwo i zmiana organizacyjna. Szczególnie w projektach <i>Big Data</i> wymaga się nowych stylów zarządzania lub praktyk, które mogą napotkać przeszkody organizacyjne, stąd konieczność dalekiej determinacji z poziomu przywództwa firmy.
Zrozumienie, czym jest <i>Big Data</i>	Niektóre firmy nadal mają problemy ze zrozumieniem tej koncepcji w całej jej złożoności. Wyzwanie to wymaga intensywnego gromadzenia wiedzy i inspiracji w dziedzinie wykorzystania <i>Big Data</i> i zautomatyzowanej analityki przez inne organizacje. Bez zrozumienia czym jest <i>Big Data</i> , nie sposób myśleć kreatywnie o możliwościach wykorzystania tego fenomenu.

Powyższe kategorie można również odczytywać jako warunki, które należy spełnić w danej firmie, aby mogła ona w sposób adekwatny gromadzić dane oraz przekształcać je w decyzje biznesowe. Organizacja taka musi zatem po pierwsze przejść przez proces kreatywnego myślenia na poziomie strategicznym, w wyniku którego dostrzeżone zostaną szanse łączenia *Big Data* i zautomatyzowanej analityki z wyzwaniami aktualnego modelu biznesowego firmy. Myślenie to jest warunkiem stworzenia konkretnej strategii *Big Data* z wykorzystaniem technologii *TDM*. Powstała strategia wykreuje realne zapotrzebowanie na talenty i umiejętności, co stworzy lub zmodyfikuje infrastrukturę danych i zintegruje ją z istniejącą infrastrukturą IT. Następnie będzie można wdrożyć strategię z uwzględnieniem całkowitego ryzyka i trudnych aspektów związanych z zarządzaniem zmianą.

Ten poziom strategicznego ulokowania inicjatyw związanych ze zautomatyzowaną analityką danych wydaje się ciągle jednym z najważniejszych wyzwań. Badania realizowane przez firmę Gartner regularnie identyfikują „określenie, jak wydobyć wartość z danych”, jako główną trudność napotykaną przez liderów biznesowych. W roku 2016 52% uczestników badania określiło ten aspekt jako problematyczny, w roku 2015 – 55%, a trzecia część tej grupy uznała go za wyzwanie priorytetowe. Analiza danych z roku 2015 pokazuje, iż odsetek liderów wskazujących ten aspekt jako problematyczny rośnie do 62% w grupie osób, które jak dotąd nie planują realizować projektów typu *Big Data* u siebie w firmach. Jednak utrzymuje się on na wysokim poziomie także wśród osób, które prowadzą już tego typu projekty.

Wykres 1.

Odsetek liderów biznesowych zgłaszających trudność związaną z określeniem wartości Big Data w poszczególnych grupach uporządkowanych ze względu na etap prowadzenia inwestycji Big Data

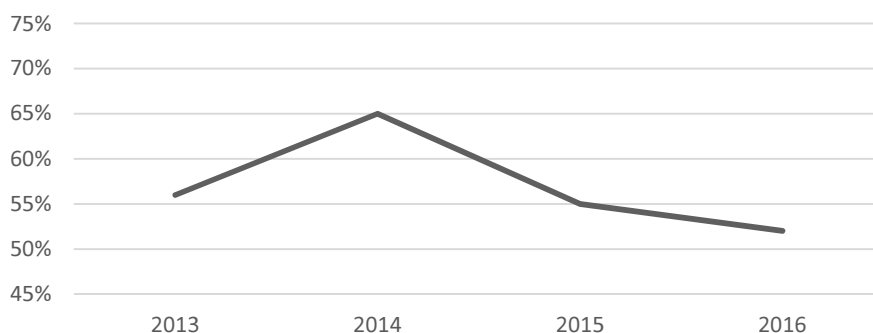


Źródło: Gartner (2015)

Wydaje się zatem, że jeden z poważniejszych problemów związanych z użytecznym wykorzystaniem danych i ich zautomatyzowanej analityki ma charakter fundamentalny. Nieco optymizmu w tym zakresie może jednak wzbudzić fakt, że intensywność tego problemu jest w trendzie spadkowym.

Wykres 2.

Odsetek liderów biznesowych określających identyfikację wartości danych Big Data jako problematyczną

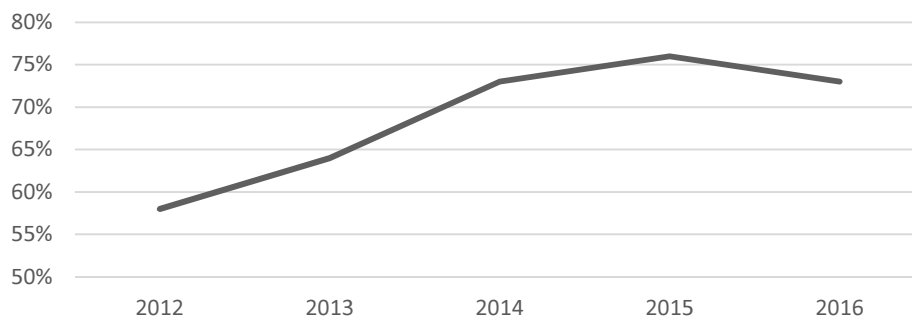


Źródło: Gartner (2013, 2014, 2015, 2016)

Zdecydowanie zwiększa się również odsetek firm, które już wdrażają bądź planują wdrożyć rozwiązania *Big Data* i *TDM*. W 2013 roku 56% badanych przez firmę Gartner rozważało prowadzenie lub prowadziło już projekty tego typu. W 2016 roku wartość ta wzrosła do 73%.

Wykres 3.

Odsetek liderów biznesowych planujących lub prowadzących wdrożenie projektów zautomatyzowanej analityki danych



Źródło: Gartner (2016)

Warto przy tym zauważyć, iż 59% wciąż pozostaje na etapie planowania lub wstępnego eksperymentowania, a jedynie w przypadku 15% badanych można mówić o zaawansowanej implementacji rozwiązań *Big Data* i *TDM*.

Schemat 3.

Etapy wdrażania projektów Big Data

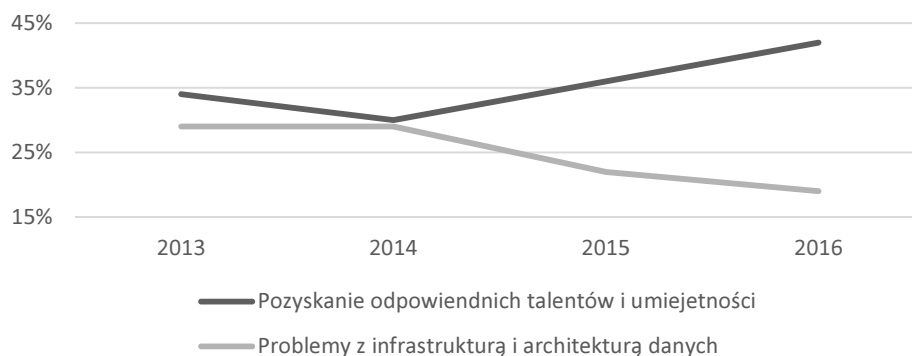


Źródło: Gartner (2016)

Warto zwrócić uwagę, że zapotrzebowanie na zasoby konkretnie związane z technologiami *TDM* będzie rosło w miarę dojrzewania dyscypliny *Big Data*. Założenie to znajduje potwierdzenie w danych wywoływanych i gromadzonych przez firmę Gartner. W miarę dojrzewania zautomatyzowanej analityki jako praktyki biznesowej firmy coraz bardziej odczuwają brak odpowiednich talentów i umiejętności, które są bardziej potrzebne w procesie translacji danych na wartość biznesową – co jest domeną procesów *TDM*. Należy także zauważyć, iż coraz bardziej odczuwalnym niedoborem umiejętności towarzyszy spadek problemów związanych z tworzeniem odpowiedniej infrastruktury i architektury danych. Trend ten został zwizualizowany na poniższym wykresie i oznacza, że wyzwania natury bardziej technicznej tracą na znaczeniu w obliczu tych związanych z umiejętnościami odpowiedniej analizy – czyli charakterystycznych dla technologii i procesów *TDM*.

Wykres 4.

Zmiana w odczuwaniu przez liderów biznesowych poszczególnych barier związanych z projektami Big Data

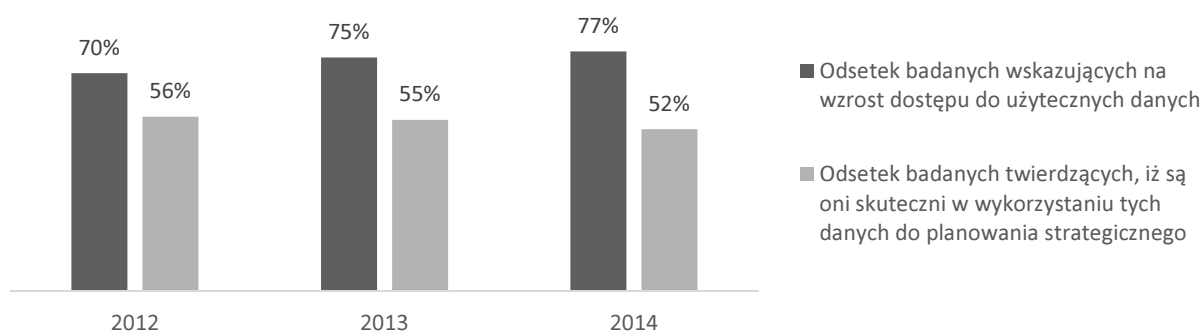


Źródło: Gartner (2016)

Kierunek, w jakim rozwijają się praktyki i potrzeby firm w odniesieniu do zautomatyzowanej analityki, sugeruje wyraźnie odczuwalną niską satysfakcję, jeśli chodzi o dotychczasowe wyniki projektów budowanych wokół idei *Big Data*. Raport wydany przez MIT Sloan Review w roku 2015 wskazuje, że z roku na rok zwiększa się odczuwalny dostęp do użytecznych danych, jednak umiejętności wykorzystania wyników analizy do realizacji celów strategicznych zmniejszają się (Ransbotham, Kiron i Prentice 2015: 3).

