

BARIERY WDROŻEŃ ROZWIĄZAŃ BIG DATA

Jędrzej WIECZORKOWSKI, Magdalena JURCZYK-BUNKOWSKA

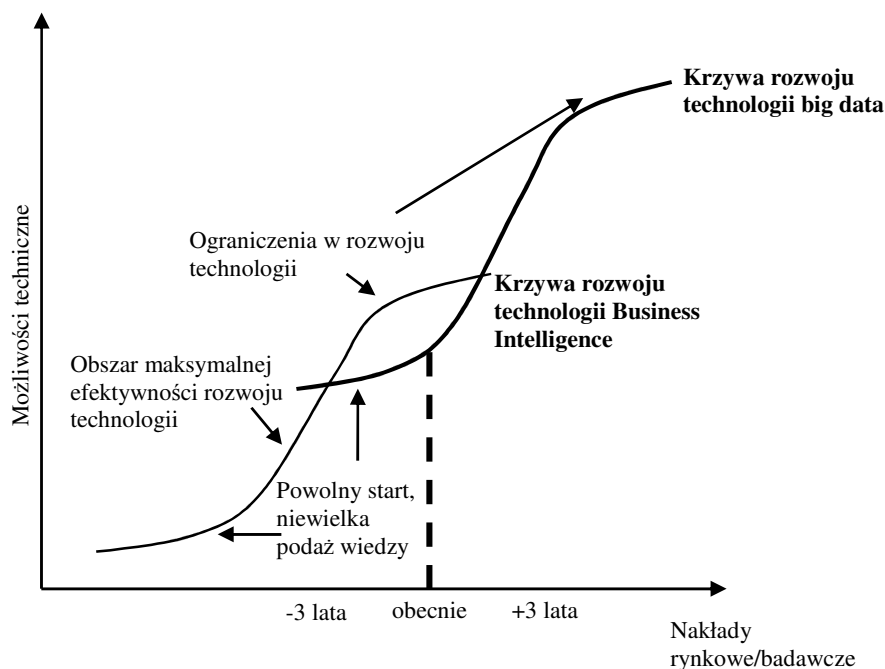
Streszczenie: Rozwój technologii big data będzie w przeciągu najbliższych lat determinował znaczące zmiany w funkcjonowaniu przedsiębiorstw produkcyjnych i usługowych, zmieniając często modele ich działalności. W artykule zidentyfikowano jego etap postępującą się krzywą Fostera (w kształcie litery S) opisującą rozwój innowacyjnych technologii. Technologię big data odniesiono do Business Intelligence. Celem artykułu jest wskazanie na bariery wdrożenia big data, które występują na obecnym etapie rozwoju tej technologii. Podzielno je na sześć grup odnoszących się do: technologii przetwarzania, ludzi, potencjału organizacyjnego, ograniczeń finansowych, dostępności danych i uwarunkowań prawnych. Stanowi to bazę do dalszych badań mających na celu zaproponowanie rozwiązań wspierających rozwój technologii big data w Polsce.

Słowa kluczowe: big data, bariery wdrożeniowe, technologie informatyczne, krzywa Fostera

1. Wprowadzenie

Współczesną gospodarkę charakteryzuje szerokie wykorzystanie technologii informatycznych. Stosowane rozwiązania stopniowo ewoluują. Obok klasycznych systemów transakcyjnych wspomagających zarządzanie w zakresie bieżącej działalności coraz szerzej stosuje się rozwiązania wykorzystujące dostępne duże wolumeny danych (big data). Wspomagają one przede wszystkim podejmowanie decyzji, np. mogą być pomocne we wspomaganie produkcyjnej działalności operacyjnej, ale również decyzjach taktycznych i strategicznych. Wdrażanie każdego systemu informatycznego wymaga spełnienia odpowiednich warunków, w innej sytuacji może nie być możliwości przystąpienia do projektu, bądź projekt skazany będzie na niepowodzenie. Uwarunkowania takie wiążą się z barierami wdrożeniowymi, czyli przeszkodami i ograniczeniami, w których funkcjonuje dana organizacja. Bariery wdrożeniową zdefiniować można jako czynnik (lub wiele czynników), którego oddziaływanie może mieć negatywny wpływ na przebieg procesu wdrożeniowego, jak również stanowić zagrożenie dla zakończenia projektu sukcesem. Bariery wdrożeniowe to występujące konkretne działania, zdarzenie lub będący ich skutkiem stan faktyczny, który ma negatywny wpływ na efektywne wdrożenie systemu [1].

Istnieją uniwersalne grupy barier wdrożeń systemów informatycznych, istotne praktycznie dla wszystkich typów oprogramowania. Jako klasyczne uniwersalne bariery wdrożeniowe (lub grupy barier ze względu na ich ogólny charakter) przedstawiane w literaturze można uznać: barierę techniczną, ekonomiczną, organizacyjną, socjopsychologiczną (ludzką) [2]. Powstaje natomiast pytanie, czy ta lista bazowych barier dotyczy w równym stopniu wdrożeń rozwiązań big data? Metody big data i związane z nimi technologie są dość młode, dlatego ich wdrożenie wymaga wielu innowacji. Tezą niniejszego artykułu jest, że znaczny udział procesów innowacyjnych o wysokim poziomie nowatorstwa ma bezpośredni wpływ na rodzaj barier wdrożeniowych technologii big data, które są charakterystyczne dla wczesnych etapów rozwoju technologii zgodnie z krzywą Fostera [3] zaprezentowaną na rys. 1, gdzie odniesiono big data do Business Intelligence (BI).



Rys. 1. Rozwój technologii Business Intelligence i big data

Źródło: opracowanie własne na podstawie [3]

Z tezy tej wynika cel artykułu, którym jest zidentyfikowanie barier (lub grup barier) wdrożeń rozwiązań big data w kontekście potencjału wdrożenia tych rozwiązań i gotowości organizacji do ich przewycięzania. Posłużono się w tym celu analizą literatury oraz aktów prawnych, a także obserwacją działalności przedsiębiorstw wykorzystujących lub rozważających wykorzystanie rozwiązań big data. W poszczególnych rozdziałach przedstawiono krótki zarys koncepcji big data, scharakteryzowano potencjał różnego typu organizacji do wdrożeń takich rozwiązań w kontekście ich innowacyjności oraz zidentyfikowano i opisano grupy barier wdrożeniowych. Następnie sformułowano wnioski dotyczące specyfiki barier w projektach związanych z wdrożeniem technologii big data.