Wykres 5.

Dostęp do danych a ich użyteczność



Źródło: Ransbotham, Kiron i Prentice (2015: 3)

Do pewnego stopnia problem ten może być związany z faktem, iż wiele firm rozwijało swoje projekty związane z *Big Data* raczej od strony technologicznej. Podejście to gwarantowało doskonałość techniczną, jednak nie zabezpieczało wyników biznesowych. Jak zauważają autorzy raportu wydanego przez firmę Mu Sigma², „wiele firm ciągle błędnie kładzie nacisk na potrzeby technologiczne ponad potrzeby związane z lepszym podejmowaniem decyzji” (2016). Intuicje te potwierdzają badania przeprowadzone przez Forbes Insights:

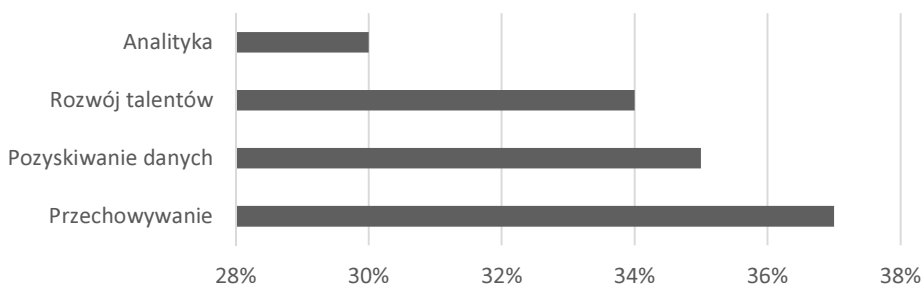
„Największy odsetek respondentów (37%) czyni zasadnicze inwestycje w infrastrukturę przechowywania danych, podczas gdy 35% skupia się na pozyskiwaniu danych. Odpowiedzią o najniższej liczbie wskazań są inwestycje w analitykę (30%), a jedynie 34% respondentów alokuje fundusze na rozwój talentów. Takie ustawienie priorytetów pokazuje, że praktyka zautomatyzowanej analityki danych jest ciągle jeszcze w powijakach.

W rzeczywistości duże inwestycje w zdolności przechowywania danych jasno pokazują, iż wiele organizacji jest ciągle na przedwstępnych etapach zarządzania, przechowywania i porządkowania wielkich zbiorów danych”.

² Firma zajmująca się analityką danych, której wartość Forbes wycenił na ponad miliard dolarów.

Wykres 6.

Priorytety względem poszczególnych aspektów inwestycji w projekty zautomatyzowanej analityki danych

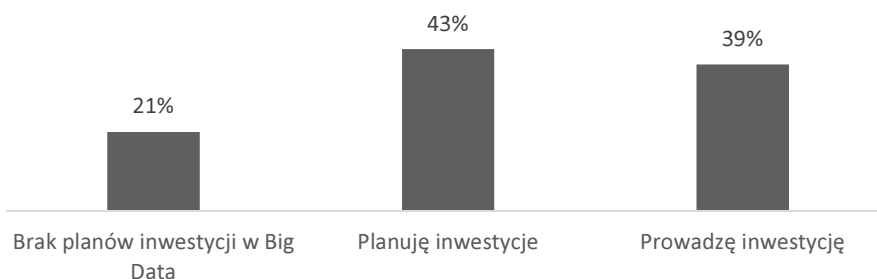


Źródło: Forbes Insights (2015: 14)

Jak jednak pokazują cytowane już tutaj badania firmy Gartner, wyzwania odczuwane przez firmy stają się o wiele bardziej praktyczne w miarę nabierania pierwszych doświadczeń we wdrażaniu zautomatyzowanej analityki. Przykładowo, jedynie 21% badanych nieplanujących inwestowania w *Big Data* wskazało w roku 2015, iż pozyskiwanie umiejętności i talentów jest dla nich wyzwaniem. Jednocześnie odsetek ten rósł ok. dwukrotnie, w miarę jak organizacja konkretyzowała swoje plany inwestycyjne, angażując się w planowanie i wdrożenie. Oznacza to, iż odpowiednie talenty i umiejętności są tym aspektem, który lekceważy się na początku wdrożenia, jednak organizacje bardzo szybko zaczynają go odczuwać po podjęciu konkretnych kroków inwestycyjnych. Trend ten pozwala przewidywać, iż umiejętności związane z eksploracją i analizą danych (specyficzne dla technik i procesów *TDM*) będą w najbliższych latach dynamicznie zyskiwać na wartości, ponieważ stanowią one warunek konieczny zrealizowania obietnicy związanej z *Big Data*.

Wykres 7.

Odsetek liderów biznesowych zgłaszających trudności ze zdobyciem umiejętności w poszczególnych grupach uporządkowanych wg zaawansowania prowadzenia inwestycji Big Data



Źródło: Gartner (2015)

Wyzwania, które również zyskują na znaczeniu, w miarę jak organizacja nabiera doświadczenia w rozwijaniu projektów zautomatyzowanej analityki danych, to zarządzanie ryzykiem (także bezpieczeństwem), integracja wielu źródeł danych i integrowanie infrastruktury *Big Data* z istniejącą infrastrukturą IT. Co ciekawe, wyzwania takie jak „określenie, jak wypracować wartość z *Big Data*” czy definiowanie strategii traci na znaczeniu, gdy organizacja wchodzi w etapy planowania i wdrażania inicjatyw *Big Data*. Oznacza to, że rozpoczęcie prac związanych z wdrożeniem bardzo często stanowi warunek konieczny dostrzeżenia realnego potencjału *Big Data* w danej firmie. Innymi słowy, potencjał ten trudniej jest dostrzec w sensie teoretycznym. W praktycznym wymiarze oznacza to, iż firmy muszą zainwestować niejako „w ciemno” i „spróbować, by się przekonać”. W związku z tym w przestrzeni publicznej wciąż mało jest przekonujących dowodów na opłacalność inwestycji w zautomatyzowaną analitykę danych.

Wykres 8.

Odsetek organizacji określających „rozumienie, jak wypracować wartość w oparciu o Big Data”, jako problematyczne w zależności od etapu wdrożenia



Źródło: Gartner (2015)

Kolejnym kluczowym wyzwaniem, które firmy coraz bardziej odczuwają, jest kultura organizacyjna dominująca w przedsiębiorstwach usiłujących skorzystać z technologii *Big Data* i *TDM*. Jak wskazuje Forbes Insights:

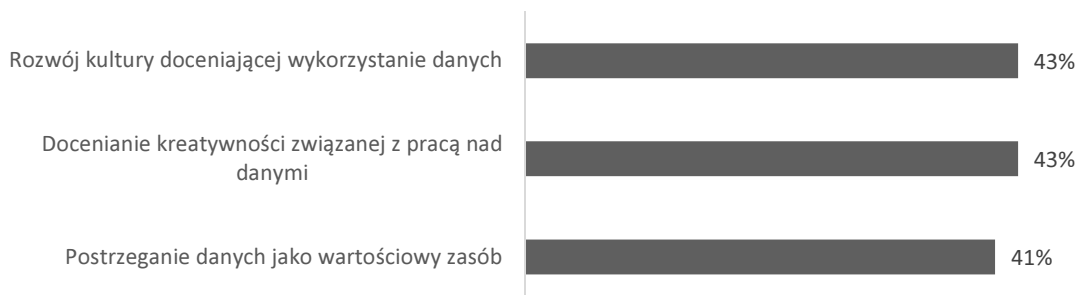
„Częścią problemu jest to, że modele biznesowe napędzane przez dane są tak odmienne od modeli biznesowych przeszłości, że ich wprowadzanie może spowodować prawdziwy wstrząs w organizacji. W zasadzie z dnia na dzień pracownicy muszą uczynić dane swoim priorytetem i muszą być gotowi dzielić się otwarcie informacjami z innymi departamentami, a także przyjąć wspólną odpowiedzialność za zbieranie danych, ich jakość i analizę. Taka reforma może spowodować, że pracownicy będą – delikatnie mówiąc – powściągliwi”.

(Forbes Insights 2015: 11-12)

Według badań aż 43% firm przebadanych przez *Forbesa* wskazuje na znaczące trudności w rozwijaniu kultury, „która docenia użycie danych” lub „docenia kreatywność i eksperymenty związane z analizą danych”. Z kolei 41% liderów biznesowych twierdzi, iż napotyka trudności, kiedy próbują sprawić, aby ich pracownicy postrzegali dane „jako wartościowy zasób”.

Wykres 9.

Trudności zgłaszane przez liderów biznesowych w odniesieniu do formowania odpowiedniej kultury organizacyjnej



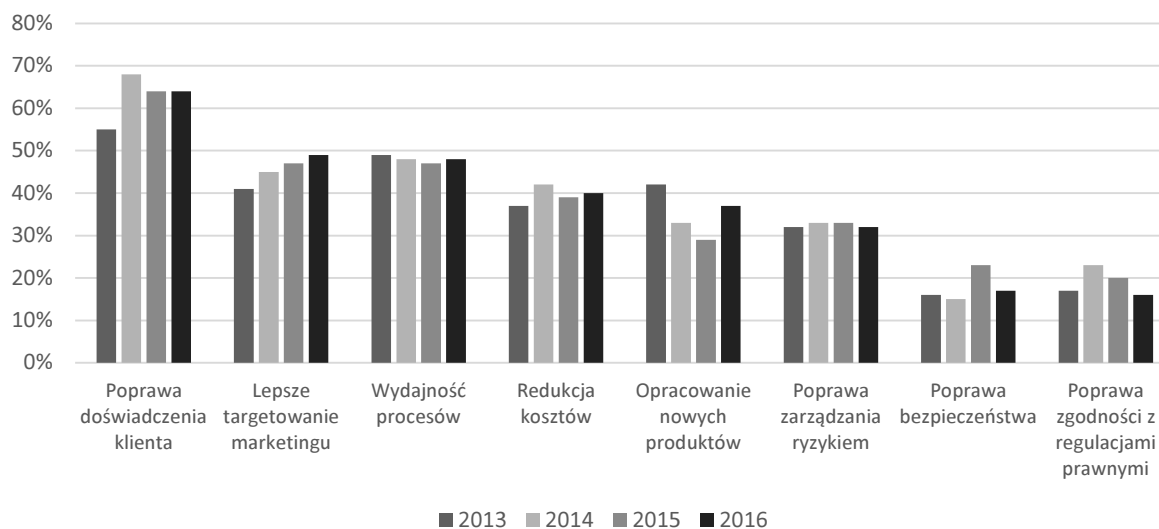
Źródło: Forbes Insights (2015: 15)

2.2 Zróżnicowanie branżowe i źródła danych

Dane wywoływane przez różne podmioty analizujące praktyki firm związane ze zautomatyzowaną analizą danych pozwalają również na wskazanie trendów w odniesieniu do bardziej precyzyjnych celów takiej analityki na poziomie mikro. Badania firmy Gartner pokazują, że trendy pozostają tutaj raczej niezmiennie od roku 2013. Głównymi celami stojącymi za inwestowaniem w projekty *Big Data* jest poprawa doświadczenia klienta, optymalizacja procesów nastawiona na poprawę efektywności oraz lepsze targetowanie działań marketingowych.

Wykres 10.

Intensywność motywacji wdrażania projektów zautomatyzowanej analityki danych



Źródło: Gartner (2016: 8)

Te trzy motywacje wydają się niezmiennie dominować. Kiedy jednak spojrzeć na problem przez pryzmat poszczególnych branż, okazuje się, przykładowo, że sektory bankowości, ubezpieczeń oraz sektor zdrowia są o wiele bardziej zainteresowane wykorzystaniem analityki danych do lepszego zarządzania ryzykiem. Ponad 56% (wartość uśredniona) badanych firm związanych z tymi sektorami wskazuje, iż zarządzanie ryzykiem jest głównym powodem inwestycji w *Big Data*, podczas gdy wskaźnik ten wynosi jedynie 26% dla przedstawicieli pozostałych sektorów. Dane sugerują także, że sektor produkcyjny, sektor firm odpowiedzialnych za dostarczanie usług komunalnych oraz sektor rządowy nie są tak bardzo zainteresowane poprawą doświadczenia użytkownika końcowego, jak inne sektory.

Tabela 5. Cele wykorzystania zautomatyzowanej analityki danych przez przedsiębiorstwa wg sektorów

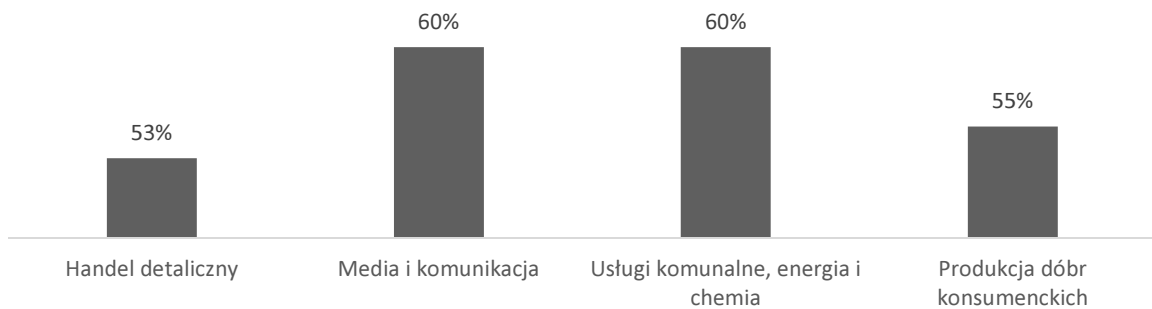
	Prod.	Media i kom.	Usługi	Admin. publiczna	Eduk.	Handel	Bank.	Ubezpieczenia	Sektor zdrowia	Transp.	Usługi komunalne
Poprawa dośw. klienta	52%	78%	66%	43%	76%	83%	77%	77%	73%	69%	44%
Lepsze targetowanie marketingu	43%	89%	53%	17%	41%	78%	66%	58%	-	38%	17%
Wydajność procesów	45%	33%	35%	49%	65%	43%	41%	50%	73%	69%	78%
Redukcja kosztów	42%	33%	35%	37%	35%	30%	41%	31%	45%	56%	61%
Opracowanie nowych produktów	23%	67%	37%	14%	24%	35%	27%	50%	-	19%	33%
Poprawa zarządzania ryzykiem	14%	22%	29%	29%	35%	22%	52%	58%	55%	31%	61%
Poprawa bezp.	17%	22%	21%	34%	29%	13%	27%	27%	9%	19%	28%
Poprawa zgodności z regulacjami prawnymi	11%	22%	18%	23%	18%	9%	25%	23%	27%	31%	44%

Źródło: Gartner (2015)

Z kolei sektory handlu detalicznego, mediów i komunikacji, dostawców usług komunalnych i energii oraz sektor dóbr konsumenckich zostały zidentyfikowane przez firmy Capgemini i Informatica jako te, w których firmy najczęściej wdrażają rozwiązania typu *Big Data* i *TDM*. Badania pokazują, iż w każdym z tych sektorów 60% badanych firm wdraża podobne rozwiązania.

Wykres 11.

*Sektory najczęściej wdrażające
zautomatyzowaną analitykę danych*

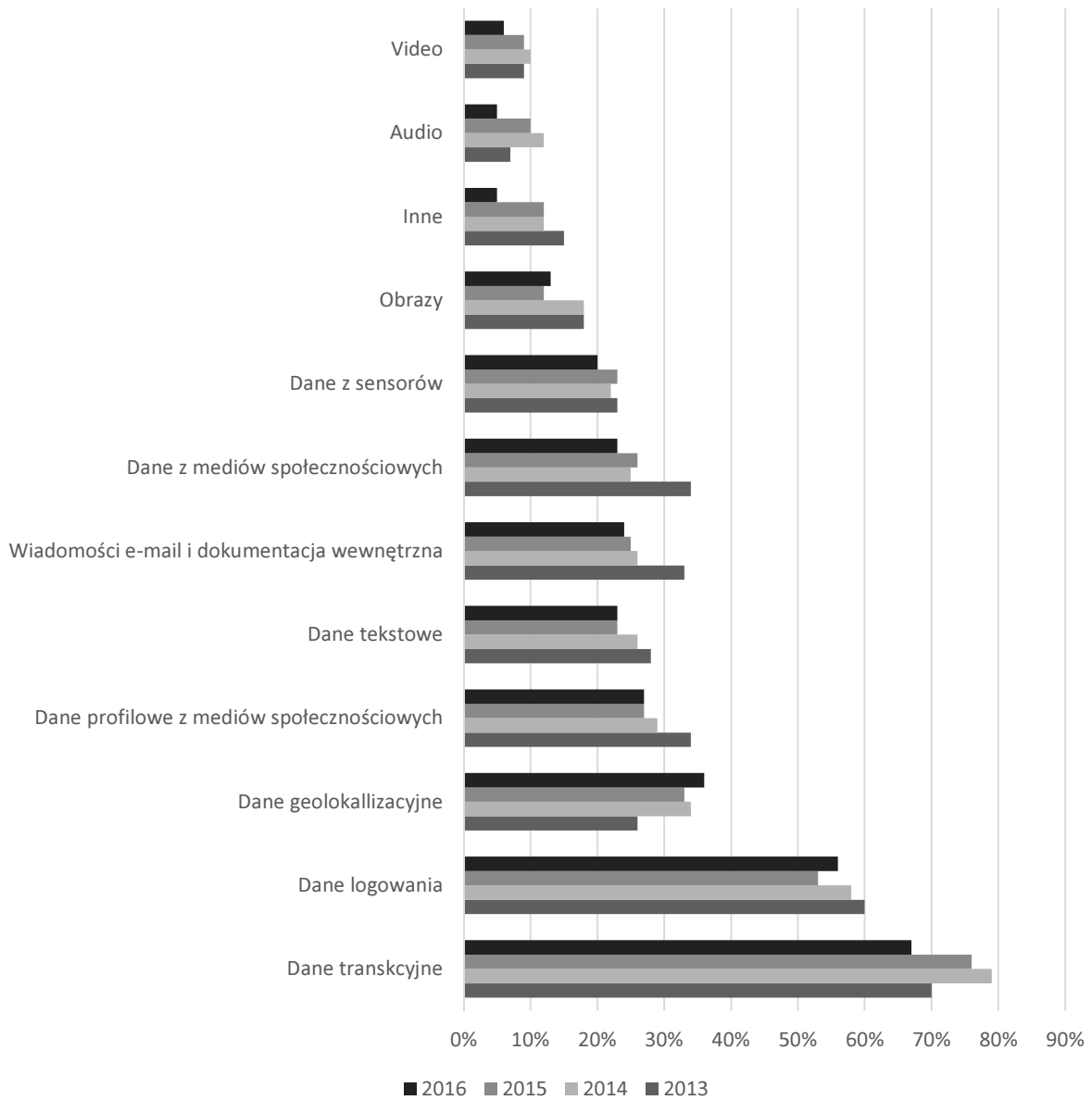


Źródło: Capgemini & Informatica (2016: 9)