2. Główne założenia koncepcji big data

Pojęcie big data jest ściśle związane z nowymi możliwościami efektywnego przetwarzania danych masowych oraz nowymi źródłami dostępu do takich danych. W szczególności powszechna digitalizacja dokumentów, zjawisko internetu rzeczy (IoT), coraz szersze stosowanie różnych sterowników, czujników, systemów monitoringu itp., rozpowszechnienie idei otwartych danych, a przede wszystkim powszechność wykorzystania w społeczeństwie internetu oraz smartfonów i innych urządzeń mobilnych wpływają na gwałtowne zwiększenie ilości różnorodnych danych, które mogą mieć znaczenie gospodarcze. Nowe wielkie zbiory danych umożliwiają nową jakość analizy statystycznej i eksploracji danych (*data mining*). W coraz szerszym zakresie wykorzystuje się korelację w badaniu różnorodnych zjawisk, poszukuje się nieznanych dotąd zależności pomiędzy danymi.

Pojęcie big data odnosi się do zbiorów danych, które ze względu na swój rozmiar i złożoność nie mogą być zarządzane i przetwarzane przy pomocy tradycyjnych technik, a także przekraczają możliwości typowych narzędzi bazodanowych w zakresie gromadzenia, przechowywania i analizowania tych danych. Koncepcja dotyczy przede wszystkim gromadzenia i przetwarzania dużych, wcześniej niewykorzystywanych na tę skalę, wolumenów danych. Zazwyczaj też łączy się ją z trzema cechami pochodzącymi z modelu 3V: *volume* (duży wolumen danych), *velocity* (zmiennosc przetwarzanych danych) i *variety*, (różnorodność przetwarzanych danych) [4], dodając jednocześnie różne kolejne cechy wzbogacające model „kilku V”. Przykładowo Gartner Group rozumie big data jako zasoby informacyjne dużych rozmiarów, szybko zmieniające się i/lub charakteryzujące się dużą różnorodnością, które wymagają efektywnych kosztowo i innowacyjnych form przetwarzania, umożliwiając poprawę wglądu w dane, podejmowanie decyzji i automatyzację procesów [5].

Dla rozwiązań big data przede wszystkim charakterystyczne jest przetwarzanie danych nieustrukturyzowanych lub słabo ustrukturyzowanych (takich jak obrazy, wideo, dźwięk) oraz danych dynamicznych, generowanych w krótkich odstępach czasu (takich jak dane sensoryczne i strumieniowe), których analiza następuje w czasie zbliżonym do rzeczywistego. Podobnie też na big data patrzy J.Surma odnajdując w koncepcji trzy podstawowe wartości dodane związane z rozwojem takich rozwiązań i ewolucją w porównaniu do koncepcji BI: możliwość składowania i przetwarzania danych niestrukturalizowanych, dostęp do danych z zewnątrz firmy pochodzący ze źródeł ogólnie dostępnych, możliwość dostępu do dużych wolumenów danych w czasie rzeczywistym [6].

Koncepcja big data dotyczy nie tylko zmiany w samym przetwarzaniu danych, ale także fundamentalnej zmiany w podejściu do gromadzenia danych wynikającym z niskiego kosztu ich pozyskiwania. Część danych, jak dawniej, zbierana jest aktywnie - w konkretnym z góry znanym celu. Jednakże inne dane zbierane są w sposób pasywny, przy okazji zbierania pozostałych danych, podczas gdy możliwości przyszłego ich wykorzystania nie są z góry znane. Jednocześnie stale brana jest pod uwagę możliwość alternatywnego wtórnego użycia wszystkich gromadzonych danych w przyszłości w nowym celu różnym od ich pierwotnego przeznaczenia. Zjawisko to jest niekiedy określane jako danetyzacja (*datafication*) [7].

W praktyce stosowanie koncepcji big data łączy ze sobą różne obszary, stąd autorzy niniejszego artykułu wyróżniają w niej trzy podstawowe aspekty: technologiczny (infrastruktura, oprogramowanie i zaawansowane metody analityczne), ekonomiczny (inaczej biznesowy, czyli zastosowanie koncepcji) oraz społeczny (społeczne konsekwencje zastosowań) [8]. Potencjał organizacji do wdrożeń rozwiązań big data oraz bariery wdrożeniowe wiążą się przede wszystkim z aspektem ekonomicznym, przenikając się jednak także z pozostałymi wymienionymi aspektami.

3. Potencjał organizacji do wdrożeń rozwiązań big data

Rozwiązania informatyczne klasy big data są w chwili obecnej podejściem dość nowym, tuż przed obszarem maksymalnego wzrostu efektywności technologii wobec nakładów (krzywa S, rys. 1) i ich wdrożenie może być w przeważającej większości przypadków traktowane jako innowacja. Co więcej, kreują one nowe możliwości, które będą pociągały za sobą kolejne innowacje. Zgodnie z klasyczną definicją J.Schumpetera [9] innowacje niosą za sobą zawsze nowe i lepsze możliwości, choć mogą przybierać różną postać np. nowych produktów, nowych metod produkcji, źródeł dostaw, nowych rynków lub metod organizacji biznesu. Powstaje pytanie do której grupy, opierając się na najbardziej znanym podziale

innowacji wyróżniającym innowacje: produktowe, procesowe, marketingowe i organizacyjne, należy zaliczyć takie wdrożenie [10]. Czy, a jeżeli tak to na ile, będzie ono wymuszało kolejne innowacje w przedsiębiorstwie?