Jeśli chodzi o źródła informacji, nie powinno dziwić, iż firmy – niezależnie od sektora – analizują głównie dane transakcyjne. Korzystanie z tych danych jest na ogół dwukrotnie częstsze niż w przypadku innych rodzajów informacji. Firmy mają nadzieję, że to właśnie w tych danych zaszyfrowane są informacje, które pozwolą na ulepszenie doświadczenia klienta, czy na lepsze dopasowanie działań marketingowych. Transakcje pozwalają bowiem śledzić nawyki zakupowe, których wykorzystanie w praktykach biznesowych stało się już relatywnie popularne. Często analizowane są także dane o zachowaniach użytkowników logujących się na różnych platformach webowych. Na znaczeniu zyskują również dane geolokalizacyjne. Jednak firmom ważne wydają się również dane pozyskiwane z mediów społecznościowych czy dokumentacji i komunikacji wewnętrznej – choć ich wykorzystanie maleje z biegiem lat, co może wskazywać na rozczarowanie przełożeniem analizy tych danych na wyniki biznesowe. Wykorzystanie danych audio, video czy danych graficznych jest według badań firmy Gartner praktyką relatywnie marginalną.

Wykres 12.

Rodzaje wykorzystywanych danych

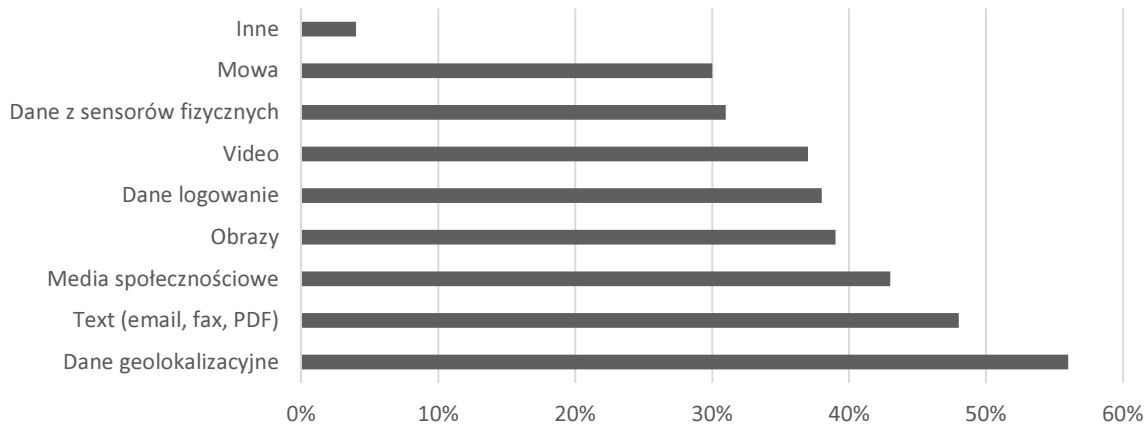


Źródło: Gartner (2016)

Nieco inne spojrzenie na rodzaje używanych danych prezentuje raport Forbes Insights. Według informacji zebranych przez ten ośrodek dane video czy graficzne analizowane są w sposób zautomatyzowany niemal równie często jak dane z mediów społecznościowych. Niemniej zarówno dane Forbesa, jak i dane firmy Gartner wskazują, że analiza tekstu jest znaczącym elementem praktyk analitycznych.

Wykres 13.

Rodzaje wykorzystywanych danych



Źródło: Forbes Insights (2015: 8)

Wydaje się, że tekst w otwartej formie jest aktywem jeszcze bardziej znaczącym na rynku europejskim, zbadanym przez firmę Fujitsu. Ponadto dynamicznie zyskuje on na znaczeniu wśród tych firm, które są już na etapie wdrożeń projektów *Big Data*.

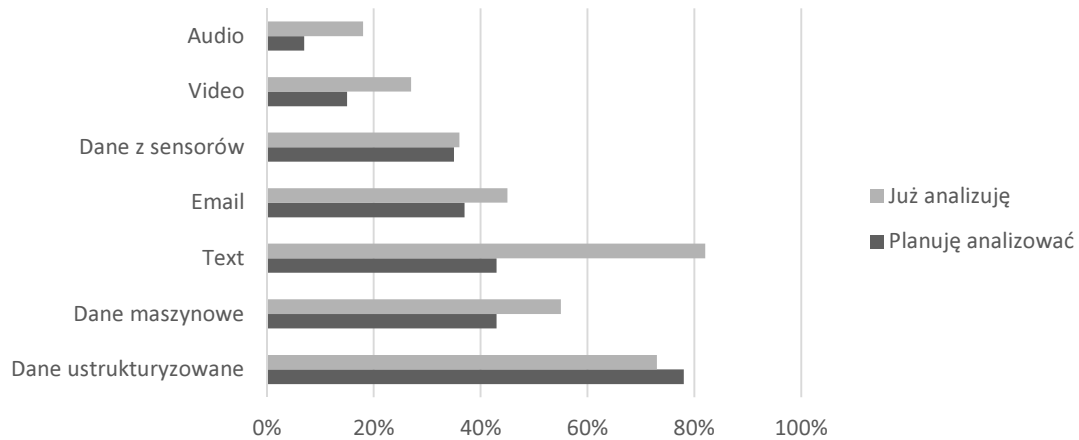
„Dane nieustrukturyzowane, szczególnie w formie otwartego tekstu, eksplorowane są dwukrotnie częściej przez doświadczonych praktyków Big Data”

Fujitsu Market Insights: Big Data (2015: 3)

Co ciekawe, po przeprowadzeniu wstępnych eksperymentów tekst okazuje się mieć większy potencjał niż dane ustrukturyzowane – jak np. dane transakcyjne. Na poniższym wykresie przedstawione są wskazane przez liderów biznesowych rodzaje danych ocenianych jako wartościowe z podziałem na dwie grupy organizacji – te, które planują dopiero wdrożyć projekty analizy *Big Data* oraz te, które już takie projekty realizują. Okazuje się, że w tej drugiej grupie dane tekstowe uznawane są za wartościowe przez prawie dwukrotnie większą liczbę liderów. Oznacza to, że dane tego rodzaju ukazują swoje znaczenie dla analizy dopiero w trakcie prowadzenia wdrożeń. Może to znaczyć, iż wartość tych danych jest dalece bardziej kontekstowa od, przykładowo, danych transakcyjnych czy danych logowania. Jest ona jednak rzeczywista i ma duży potencjał biznesowy.

Wykres 14.

Rosnąca wartość danych tekstowych



Źródło: Fujitsu (2014: 3)

Rozpoznaną wartość analizy nieustrukturyzowanego tekstu potwierdza ostatnie badanie firmy Gartner, z którego dowiadujemy się, iż jedynie 23% badanych organizacji aktualnie analizuje dane tego typu, a 33% planuje w najbliższym czasie dodać tego typu dane do swojego miksu analitycznego (Gartner 2016: 15).

2.3 Konkluzja – ludzkie talenty analityczne niezbędnym składnikiem konwersji *Big Data* na wartość gospodarczą

Praktyki, bariery i trendy stosowania technologii *Big Data* i *TDM* w przedsiębiorstwach opisaliśmy w tym rozdziale na podstawie informacji przekazywanych głównie przez liderów biznesowych i gromadzonych przez poszczególne firmy analityczne, na których raporty się powołujemy. Wyniki tych analiz pozwalają wyciągnąć wniosek, iż zautomatyzowana analiza danych nie jest już praktyką niszową, a raczej w relatywnie krótkim czasie stała się pełnoprawną częścią procesów biznesowych. Mimo jednak dużej aktywności firm w tym zakresie, należy stwierdzić, iż stadium zaawansowania tych procesów – i w związku z tym umiejętności ich wykorzystania na rzecz budowania wartości gospodarczej – jest jeszcze wczesne. Wiele organizacji wciąż zmaga się z warunkami wstępnymi skutecznego wykorzystania technologii *Big Data*, a zatem określeniem, jakiego rodzaju wartość może być wynikiem inwestycji firmy na tym polu. Wiele firm eksperymentuje bez jasnej wizji i strategii dotyczącej tego, jak doprowadzić do kapitalizacji danych. Doprowadza to do sytuacji, w której firmy inwestują w pierwszej kolejności w technologie przechowywania i zarządzania danymi, licząc, że aktywność ta sama w sobie przyniesie określone profity. Jednak – jak pokazują przedstawione tu badania – ten tryb postępowania szybko prowadzi do rozczarowania. Okazuje się bowiem, iż bez odpowiedniej jakości talentów eksploratorskich i analitycznych nawet najlepsza infrastruktura pozostaje jedynie kosztem dla firmy i nie można jej zamienić na aktyw o realnej wartości rynkowej.