W większości przypadków rozwiązania big data wdrażane są w celu poprawy systemu zarządzania organizacją dzięki lepszymu dostępowi do informacji zarządczej z wykorzystaniem danych teoretycznie dotąd posiadanych, lecz nie w pełni właściwie wykorzystywanych. Możliwe jest więc powtórne użycie danych stosowanych dotychczas w innym zakresie lub danych wcale niewykorzystywanych, a zbieranych w powiązaniu z tymi pierwszymi (tzw. danych resztkowych). Dane te mogą wtedy służyć do różnorodnych analiz na potrzeby wewnętrzne. W takim przypadku należy mówić o innowacjach organizacyjnych. Szczegółowe posiadane dane masowe często dotyczą klientów, ich zachowań i transakcji handlowych. Mogą być wtedy wykorzystane do udoskonalenia metod sprzedaży i w takim przypadku wdrożenie byłoby traktowane jako innowacja marketingowa.

Z inną sytuacją można mieć do czynienia w przypadku wykorzystywania szczegółowych masowych danych zbieranych bezpośrednio w procesie produkcyjnym. Dane takie pochodzą z taśmy produkcyjnej np. z różnych czujników, termowizji, itp. i mogą w celu poprawy jakości wytwarzanych produktów zostać wykorzystane w czasie rzeczywistym, co jest typowe dla rozwiązań big data. Charakterystyczne dla big data byłoby także stosowanie danych nieustrukturyzowanych, jak obraz wideo z kamer przemysłowych. Dane mogą być łączone z danymi z systemów informatycznych np. klasy ERP, MES, SCADA, CRM, PLM i in. Dzięki takim rozwiązaniom wykorzystującym zestawy danych, których rozmiar przekracza zdolności typowych narzędzi analitycznych, udoskonala się proces produkcyjny. Taka innowacja zaliczona zostałaby do grupy procesowej.

Metody big data mogą być także wykorzystywane w celu opracowania nowych produktów i być wtedy związane z innowacjami produktowymi. Z taką sytuacją mamy do czynienia w przypadku zbliżonym poprzedniego, lecz gdy wynikiem nie jest udoskonalenie procesu, a przykładowo opracowanie nowego procesu na potrzeby wprowadzenia nowego produktu. Charakterystyczne dla big data będą także nowe produkty informacyjne polegające na sprzedaży zebranych danych masowych lub świadczenia nowych usług (informacyjnych, doradczych) wykorzystujących te dane. Możliwe jest również świadczenie nowych usług wykorzystujących pozyskane dane zewnętrzne - zakupione lub pobrane nieodpłatnie np. z wykorzystaniem danych otwartych (*open data*), w tym dotyczących sektora publicznego (*open government data*). W omawianym przypadku także można mówić o innowacjach produktowych.

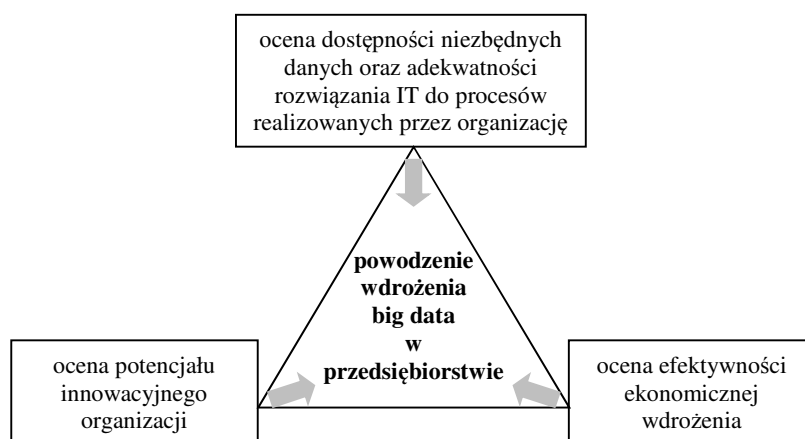
Jak powyżej wykazano rozwiązania big data mogą być podstawą do wdrożenia innowacji wszystkich czterech podstawowych typów. Ze względu na nowatorstwo koncepcji big data wdrażane rozwiązania stosunkowo często mogą przybierać charakter innowacji radykalnych. Takie innowacje wymagają zupełnie nowej wiedzy, często zaawansowanej technologii, w odróżnieniu od innowacji przyrostowych bazujących zazwyczaj na wiedzy istniejącej i względnie łatwo dostępnej [11]. Innowacje o wyższym poziomie nowatorstwa, w szczególności radykalne, niosą za sobą wyższy potencjał poprawy odnoszonej do zmiany jakości lub kosztu.

Kluczowym czynnikiem powodzenia wdrożenia rozwiązania big data jest uzyskanie dostępu do niezbędnych danych – wewnętrznych lub zewnętrznych. Jednakże innym podstawowym czynnikiem jest opracowanie nowego lub modyfikacja dotychczasowego modelu biznesu, ewentualnie pomysł na udoskonalenie funkcjonujących w organizacji procesów, w szczególności zarządczych i produkcyjnych. Powyższy czynnik jest kluczowy dla oceny potencjału wdrożeń innowacyjnych rozwiązań, w praktyce nie tylko na potrzeby

rozwiązań big data. Tym niemniej, ze względu na nowatorstwo tych metod oraz ich zastosowań biznesowych, w przypadku rozwiązań big data jest on szczególnie istotny. Wiąże się to także z oceną zdolności organizacji do realizacji projektów innowacyjnych, w szczególności w obszarze big data. Obok umiejętności realizacji projektów niezbędna jest umiejętność późniejszego przyswojenia ich produktów, co wiąże się z koniecznością modyfikacji dotychczasowych procesów, w szczególności zarządczych lub produkcyjnych. Niezależnie od powyższych, podobnie jak w każdym projekcie, elementem oceny prawdopodobieństwa powodzenia wdrożenia rozwiązania jest ocena jego efektywności ekonomicznej. Podejmowanie ryzyka jest wpisane w zarządzanie procesami innowacji, tym niemniej niezbędne jest dokonywanie oceny powodzenia każdego innowacyjnego projektu.

Stąd autorzy uważają, że ocena powodzenia wdrożenia rozwiązania big data, a co za tym idzie potencjału wdrożenia rozwiązania, składa się, podobnie jak innych innowacyjnych projektów informatycznych, z następujących elementów (rys.2) [12]:

- oceny dostępności niezbędnych danych oraz adekwatności rozwiązania do procesów realizowanych przez organizację,
- oceny efektywności ekonomicznej wdrożenia,
- oceny potencjału innowacyjnego organizacji.