Dodatkowym wyzwaniem jest fakt, iż umiejętności eksploracji i analizy informacji wymagają wzmocnienia nie tylko w tych grupach pracowników, którzy odpowiedzialni są bezpośrednio za pracę z danymi. Umiejętność ta powinna stać się ważnym elementem kultury organizacyjnej – dopiero

spełnienie tego warunku umożliwi szybkie i skuteczne wykorzystanie danych do podejmowania decyzji we wszystkich departamentach danego przedsiębiorstwa. Niestety reforma kultury organizacji w stronę większej przychylności danym i analityce jest aktualnie jednym z poważniejszych wyzwań. Bariera ta sprawia, iż wyniki pracy nawet dobrego zespołu analitycznego nie są konsumowane przez organizację. Jak określają to eksperci *MIT Sloan Management Review*, „konsumpcja” analityki jest problemem typu „ostatniej mili”. Badania przeprowadzone przez ten ośrodek pokazały, że jedynie 27% organizacji zadeklarowało odniesienie sukcesu na polu integracji talentów analitycznych z innymi kluczowymi departamentami biznesowymi (Ransbotham, Kiron, Prentice 2014). Aby zaradzić temu wyzwaniu, 49% organizacji wdraża szkolenia menedżerów, aby podwyższyli oni swoje umiejętności analizy i rozumienia danych, a 34% szkoli zespoły analityków odpowiedzialnych za wykonanie *TDM*, aby lepiej zrozumieli, jak działa cała organizacja i jej poszczególne działy. Okazuje się zatem, iż aby skorzystać z możliwości zaawansowanych technologii analitycznych, nie można pomijać komponentu ludzkiego. Być może powinien on nawet stać się punktem wyjścia wdrażania zautomatyzowanej analityki danych w przedsiębiorstwach.

3. Rynkowa i gospodarcza wartość *Big Data* i *TDM*

W poprzednich rozdziałach zwracaliśmy uwagę na to, jak ważne dla firm jest rozwinięcie talentów analitycznych oraz przekształcenie kultury organizacyjnej, aby generalne postawy względem zautomatyzowanej analityki danych były pozytywne wśród wszystkich pracowników merytorycznych organizacji. Osiągnięcie tych dwóch celów oznacza wprowadzenie daleko idącej zmiany w działaniu danego przedsiębiorstwa. Zmiana ta jest prowadzona z wykorzystaniem pomocy zewnętrznej i wiąże się ze zgłaszaniem przez firmy popytu na wiedzę, technologie, dane czy usługi pozwalające na wdrażanie projektów zautomatyzowanej analityki danych. Potrzeba tej zmiany tworzy zatem rynek – w tym rozdziale szacujemy jego wartość, pokazując perspektywy jego rozwoju globalnie i w kontekście europejskim. Przytaczając różne szacunki finansowe, zwracamy również uwagę, iż wartość rynku nie jest równoznaczna z jego całościowym znaczeniem dla ogółu gospodarki. Ten holistyczny ogląd wymaga bowiem wzięcia pod uwagę różnego rodzaju efektów, które mogą pozytywnie oddziaływać na całość ekosystemu ekonomicznego. Kwestię tę przybliżamy w ostatnich częściach tego rozdziału.

Prezentując estymacje, warto także przypomnieć, iż znajomość wartości rynku *Big Data* pozwala również zrozumieć wartość technologii *TDM*. Część bowiem rynku związana jest bardziej z tworzeniem i rozwojem infrastruktury danych. Natomiast druga jego część skupia się bardziej na eksploracji i analizie danych – co jest specyficzne dla *TDM*.

Ramka 6. Przypomnienie definicji *Big Data*

Sformułowanie *Big Data* odnosi się do wielkich zbiorów danych. Różne organizacje zaczęły wdrażać rozwiązania *Big Data*, aby uzyskać dostęp do tych wielkich zbiorów i móc je analizować. Zbiory te podczas analizy są w stanie dostarczyć wniosków praktycznych o charakterze *actionable intelligence*. Wnioski te pomagają organizacjom podejmować szybkie decyzje, co pozwala jeszcze lepiej dopasować ich ofertę do potrzeb klienta. W ciągu ostatnich lat fenomen *Big Data* zyskał bardzo dużo uwagi. Co więcej, efektywność kosztowa osiągnięta dzięki rozwiązaniom *Big Data* oraz spełnienie warunków bezpieczeństwa danych doprowadziły do znaczącego rozwoju rynku.

Rynek *Big Data* ukonstytuował się w roku 1999, kiedy po raz pierwszy użyto tego sformułowania w publikacji zatytułowanej „*Visually Exploring Gigabytes in Real Time*”. Rok później rozmiar danych rzeczywiście osiągnął skalę gigabajtów i zaczęto oferować usługi pozwalające na opanowanie takiej ilości danych. Ponadto ukończono termin „trzech V” do opisanie danych typu *Big*: prędkość (*velocity*), wielkość (*volume*) i różnorodność (*variety*).

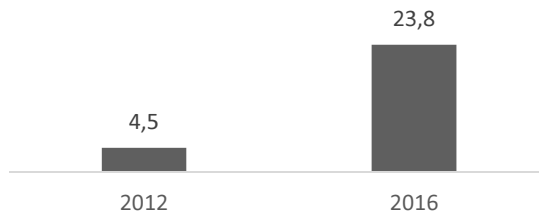
Źródło: *Research and Markets (2015)*

3.1. Szacunki globalne i europejskie

Rynek *Big Data* rośnie w bardzo szybkim tempie. Globalna firma doradcza IDC wyliczyła, iż w roku 2010 rynek ten wart był 4,5 miliarda dolarów. Przewidywano wówczas, iż urośnie on do rozmiaru 23,8 miliarda dolarów w roku 2016 (IDC 2012).

Wykres 15.

Wartość globalnego rynku Big Data w mld USD



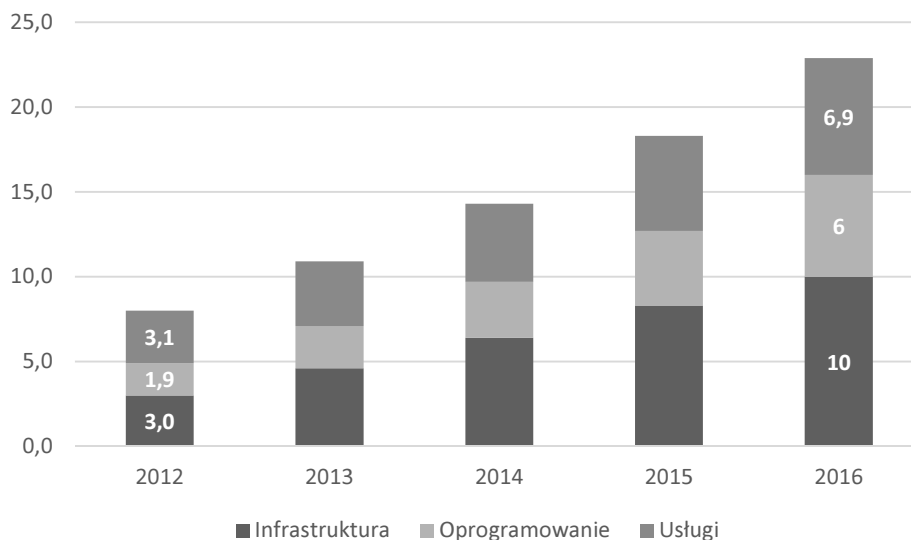
Źródło: IDC (2012: 5-6)

Analizując te wyniki, należy pamiętać, iż nie jest to szacunek wpływu ekonomicznego rozumianego jako suma zysków i wartości otrzymanej dzięki zautomatyzowanej analizie danych. Kwoty te są spojrzeniem wąskim, czyli reprezentują to, ile firmy wydają na technologie związane z *Big Data* i *TDM* w danym roku kalendarzowym.

Metodologia zbierania informacji od firm pozwala na rozbitcie tych wydatków na trzy główne kategorie: (1) infrastrukturę, (2) oprogramowanie oraz (3) usługi. W roku 2012 firmy wydały 3,1 miliarda dolarów na zakup produktów z pierwszej kategorii, 1,9 miliarda z drugiej i 3 miliardy z trzeciej kategorii. Oszacowano wówczas, iż wydatki te dynamicznie wzrosną, osiągając średni roczny wskaźnik wzrostu na poziomie 31,7% w latach 2011-2016.

Wykres 16.

Wzrost wartości rynku Big Data dla produktów z poszczególnych kategorii w mld USD



Źródło: IDC (2012: 5-6)

Wydatki na infrastrukturę obejmują wszelkie inwestycje związane z budową rozwiązań służących do przechowywania danych i zarządzania dostępem do nich. Ta kategoria wydatków obejmuje także środki inwestowane w taki sprzęt jak serwery czy moc obliczeniowa. Poniżej przedstawiamy szczegółową charakterystykę tych wydatków.