Rys. 2. Ocena potencjału wdrożenia rozwiązania big data
źródło: opracowanie własne na podstawie [12]

Biorąc pod uwagę powyższe elementy należy stwierdzić, że potencjał wdrożenia rozwiązań big data obejmuje potencjał projektu i jego produktów oraz potencjał organizacji. Ten drugi z kolei obejmuje potencjał organizacji na etapie wdrażania rozwiązania (zdolność do przeprowadzenia projektu oraz wprowadzenia zmian niezbędnych do wdrożenia jego produktów) oraz na etapie późniejszego jego utrzymania (zdolność do utrzymania produktów projektu oraz umiejętność ich wykorzystania). Można więc mówić o gotowości do przyswajania takich innowacji, zaś niewystarczający potencjał równoważny jest wystąpieniu bariery wdrożeń rozwiązań big data.

4. Uniwersalne bariery wdrożeniowe w przypadku rozwiązań big data

Koncepcja big data mogła wejść w życie dzięki postępowi technicznemu i spadkowi cen sprzętu niezbędnego do przetwarzania wielkich zbiorów danych. W konsekwencji bezkosztowość lub bardzo niski koszt jednostkowego przetwarzania można traktować jako podstawowy paradygmat big data. Jednakże wymagania infrastrukturalne w przypadku metod big data są znacząco wyższe niż dla typowych zastosowań transakcyjnych. I **barierę technologiczną** należy traktować jako pierwszą z barier wdrożenia rozwiązań big data.

Konieczne jest rozwiązanie takich typowych dla big data problemów jak: zapewnienie odpowiedniej wydajności przetwarzania danych, zintegrowanie oprogramowania z dotychczasowymi systemami transakcyjnymi i BI; analiza strumieni danych w czasie rzeczywistym; opracowanie narzędzi zapewniających kompleksową wizualizację danych. W zakresie technologii stosuje się więc przede wszystkim rozproszone systemy plików (*Distributed File Systems* - DFS) i przetwarzanie równoległe (*Massively Parallel Processing* - MPP). Podstawą są bazy danych NoSQL, w których obsługa dużych zbiorów danych realizowana jest poprzez partycjonowanie, kosztem częściowej rezygnacji z zasady integralności lub dostępności w pełni aktualnych danych [13].

Infrastruktura i narzędzia informatyczne wymagają odpowiedniej obsługi, co związane jest z **barierą ludzką** (kadrową). Narzędzia wizualizacji mają maksymalnie uprościć interpretację danych i umożliwić wykorzystywanie oprogramowania bezpośrednio przez decydentów. Jednocześnie jednak w wielu przypadkach do danych źródłowych trzeba właściwie dobrać stosowane metody. Także interpretacja zależności wymaga odpowiedniej wiedzy. Dopiero wiedza i kompetencje pozwalają wykorzystywać potencjał big data w organizacji. Niezbędni są więc specjaliści od danych, tzw. *data scientists*, zastępujący dotychczasowych analityków danych tradycyjnie zajmujących się wyłącznie metodami analizy statystycznej. Powinni oni być doskonale obeznani z narzędziami informatycznymi, dobrze operować danymi niespójnymi i heterogenicznymi. Dodatkowo często oczekuje się od nich merytorycznej wiedzy o danych specyficznych dla obszaru, którym się zajmują, jak np. dane z mediów społecznościowych o potencjalnych klientach, dane naukowe itp. W chwili obecnej zawód ten jest bardzo deficytowy, choć wcześniejsze badania [14] świadczą, że oferta uczelni wyższych w zakresie studiów głównie magisterskich i podyplomowych jest coraz szersza. Z drugiej strony wymagania intelektualne wobec *data scientists* są bardzo wysokie, co w naturalny sposób ogranicza dostępność zawodu.

Kolejnym zagadnieniem łączonym z barierą ludzką jest konieczność zmiany podejścia mentalnego zarządzających i dostosowanie go do nowej kultury organizacyjnej, co związane jest z kolejną **barierą – organizacyjną** (procesową). Zarządzający powinni zaakceptować sytuację, w której podejmują niektóre bazujące na big data decyzje, których w pełni nie są w stanie zrozumieć. Jednocześnie można mieć do czynienia z sytuacją, w której trafniejsze decyzje będzie podejmowała osoba bez odpowiedniego doświadczenia zarządczego opierająca się na analizie zbiorów danych niż doświadczony manager. Niezbędna jest więc czasem akceptacja faktu, że wystarczająca jest wiedza na temat korelacji pomiędzy danymi, bez znajomości przyczynowości zjawiska. W innych sytuacjach, w oparciu o opracowane wcześniej reguły działania, decyzje mogą być podejmowane automatycznie bez udziału człowieka (na etapie podejmowania decyzji), co oznacza użycie tzw. „złotej pętli” decyzyjnej, gdzie wygenerowana propozycja nie trafia do menedżera, ale jest wykonywana automatycznie [6]. Pozyskiwanie danych w czasie zbliżonym do rzeczywistego oraz automatyzacja podejmowania decyzji może prowadzić do znaczącego przyspieszenia procesów zarządczych. Wdrożenie rozwiązań big data wiązać się może z problemem braku

zrozumienia tej idei i oporem ze strony kierowników różnego szczebla. Wymóg doświadczenia przesuwają się więc częściowo z decydenta na analityka, co w kulturze organizacyjnej może być nieakceptowalne. Potrzebna może więc być nowa kultura organizacyjna oparta na danych.

Uzyskanie pełnej wartości biznesowej z wdrożenia big data wymaga ponownego zdefiniowania przez przedsiębiorstwo swoich procesów i sposobu działania [15]. Zakres i poziom wykorzystania metod big data może się znacząco różnić pomiędzy organizacjami. Przykładowo można stosować metody big data na kilku coraz bardziej zaawansowanych poziomach interpretacji danych - analityki: opisowej, diagnostycznej, predykcyjnej i dyspozycyjnej. Analityka opisowa w oparciu o zgromadzone dane odpowiada na pytanie: "Co się stało?". Analityka diagnostyczna na podstawie danych archiwalnych odpowiada na pytanie "Dlaczego to się stało?". Analiza predykcyjna określa prawdopodobne scenariusze tego, „Co może się zdarzyć?”. I wreszcie analityka dyspozycyjna określa jakie działania należy podjąć odpowiadając na pytanie: "Co należy z tym zrobić?". [16]. Wszelki opór przed dostosowaniem procesów do możliwości metod big data wiąże się ze ograniczeniem wspomnianego wcześniej potencjału innowacyjnego organizacji oraz zdolności absorpcji innowacji i zalicza się do bariery organizacyjnej. Do tej grupy zaliczają się również inne problemy ograniczające opisany wcześniej potencjał organizacji do wdrożeń rozwiązań big data

Pochodną niektórych wymienionych powyżej barier wdrożeniowych jest kolejna – **bariera finansowa** (ekonomiczna). Jak powyżej zauważono, metody big data wymagają nowoczesnych rozwiązań infrastrukturalnych. Zakup odpowiedniej technologii (infrastruktury sprzętowej wraz z odpowiednimi narzędziami programistycznymi) jest kosztowny. Jednocześnie jednak odnotować należy znaczną poprawę dostępności takich rozwiązań w ostatnim czasie, co znacząco wpływa na obniżenie kosztów i w konsekwencji wzrost popularności zastosowań metod big data. Określane jest to niekiedy jako demokratyzacja analityki, czyli szerszy dostęp do narzędzi analitycznych, a w konsekwencji wykorzystywanie big data także przez mniejsze firmy i start-up'y.

Przykładowo jako alternatywę zakupu stosownej technologii, należy odnotować możliwość dostępu do niezbędnych rozwiązań w modelu chmury obliczeniowej obniżając istotnie początkowe koszty stosowania metod big data. Na barierę finansową wpływa także opisana bariera ludzka. Pozyskanie niezbędnych specjalistów także jest kosztowne. Ponadto oprogramowanie o otwartym kodzie (*open source*), które jest popularnym modelem licencjonowania narzędzi informatycznych stosowanych w rozwiązaniach big data, także często przenosi koszty z zakupu technologii na koszty osobowe. W zakresie bariery finansowej wspomnieć należy także o ewentualnych kosztach dostępu do zewnętrznych źródeł danych.

5. Specyficzne bariery wdrożeń rozwiązań big data

Podstawową potencjalną barierą jest zdaniem autorów ta, która niejako wynika z istoty pojęcia big data, czyli **bariera danych**. Oczywistym problemem może być pozyskanie odpowiednich danych dotyczących okresu adekwatnego do potrzeb. Praktyczne zastosowanie metod big data wymaga więc od samego początku odpowiedniej skali przetwarzania danych oraz często dostępu do odpowiednich danych historycznych. Jest to swoisty próg wejścia dla wykorzystania takich rozwiązań. Niektóre nowe usługi informacyjne mogą istnieć wyłącznie w dużej skali, nie są zaś prostym rozwinięciem małej skali. Z pomocą przychodzi tu gromadzenie przeróżnych własnych danych z myślą o

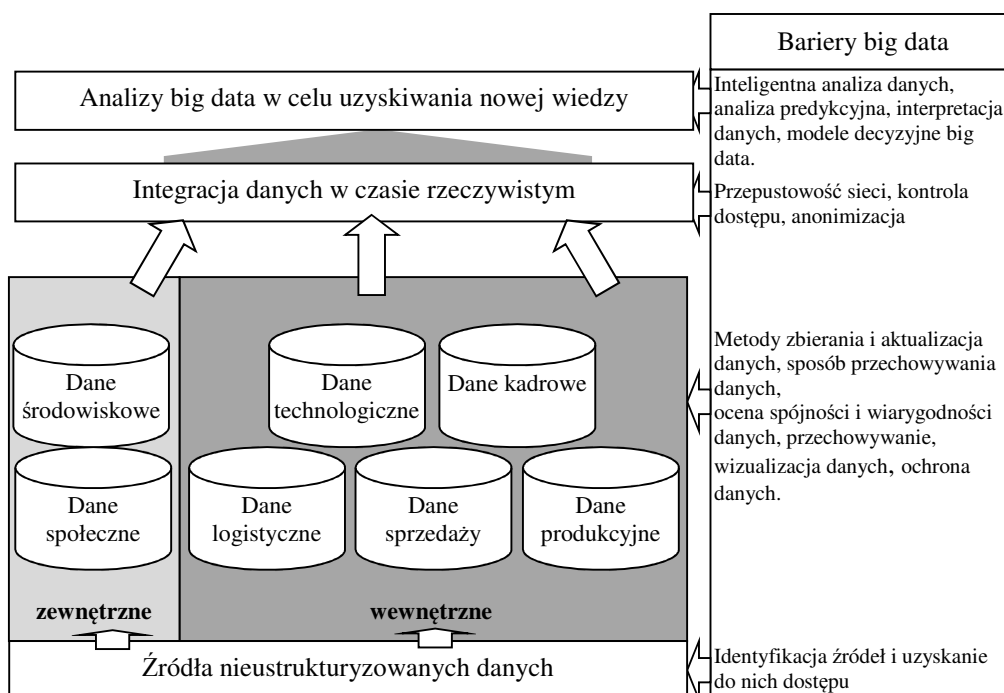
alternatywnym wykorzystaniu ich w przyszłości lub uzyskanie dostępu do danych zewnętrznych (odpłatnie lub nieodpłatnie, w tym w modelu danych otwartych).