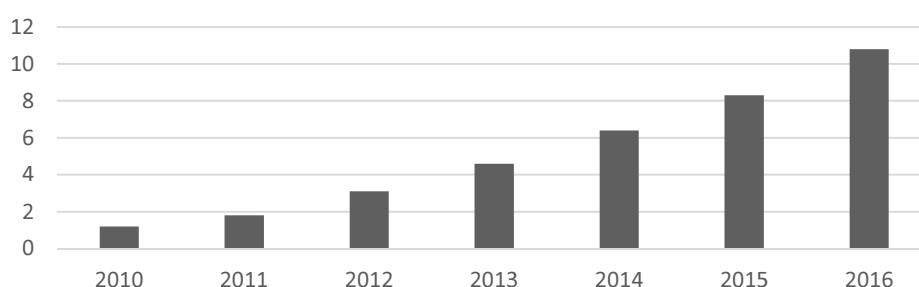
Tabela 6. Charakterystyka wydatków *Big Data* w kategorii Infrastruktura

1	Wydatki zewnętrzne związane z zakupem przestrzeni do przechowywania danych – tak wydatki na przestrzeń w chmurze, jak i zakup dysków twardej przez dużych dostawców usług chmurowych. Kategoria ta zawiera także wydatki na oprogramowanie wspierające przechowywanie czy ochronę danych.
2	Zakup serwerów wewnętrznych i oprogramowania wspierającego ich funkcjonowanie.
3	Wydatki na komponenty pozwalające na utrzymanie sieci.
4	Wydatki na infrastrukturę chmurową pozwalającą na łączenie serwerów z usługami sieciowymi dostarczanymi w ramach usług chmury publicznej.
<i>Źródło: IDC (2012: 4)</i>	

Jak pokazano, w roku 2012 firmy wydały 3,1 miliarda dolarów na zabezpieczenia infrastruktury *stricte* do adekwatnego przechowywania danych. Oszacowano również, iż wydatki te wzrosną do 10,8 miliarda dolarów w roku 2016 – osiągając średni roczny wskaźnik wzrostu na poziomie 43% w tym okresie. Tempo tego wzrostu jasno wskazuje na dynamicznie rosnące wśród firm zainteresowanie tworzeniem tego typu infrastruktury organizującej wszelakie dane cyfrowe gromadzone przez przedsiębiorstwa. Należy jednak pamiętać, iż inwestycje tego typu są jedynie początkowym etapem procesu tłumaczenia danych na wartość rynkową.

Wykres 17.

Wzrost wydatków na infrastrukturę Big Data



Źródło: IDC (2012: 5-6)

O ile wydatki z kategorii infrastruktura są relatywnie klarowne i zrozumiałe, wydatki z kategorii oprogramowanie są nieco bardziej złożone. Zawierają one bowiem wydatki na zakup i tworzenie algorytmów wspierających organizację danych i zarządzanie nimi, oprogramowanie służące eksploracji i analizie danych oraz aplikacje specyficzne dla danego sektora czy branży – patrz poniżej.

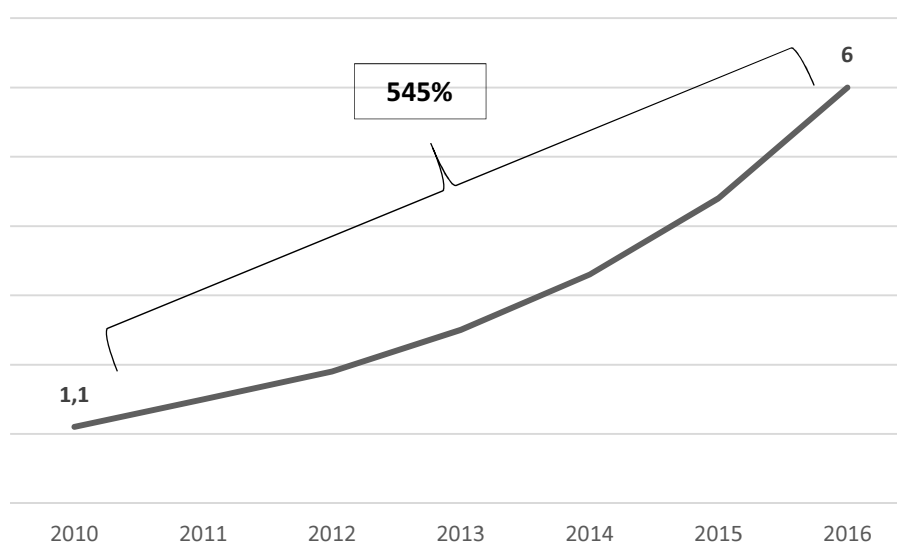
Tabela 7. Charakterystyka wydatków *Big Data* w kategorii Oprogramowanie

1	Oprogramowanie służące organizacji i zarządzaniu danymi – włączając w to wydatki niezbędne do zarządzania równoległego bazami danych, systemy zarządzania treścią, bazy danych XML, obiektowe bazy danych, programy pozwalające na integrację danych <i>etc.</i>
2	Oprogramowanie pozwalające na eksplorację i analizę danych oraz programy służące do wizualizacji danych.
3	Aplikacje pozwalające na zarządzanie procesami biznesowymi, a także aplikacje specyficzne dla poszczególnych branż.
Źródło: IDC (2012: 4)	

W roku 2012 firmy wydały na powyższe cele ok. 1,9 miliarda dolarów i oszacowano, że wartość ta osiągnie 6 miliardów dolarów w roku 2016.

Wykres 18.

Wzrost wydatków na oprogramowanie związane z *Big Data*



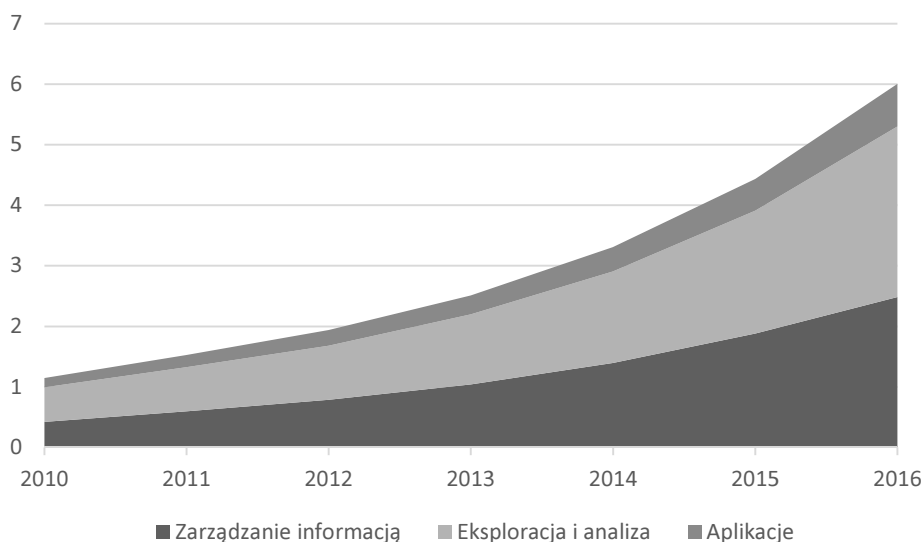
Źródło: IDC (2012: 6)

Należy powiedzieć, iż dynamika wzrostowa w tej kategorii jest na wysokim poziomie. Choć w latach 2010-2016 wydatki te nie rosły tak szybko jak wydatki z kategorii „infrastruktura”, dane zaktualizowane w roku 2015 wskazują, iż będą one rosnąć najszybciej latach 2016-2019 (IDC 2015). Dane te potwierdzają obraz rozwoju praktyk zautomatyzowanej analityki danych nakreślony we wcześniejszych częściach tego opracowania – w pierwszym okresie rozwoju *Big Data* firmy o wiele bardziej skupiały się na rozwoju infrastruktury, a dopiero teraz wkraczają w okres, gdzie na znaczeniu zyskują siły eksploracyjne i analityczne, które wiążą się bezpośrednio z technikami i technologiami *TDM*.

Dane pozwalają poznać subkategorie wydatków – i tak w roku 2012 firmy wydały 786,7 miliona dolarów na oprogramowanie do zarządzania informacjami, które wiąże się z tworzeniem i utrzymaniem infrastruktury danych. Firmy wydały 895,2 miliona dolarów na oprogramowanie do eksploracji i analizy danych, a 254,7 miliona na konkretne aplikacje charakterystyczne dla danych branż.

Wykres 19.

Wydatki na oprogramowanie dot. Big Data wg poszczególnych subkategorii w mld USD

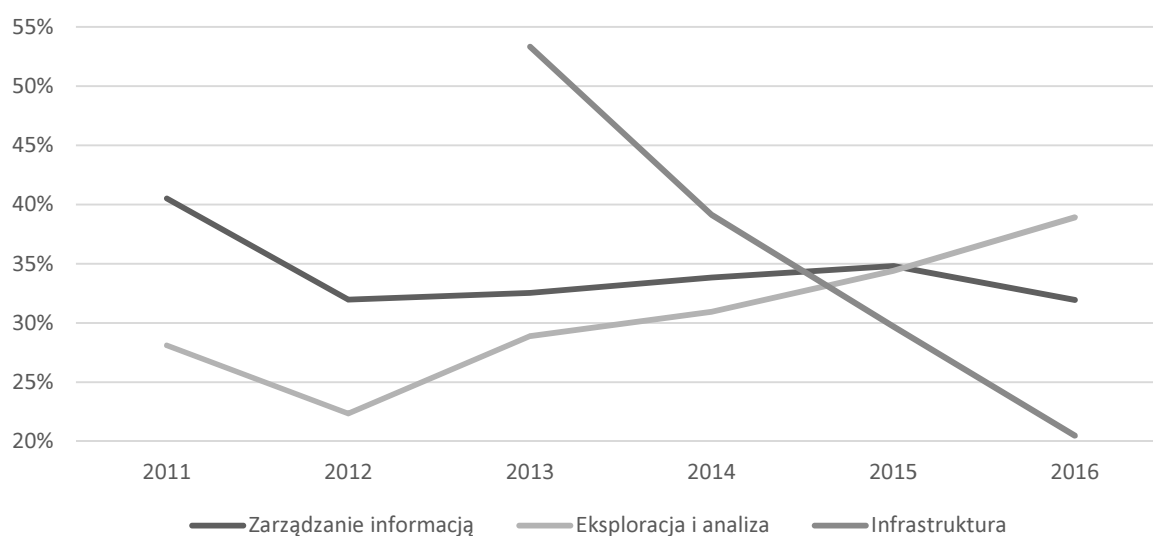


Źródło: IDC (2012: 6)

Zestawienie rocznych stóp wzrostu dla dwóch kategorii wydatków, z których jedna dotyczy wspierania infrastruktury danych, a druga eksploracji i analizy, pokazuje rosnące znaczenie wydatków na *TDM*. Stopa wzrostu tej drugiej kategorii rośnie szybciej – w ostatnim roku analizy stopa wzrostu dla pierwszej kategorii zmalała. Drastycznie również maleje stopa wzrostu dla ogólnej kategorii „infrastruktura”. Fakty te sugerują, że być może powoli zaczynamy mieć do czynienia z nasyceniem rynku *Big Data* w zakresie tworzenia infrastruktury danych. Potwierdza to deklaracje liderów biznesowych, iż rośnie ich zainteresowanie talentami, umiejętnościami i rozwiązaniami pozwalającymi na tłumaczenie danych na wartość biznesową.

Wykres 20.

Stopy wzrostu wydatków na oprogramowanie TDM vs. oprogramowanie dot. utrzymania infrastruktury danych vs. „infrastruktura”



Źródło: IDC (2012)

Dalsza analiza danych finansowych pozwala się domyślić, jak duże kwoty wydawane są na techniki i technologie TDM w każdej z subkategorii wydatków na oprogramowanie. 2,5 miliarda dolarów (w roku 2012) na zarządzanie informacją (nie jest to wydatek TDM), 2,8 miliarda na eksplorację i analitykę (w całości wydatek TDM) oraz 0,7 miliarda dolarów w kategorii „aplikacje” (z czego 0,3 mld można uznać za wydatek TDM³).

Tabela 8. Wydatki na TDM w kategorii „oprogramowanie” (mld USD)

Subkategoria wydatków	Kwota	Z tego na TDM
Zarządzanie informacją	2,5	0
Eksploracja i analityka	2,8	2,8
Aplikacje	0,7	0,3
Suma	6	3,1

Źródło: IDC 2012

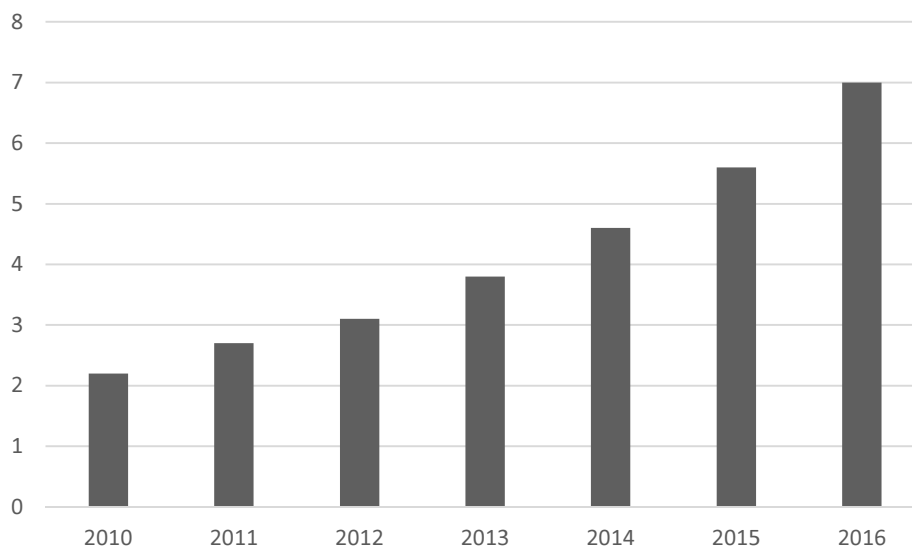
Ostatnią kategorią wydatków w ramach rynku *Big Data* badaną przez firmę IDC jest kategoria nazwana „usługi”. Zawiera ona wydatki na „konsulting biznesowy, outsourcing niektórych procesów, usługi związane z bezpieczeństwem, wsparcie w zakresie wykorzystania oprogramowania i sprzętu oraz szkolenia kadr dotyczące wdrażania i rozwijania projektów *Big Data*” (IDC 2012). Warto dodać, iż jest

³ Dane gromadzone przez IDC pozwalają jedynie na szacunki względem wydatków tej ostatniej subkategorii. Obejmuje ona bowiem zarówno aplikacje analityczne specyficzne dla danych branż, jak i aplikacje do zarządzania procesami biznesowymi. Przyjmując strategię konserwatywnego podejścia, należy ocenić, iż ok. 60% wydatków dotyczy zarządzania procesami biznesowymi niezwiązanymi bezpośrednio z technikami i technologiami TDM.

to znacząca kategoria wydatków, która osiągnęła 3,1 miliarda dolarów w roku 2012 i stanowi 39% rynku. Szacuje się, że wydatki z tej kategorii wzrosły do 7 miliardów dolarów w roku 2016 przy średniej rocznej stopie wzrostu 21,1%.

Wykres 21.

Wydatki na usługi w ramach rynku Big Data



Źródło: IDC (2012: 6)

Sposób, w jaki gromadzi się informacje na temat tych wydatków, uniemożliwia zrozumienie wprost, jaka ich część odnosi się do budowania i utrzymywania infrastruktury danych, a jaka do celów związanych z ich eksploracją i analizą. Bazując na poprzednich estymacjach, uznajemy, iż – przyjmując założenia konserwatywne – nie jest irracjonalnym założeniem, że 15% z tej kwoty dotyczy serwisowania wszelkich zakupów związanych z *TDM*. Dodatkowo – odwołując się do coraz silniejszej presji na szkolenie kadr i wdrażanie kultury organizacyjnej przyjaznej analityce danych – szacujemy, iż ok. 2 miliardy dolarów z kategorii „usługi” są wydatkowane na szkolenia powiązane z *TDM*.

Szacunki przedstawione jak dotąd pozwalają wyciągnąć wniosek, iż w roku 2016 rynek *Big Data* był wart 23,8 miliarda dolarów, z czego 6,2 miliarda dolarów wydano na technologie, usługi i inne produkty dotyczące eksploracji i analizy danych. Dalszych 10,8 miliarda dolarów wydano na stworzenie odpowiedniej infrastruktury danych, 2,5 miliarda na oprogramowanie służące zarządzaniu informacją, a 4,3 miliarda dolarów na usługi innego rodzaju. Jednak trzeba podkreślić, iż szacunki te bazują na danych i estymacjach firmy IDC poczynionych w roku 2012, ponieważ jedynie dane z tego roku pozwalają na określenie, które z wydatków na rynku *Big Data* odwołują się do tworzenia infrastruktury, a które są partykularne dla *TDM*. Niemniej w roku 2015 IDC zaktualizowało swoje szacunki dotyczące całego rynku, raportując, iż w 2016 r. wartość globalnego rynku wyniesie 26,1 miliarda dolarów – czyli o 2,3 miliarda dolarów więcej, niż szacowano w roku 2012. Stąd wykorzystując wiedzę o proporcjach wydatków z roku 2012 oraz najbardziej aktualne informacje na temat wielkości rynku, można powiedzieć, iż wartość rynkowa *TDM* wyniosła w roku 2016 ok. 7 miliardów dolarów i stanowiła 27% całości rynku *Big Data*.

Warto również nadmienić, iż między rokiem 2014 a 2019 wydatki z kategorii „oprogramowanie” będą rosły najszybciej – w tempie 26,2% średniej rocznej stopy wzrostu, w porównaniu ze stopą 22,7% dla

kategori „usługi” i 21,7% dla kategorii „infrastruktura”. Informacje te pozwalają szacować, iż w roku 2019 *TDM* będzie stanowił 35% całego rynku *Big Data* i osiągnie wartość 17 miliardów dolarów.

Dane europejskie pozwalają przy tym stwierdzić, iż w roku 2016 europejski rynek *Big Data* osiągnął wartość 9,37 miliarda dolarów i urosnie średnio o 25,7%, osiągając wartość 29,4 miliarda dolarów w roku 2021 (Research and Markets, 2016). W tym układzie europejski *TDM* był wart ok. 2,5 miliarda dolarów w roku 2016 i będzie rósł przynajmniej do 10,3 miliarda w roku 2021.