Od danych, które mają być wykorzystane z zastosowaniem metod big data, oczekuje się nie tylko wystarczającej ich ilości, ale i ich satysfakcjonującej jakości. Te dwie cechy są jednak częściowo ze sobą sprzeczne. Polityka gromadzenia jak największej ilości danych (w konkretnym obszarze tematycznym) pochodzących z różnych źródeł niesie za sobą ryzyko pozyskiwania wśród nich danych niskiej jakości. Mogą one w jakimś stopniu być błędne, nieaktualne, wzajemnie niespójne (szczególnie gdy pochodzą z różnych źródeł). Decyzje podjęte na bazie analizy błędnych danych mogą spowodować poważne problemy funkcjonowania przedsiębiorstwa. Przy różnorodności źródeł, typową sytuacją jest różna jakość oraz niespójność danych. Problem dotyczy przede wszystkim danych pochodzących ze źródeł zewnętrznych, w szczególności ogólnodostępnych. Wtedy przede wszystkim wymaga się stosowania rozwiązań weryfikujących ich jakość. Proces porządkowania i poprawiania danych przy ich znacznej ilości może być jednak praktycznie niewykonalny, kosztowny i nieuzasadniony ekonomicznie. Idea big data co do zasady stawia ilość nad jakość, a algorytmy ich przetwarzania muszą być do pewnego stopnia odporne na błędy w danych. Można więc postawić tezę, że koncepcja big data, częściowo przeczy powszechnie wcześniej akceptowanej zasadzie, że niska jakość danych na wejściu, powoduje niską jakość na wyjściu (reguła GIGO *garbage in, garbage out*, czyli śmieci na wejściu to śmieci na wyjściu). Tym niemniej, w zależności od zastosowania, istnieje pewien minimalny akceptowalny poziom jakości danych i można przyjąć dwa podstawowe podejścia: dbanie o dobór danych wysokiej jakości ograniczając ilość danych źródłowych lub częściowe ignorowanie problemu jakości danych źródłowych, przedkładając nad nią ilość pozyskiwanych danych. Inaczej mówiąc można zwiększać złożoność procesów po stronie gromadzenia danych (np. filtrowanie) lub po stronie analizy danych. W praktyce jednak rozwiązania big data znajdują się pomiędzy tymi skrajnościami.

Kolejną, powiązaną z poprzednią, istotną barierą wdrożeń rozwiązań big data jest **bariera prawna**. Gromadzenie danych i ich przetwarzanie musi być zgodne z systemem prawnym i nie wszystko, co jest dostępne informacyjnie i technologicznie, możliwe jest do zastosowania. Pierwszym istotnym aspektem prawnym jest gromadzenie i przetwarzanie danych prywatnych, w szczególności osobowych, w tym wrażliwych, drugim zaś dostęp w celach biznesowych do cudzych, często publicznych lub ogólnodostępnych danych wraz z kwestiami praw autorskich lub licencji.

Wyjątkowo ściśle ograniczenia dotyczą możliwości przetwarzania danych osobowych, które są kluczowe dla zastosowania metod big data w wielu obszarach, w szczególności obejmujących różnorodne działania marketingowe przy działalności detalicznej. Problem jest obecnie bardzo aktualny ze względu na zmiany prawne wchodzące w życie w 2018 roku. Zgodnie z prawem jako dane osobowe traktuje się wszelkie informacje dotyczące zidentyfikowanej lub możliwej do zidentyfikowania osoby fizycznej, a osoba możliwa do zidentyfikowania to osoba, której tożsamość można ustalić bezpośrednio lub pośrednio, szczególnie przez powołanie się na numer identyfikacyjny lub jeden, bądź kilka szczególnych czynników określających jej fizyczną, fizjologiczną, umysłową, ekonomiczną, kulturową lub społeczną tożsamość. Natomiast przetwarzanie danych osobowych oznacza każdą operację lub zestaw operacji dokonywanych na danych osobowych przy pomocy środków zautomatyzowanych lub innych, jak np. gromadzenie, rejestracja, porządkowanie, przechowywanie, adaptacja lub modyfikacja, odzyskiwanie, konsultowanie, wykorzystywanie, ujawnianie poprzez transmisję, rozpowszechnianie lub udostępnianie w inny sposób, układanie lub kompilowanie, blokowanie, usuwanie lub niszczenie [17].

Specyficzną grupę danych osobowych stanowią ponadto dane wrażliwe (sensytywne), czyli dotyczące w szczególności stanu zdrowia, pochodzenia rasowego lub etnicznego, poglądów politycznych, religijnych i światopoglądowych, nałogów, życia seksualnego oraz orzeczeń w postępowaniach sądowych.



Rys. 3. Szczegółowe bariery i problemy rozwoju technologii big data
Źródło: opracowanie własne

W polskim systemie prawnym zasady te były dotąd opisane w Ustawie o ochronie danych osobowych [17]. Jednakże nowe wyzwania technologiczne wymusiły powstanie nowych regulacji dotyczących prywatności. W 2016 r. uchwalono rozporządzenie Parlamentu Europejskiego i Rady o ochronie danych (tzw. RODO - Rozporządzenie o Ochronie Danych Osobowych, GDPR - *General Data Protection Regulation*) [18], które jest bezpośrednio skutecznym aktem prawa unijnego w Polsce i w szerokim zakresie wejdzie w życie w 2018. W szczególności dokładnie zdefiniowano tam niektóre pojęcia związane ze współczesną technologią. Wprowadzono ściślejsze ograniczenia dotyczące profilowania, oznaczającego dowolną formę zautomatyzowanego przetwarzania danych osobowych, które polega na wykorzystaniu takich danych do oceny niektórych czynników osobowych osoby fizycznej, w szczególności do analizy lub prognozy aspektów dotyczących efektów pracy tej osoby, jej sytuacji ekonomicznej, zdrowia, osobistych preferencji, zainteresowań, wiarygodności, zachowania, lokalizacji lub przemieszczania się. Jednocześnie określono wymogi dotyczące pseudonimizacji danych oznaczającej przetwarzanie danych osobowych w taki sposób, by nie można ich było już przypisać konkretnej osobie, której dane dotyczą, bez użycia dodatkowych informacji, pod warunkiem że takie dodatkowe informacje są przechowywane osobno i są objęte środkami technicznymi i organizacyjnymi uniemożliwiającymi ich

przypisanie zidentyfikowanej lub możliwej do zidentyfikowania osobie fizycznej [18]. Istotną barierą stały się prawne wymagania technologiczne i organizacyjne niosące za sobą wysokie koszty dla podmiotów gromadzących i przetwarzających dane osobowe.

Drugi wymieniony aspekt bariery prawnej dotyczący zastosowań rozwiązań big data to dostęp do danych zewnętrznych, który może być realizowany dzięki danym otwartym, w szczególności pochodzącym z sektora publicznego. Koncepcja otwartych danych opisuje zbiory danych, które są bez ograniczeń dostępne i mogą być używane, jak również przetwarzane i publikowane przez każdego bez restrykcji dotyczących praw autorskich lub patentów, z zastrzeżeniem co najwyżej do wymogu podania źródła, czy umożliwienia dalszej dystrybucji przetworzonych treści na takich samych zasadach [19]. Natomiast idea otwartości danych sektora publicznego wiąże się z dyskusją na ile informacje gromadzone przez instytucje publiczne powinny być powszechnie dostępne, czy i w jakim stopniu prawo powinno zezwalać na ponowne wykorzystywanie tego, co powstało za publiczne środki. Celem inicjatywy jest w szczególności udostępnienie wszystkich danych nienaruszających bezpieczeństwa oraz praw do ochrony danych osobowych i tajemnicy handlowej posiadanych przez organizacje rządowe i samorządowe, co m.in. ma pobudzić innowacyjność [20]. W Polsce do opisywanej koncepcji odnoszą się w szczególności dwie ustawy: Ustawa o dostępie do informacji publicznej [21] oraz Ustawa o ponownym wykorzystywaniu informacji sektora publicznego [22].

Pierwsza z ustaw dotyczy prawa zaliczanego do kategorii praw wolnościowych i daje uprawnienie do wiedzy o działaniach władz publicznych (transparentność władzy), natomiast druga ma charakter prawa gospodarczego dającego prawo do tworzenia wartości dodanej w oparciu o informacje gromadzone przez sektor publiczny [23]. Zgodnie z nią „ponowne wykorzystywanie” oznacza wykorzystywanie przez osoby fizyczne lub prawne dokumentów będących w posiadaniu organów sektora publicznego, do celów komercyjnych lub niekomercyjnych innych niż ich pierwotne przeznaczenie w ramach zadań publicznych, dla których te dokumenty zostały wyprodukowane [22]. W praktyce dane zawarte w centralnym repozytorium informacji publicznej (<http://danepubliczne.gov.pl/>) [24] mogą być podstawą dla modelu biznesowego opartego na koncepcji big data, lecz tylko w ramach ograniczeń powyższego prawa.

Wybrane szczegółowe bariery i problemy związane z wdrażaniem rozwiązań big data przedstawiono na rysunku 3.

Pytaniem pozostaje jak powyższe grupy barier odnoszą się do konkretnych szczegółowych barier i problemów artykułowanych przez przedsiębiorców. Kilka lat temu badania na ponad 300 podmiotach zainteresowanych analityką big data wykonano w TDWI Research [25]. Badano w praktyce szczegółowe bariery z grup technologicznej, ludzkiej i finansowej. Najczęściej zgłaszane konkretne problemy dotyczyły bariery ludzkiej (głównie nieodpowiedni personel i brak umiejętności) oraz finansowej (wysokie koszty), w dalszej kolejności technologicznej (problemy architektoniczne i narzędziowe). Należy jednak zauważyć, że nie pytano przede wszystkim o problemy z dostępnością danych, wychodząc z założenia, że ankietowane podmioty mają już ten problem rozwiązany. Choć badanie było przeprowadzone jakiś czas temu, można przypuszczać, że obecnie problemy byłyby podobne, gdyż zastosowania big data intensywnie się rozwijają i, mimo szkolenia nowych specjalistów, problem kadrowy pozostaje niezmiennie bardzo istotny.

6. Podsumowanie i kierunki dalszych badań

Analityka big data jest ciągle jeszcze w początkowej fazie rozwoju. Zaprezentowane w niniejszym artykule grupy barier wdrożeń rozwiązań big data w dużym stopniu pokrywają się z ogólnymi uniwersalnymi barierami wdrożeniowymi systemów informatycznych: technologiczną, ludzką, organizacyjną, finansową. Dwie dodatkowo zaproponowane specyficzne bariery to: bariera danych (szczególnie ważna ze względu na istotę metod big data opierających się na wielkich wolumenach często historycznych danych, pochodzących z zewnętrznych źródeł), a także bariera prawna (istotna ze względu na wagę przetwarzania danych osobowych oraz zewnętrznych). Ta ostatnia jest niezmiernie istotna ze względu na społeczny wymiar analizy danych. Technologie big data nie mogą traktować w sposób instrumentalny człowieka z jego wiedzą i prawem do prywatności. Oczywiście bariery analogiczne do obu powyższych barier specyficznych występują także przy innych wdrożeniach w obszarze IT, tym niemniej to w przypadku big data odgrywają szczególną rolę. Choć pozostałe cztery zaproponowane bariery są uniwersalne, lecz w artykule przedstawiono ich cechy charakterystyczne dla rozwiązań big data. Zdaniem autorów szczególną uwagę należy zwrócić na zaawansowanie rozwiązań technologicznych, a także wysokie oczekiwania wobec osób zajmujących się opracowywaniem modeli i analizą dużych wolumenów danych (*data scientists*), co wpływa na wysokie koszty implementacji rozwiązań. Ważna jest też konieczność dostosowania procesów organizacyjnych do sytuacji, w której decyzje są często podejmowane w oparciu o korelację danych, zamiast zrozumienia przyczynowości zjawiska.

Ponadto należy stwierdzić że:

1. Tworzenie i wykorzystywanie nowej wiedzy jest koniecznym warunkiem dla powstania innowacji. Tym samym technologie big data, których celem jest ujawnienie nowych, nieznanych, a jednak istotnych korelacji pomiędzy danymi z różnych źródeł wpisują się wprost w rozwijanie działalności gospodarczej.
2. Technologie big data wpłyną w ten sam sposób na pojawianie się innowacji różnych typów: procesowych, produktowych, organizacyjnych i marketingowych, jak stało się to w przypadku powszechnego wykorzystania sieci Internet i będą tak samo znaczące dla rozwoju poszczególnych przedsiębiorstw i całych gospodarek.
3. Jednym z podstawowych zadań technologii big data jest pomoc w zmniejszeniu niepewności w podejmowaniu decyzji operacyjnych i inwestycyjnych poprzez łączenie danych dotyczących czynników produkcyjnych, ekonomicznych i środowiskowych.