3.2 Szerszy wpływ danych na gospodarkę

Warto pamiętać, iż kalkulacje przedstawione powyżej prezentują wąskie spojrzenie na wartość finansową technologii *Big Data* i *TDM*. Innymi słowy jest to wpływ bezpośredni, który należy rozumieć jako „wpływ generowany bezpośrednio przez branżę *Big Data*; jest to aktywność wygenerowana przez firmy aktywne w tym sektorze. W tym ujęciu wpływ bezpośredni jest mierzony wpływami finansowymi z tytułu sprzedaży produktów i usług *Big Data*” (IDC 2015b: 145). Jednak warto pamiętać, że rynki *Big Data* to w większości rynki typu B2B. Oznacza to, że ich istnienie jest warunkowane potrzebami innych firm i popytem generowanym przez nie – a nie przez konsumentów końcowych. W związku z tym firmy będą zgłaszać zapotrzebowanie o tyle, o ile skorzystanie z tych produktów przełoży się w jakimś sensie na poziom ich konkurencyjności.

Stąd analizując wpływ tych technologii, należy także wziąć pod uwagę efekty pośrednie wprost (*forward indirect impacts*), które będą zawierały wzrost gospodarczy spowodowany zastosowaniem zautomatyzowanej analityki danych przez firmy wytwarzające produkty dla klienta końcowego. Dla tych firm bowiem rozwiązania danowo-analityczne stają się znaczącym czynnikiem produkcji dóbr i usług, który pozwala osiągać różnego rodzaju przewagi konkurencyjne i zyski społeczne związane z większą produktywnością i innowacyjnością (*Ibidem*). Można również wziąć pod uwagę tzw. efekty pośrednie odwrócone (*reverse indirect impacts*) – czyli zyski generowane przez dostawców towarów i usług dla firm z branży *Big Data*.

Ostatnią ważną kategorią efektów ekonomicznych są tzw. efekty indukowane. Efekty te zawierają aktywność gospodarczą wygenerowaną przez pensje dla pracowników w branży *Big Data*. Część tej aktywności będzie odnosić się do dóbr i usług konsumenckich, co prowadzi do wzrostu gospodarczego – szczególnie w takich branżach, jak handel detaliczny, bankowość czy rozrywka.

Cytowana tu wielokrotnie firma IDC zaprezentowała w roku 2013 model, według którego całościowy wpływ gospodarczy danych w ogóle został obliczony na 321,1 miliarda dolarów, co stanowiło 1,8% europejskiego PKB w roku 2013. W 2013 wartość europejskiego rynku danych (czyli rynku wszelkich danych – nie tylko tych o charakterystyce *Big*) wyliczono na 61,6 miliarda dolarów. Kalkulacje te oznaczają, iż 1 dolar inwestycji firm w produkty i usługi związane z wszelkimi danymi i ich analityką dawał zwrot gospodarczy na poziomie 5,2 dolara.

Jednak dane o charakterystyce *Big* mają o wiele większy potencjał. Zgodnie z modelem opracowanym przez Buchholtz (*et al.*) szacuje się, iż fenomen *Big Data* dla europejskiej gospodarki dodatkowych 200 miliardów euro w roku 2020. Na poprzednich stronach prezentowaliśmy z kolei szacunki, iż europejski rynek *Big Data* będzie wart ok. 18,7 miliarda euro w roku 2020. Oznacza to, iż 1 euro zainwestowane w produkty i usługi *Big Data* przyniesie średnio 10,7 euro zwrotu dla gospodarki (biorąc pod uwagę wszystkie kategorie efektów). Wyliczenia te pokazują, iż wpływ ekonomiczny procesów związanych z

Big Data (w tym *TDM*) jest ponad dwukrotnie większy niż wpływ bardziej tradycyjnych danych. Używając tych szacunków, można stwierdzić, że w roku 2016 *TDM* w formie produktów i usług był wart 2,5 miliarda dolarów na rynkach europejskich, jednak jego wpływ na całość gospodarki wyniósł do 26,7 miliarda dolarów. Wpływ ten może sięgnąć 110,1 miliarda dolarów w roku 2020.

Bibliografia

- Bean, Charles. 2016. *Independent Review of UK Economic Statistics*.
- Barel, Yannick. *Big Data. Big Opportunity*, IBM: Power Systems.
- Bradshaw, David, Gabriella Cattaneo, Rosanna Lifonti, John Simcox. 2014. *Uptake of Cloud in Europe. Follow-up of IDC Study on Quantitative estimates of the demand for Cloud Computing in Europe and the likely barriers to take-up*, Framingham: IDC Analyze the Future.
- Brown, Brad, Jacques Bughin, Angela Hung Byers, Michael Chui, Richard Dobbs, James Manyika, Charles Roxburgh. 2011. *Big data: The next frontier for innovation, competition and productivity*, McKinsey Global Institute.
- Brynjolfsson, Erik, Hitt Lorin M. and Kim Heekyung Hellen (2011), *How Does Data-Driven Decision-making Affect Firm Performance?*
- Buchholtz, Sonia, Maciej Bukowski, Aleksander Śniegocki. 2014. *Big and open data in Europe. A growth engine or a missed opportunity?*, Warszawa: demosEUROPA, Warsaw Institute for Economic Studies.
- Buytendijk, Frank. 2014. *Hype Cycle for Big Data*, Stamford: Gartner.
- Capgemini and Informatica. 2016. *The Big Data Payoff: Turning Big Data into Business Value*.
- Cattaneo, Gabriella, Mike Glennon et al. 2015. *European Data Market SMART 2013/0063, D6 – First Interim Report*, Framingham: IDC Analyze the Future.
- Cebr. 2012. *Data Equity. Unlocking the value of big data*, London: Centre for Economics and Business Research Ltd.
- Danson, Forrest et al. (red.). 2016. *Analytics Trends 2016. The Next Evolution*, New York: Deloitte.
- EMC Corporation. 2014. *Redefine. EMC Big Data: Become Data-Driven*, Warszawa: EMC Forum 2014.
- Euro Federico (2009), *The Economic Impact of Cloud Computing on Business Creation, Employment and Output in Europe: An Application of the Endogenous Market Structures Approach to a GPT Innovation*.
- European Commission, *Digital Market Factsheet (2015)*.
- European Parliament. 2016. *Towards a thriving data-driven economy*, Strasburg.
- EY Building a better working world, Europe Unlimited, IDC Analyze the Future, tech4i2. 2013. *Business Opportunities: Big Data*.
- Fayyad, Usma, Gregory Piatetsky-Shapiro and Padhraic Smyth. 1994. *From Data Mining to Knowledge Discovery in Databases*, „AI Magazine”, nr 17/3: 37-54.
- Forbes Insights. 2015. *Betting on Big Data. How the right culture, strategy and investments can help you leapfrog the competition*. Washington.
- Fujitsu. 2014. *Fujitsu Market Insights. Big Data*, Fujitsu Technology Solutions GmbH.
- Georghiou, Luke. 2015. *Value of Research. Policy Paper by the Research, Innovation and Science Policy Experts (RISE)*, Luxembourg: Publications Office of the European Union.
- Girard, Greg. *From Big Data to Real-Time Business Value*, Framingham: IDC Analyze the Future.

Heudecker, Nick, Lisa Kart. 2014. *Survey Analysis: Big Data Investment Grows but Deployments Remain Scarce in 2014*, Stamford: Gartner.

Heudecker, Nick, Merv Adrian. 2015. *Survey Analysis: Hadoop Adoption Drivers and Challenges*, Stamford: Gartner.

Heudecker, Nick, Lisa Kart. 2015. *Survey Analysis: Practical Challenges Mount as Big Data Moves to Mainstream*, Stamford: Gartner.

Hexagon Mining, Level 3 Communications and Liquid Telecom, Maptek, Reflex SAP. 2015. *Mining Big Data*, London: Aspermont Media.

Intel IT Center. 2012. *Peer Research. Big Data Analytics. Intel's IT Manager Survey on How Organizations Are Using Big Data*, United States.

Kart, Lisa, Nick Heudecker, Frank Buytendijk. 2013. *Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype*, Stamford: Gartner.

Kart, Lisa. 2015. *Big Data Industry Insights*, Stamford: Gartner.

Kelly, Jeff. 2015. *Big Data Vendor Revenue and Market Forecast. 2011-2026*, Wikibon.

McKinsey (2011), *Are you ready for the era of big data?* Retrieved from www on 26th August 2016, <http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data#0>.

Pietruszyński, Piotr. 2014. *Big Data+. Systemy analityki wielkich zbiorów danych w polskich organizacjach*. Computerworld.

Ransbotham, Sam, David Kiron, Pamela Kirk Prentice. 2015. *The talent dividend. Analytics talent is driving competitive advantage at data-oriented companies*. MIT Sloan Management Review.

SNS Research Market Intelligence & Consultancy Solutions. *The Big Data Market. Opportunities, Challenges, Strategies, Industry Verticals & Forecasts*.

SVB Analytics. 2015. *Big Data Next: Capturing the Promise of Big Data*, United States: Silicon Valley Bank.

Vesset, Dan *et al.* 2012. *Market Analysis. Worldwide Big Data Technology and Services 2012–2015 Forecast*, Framingham: IDC Analyze the Future.

Vesset, Dan *et al.* 2015. *IDC FutureScape: Worldwide Big Data and Analytics 2016 Predictions*, Framingham: IDC Analyze the Future.

Vasset, Dan, Suya Xiong. 2015. *IDC MaturityScape Benchmark: Big Data and Analytics in the United States*, Framingham: IDC Analyze the Future.

Watson Healy Leigh. 2016, *Information Industry Outlook 2017: All Data, Nothing But Data*. Outsell.

Zwakman, Greg, Matt Aslett, Jason Stamper, Krishna Roy. 2015. *Total Data market expected to hit \$115bn by 2019*. 451 Research LLC.