Początkowa faza rozwoju technologii charakteryzuje się stosunkowo powolnym wzrostem możliwości względem ponoszonych nakładów. Potrzebne jest zatem odpowiednie zarządzanie ukierunkowane na systemowe przełamywanie barier i minimalizację ryzyka ich negatywnego oddziaływania. Dalsze planowane prace autorów ukierunkowane będą na zaproponowanie rozwiązań wspierających rozwój technologii big data w polskich przedsiębiorstwach i ośrodkach naukowych. Najbliższym elementem tych prac będzie ocena wpływu poszczególnych grup barier na wdrożenia tej innowacyjnej technologii. Planuje się przeprowadzenie stosownych badań ankietowych, a następnie przygotowanie macierzy, dla której obszarów zaproponowane zostaną rozwiązania wspierające rozwój technologii big data bazujące na sprawdzonych rozwiązaniach np. rozwoju komputerowo zintegrowanych systemów informatycznych wspierających zarządzanie przedsiębiorstwem.

Literatura

1. Parys T., Scenariusze wdrożenia a bariery wdrożeniowe zintegrowanych systemów informatycznych, *Innowacje w zarządzaniu i inżynierii produkcji*, tom 2, red. R. Knosala, Oficyna Wydawnicza Polskiego Towarzystwa Zarządzania Produkcją, Opole 2016, s.845-853, http://www.ptzp.org.pl/files/konferencje/kzz/artyk_pdf_2016/T2/t2_0845.pdf
2. Kisielnicki J., *Informatyczna infrastruktura zarządzania*, PWN, Warszawa 1993.
3. Foster M. J.: The value of formal planning for strategic decisions: A comment. *Strategic Management Journal*, vol. 7, 1986, s. 179–182.
4. Laney D.: *Application delivery strategies*. META Group, Stamford, 2001.
5. <http://www.gartner.com/it-glossary/big-data/>
6. Surma J., *Cyfryzacja życia w erze big data*, PWN, Warszawa 2017.
7. Mayer-Schonberger V., Cukier K.: *Big data – A revolution that will transform how we live, work, and think*. An Eamon Dolan Book Houghton Mifflin Harcourt, Boston, 2013.
8. Wieczorkowski J., Polak P.: Big data: Three-aspect approach, “*Online Journal of Applied Knowledge Management*”, vol. 2, issue 2, *International Institute for Applied Knowledge Management*, 2014, s. 182-196. http://www.iiakm.org/ojakm/articles/2014/volume2_2/OJAKM_Volume2_2pp182-196.pdf
9. Schumpeter J., *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*, Harvard University Press, Harvard 1934.
10. OECD, *Podręcznik Oslo. Zasady gromadzenia i interpretacji danych dotyczących innowacji. Pomiar działalności naukowej i technicznej*, Wydanie trzecie, Komisja Europejska, 2005.
11. Abernathy W., Utterback J., *Patterns of industrial innovation*, „*Technology Review*”, No 80/7, 1978, s.40-47.
12. Wieczorkowski J., Jurczyk-Bunkowska M., *Ocena potencjału innowacyjnego organizacji pod kątem wdrożenia rozwiązań IT – założenia modelu*, „*Ekonomiczne Problemy Usług*” nr 122, 2016, s. 391-400, http://www.wziew.pl/zn/epu/122/epu_122.pdf
13. Brewer E.A., *A Certain Freedom: Thoughts on the CAP Theorem*, *Proceeding of the XXIX ACM SIGACT-SIGOPS symposium on Principles of distributed computing*. — N. Y., ACM, 2010.
14. Pawełszek I., Wieczorkowski J., *Big data as a business opportunity: an Educational Perspective*, “*Annals of Computer Science and Information Systems*”, Volume 5, *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, Polskie Towarzystwo Informatyczne, IEEE Computer Society Press, 2015, s.1563-1568, <https://fedcsis.org/proceedings/2015/pliki/365.pdf>
15. McAfee A., Brynjolfsson E., *Big Data the Management Revolution*, “*Harvard Business Review*”, vol. 90. Issue 110, 2012, s. 60-68.
16. Erl T., Khattak W., Buhler P., *Big Data Fundamentals: Concepts, Drivers & Techniques*, Pearson Education, 2015.
17. Ustawa z dnia 29 sierpnia 1997 r. o ochronie danych osobowych
18. Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych).
19. Kozierski P., Kabaciński R., Lis M., Kaczmarek P., *Open Access. Analiza zjawiska z punktu widzenia polskiego naukowca*. Oficyna Wydawnicza "Impuls", Kraków 2013.

20. Pawełoszek I., Wybrane problemy wdrożenia koncepcji otwartych danych w e-administracji, „Roczniki Kolegium Analiz Ekonomicznych” 33/2014, Szkoła Główna Handlowa w Warszawie 2014, s.455-470.
http://rocznikikae.sgh.waw.pl/p/roczniki_kae_z33_26.pdf
21. Ustawa z dnia 6 września 2001 r. o dostępie do informacji publicznej.
22. Ustawa z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego.
23. Gałach A. i in, Dostęp do informacji publicznej a prawo do prywatności, Wydawnictwo C.H.Beck, Warszawa 2015.
24. <http://danepubliczne.gov.pl>
25. Russom P., Big data analytics, TDWI Research, Renton 2011.

Dr Jędrzej WIECZORKOWSKI
Instytut Informatyki i Gospodarki Cyfrowej
Kolegium Analiz Ekonomicznych
Szkoła Główna Handlowa w Warszawie
02-513 Warszawa, Madalińskiego 6/8
e-mail: jedrzej.wieczorkowski@sgh.waw.pl

Dr inż. Magdalena JURCZYK-BUNKOWSKA
Instytut Innowacyjności Procesów i Produktów
Politechnika Opolska
45-370 Opole, ul. Ozimska 75
tel.: (0-77) 423 40 44
e-mail: m.jurczyk-bunkowska@po.opole.pl