

Sławomir Leciejewski

## PROBLEM *BIG DATA* W NAUKACH EKSPERYMENTALNYCH

<https://doi.org/10.37240/FiN.2020.8.1.11>

### STRESZCZENIE

W artykule opisuję fenomen *big data* i jak ma się on do pracy badawczej realizowanej w ramach nauk eksperymentalnych. Poszukuję odpowiedzi na dwa ważne pytania. Czy metody badawcze zaproponowane w ramach paradygmatu *big data* mają zastosowanie w naukach eksperymentalnych? Czy zastosowanie metod badawczych z paradygmatu *big data* w konsekwencji prowadzi do nowego rozumienia tego, czym jest nauka?

**Słowa kluczowe:** *big data*, nauki eksperymentalne, filozofia nauki, metodologia nauk.

Nauki eksperymentalne generują olbrzymie ilości danych empirycznych (*big data*). Na przykład w latach 2011–12 Wielki Zderzacz Hadronów (LHC) działający w laboratorium CERN pod Genewą wygenerował 30 petabajtów danych. Samo przeczytanie takiej ilości danych zajęłoby około 500 mld lat, przy założeniu, że dane te zostałyby przekonwertowane na standardowy maszynopis oraz, że udałoby się czytać codziennie 100 stron takiego maszynopisu.

W tak zarysowanym kontekście zagadnienia *big data* w naukach eksperymentalnych, problemem badawczym niniejszego artykułu będzie poszukiwanie odpowiedzi na dwa ważne pytania. Czy metody badawcze zaproponowane w ramach paradygmatu *big data*<sup>1</sup> mają zastosowanie w naukach eksperymentalnych? Czy zastosowanie metod badawczych z paradygmatu *big data* w konsekwencji prowadzi do nowego rozumienia tego, czym jest nauka? Aby odpowiedzieć na powyższe pytania zacznę od scharakteryzowania tego, czym jest *big data*. Następnie porównam metody badawcze zaproponowane w ramach paradygmatu *big data* z dotychczasowymi metodami pracy badawczej w naukach eksperymentalnych. Następnie będę argumen-

---

<sup>1</sup> Poprzez określenie „paradygmat *big data*” w swoim artykule będę rozumiał sposób rozwiązywania problemów (składowanie, przetwarzanie i analizowanie dużych ilości danych) przez entuzjastów *big data*. Nawiązuję zatem do węższego rozumienia paradygmatu w sensie Thomasa Kuhna rozumianego jako wzorzec rozwiązywania problemów. Por. K. Jodkowski, *Wspólnoty uczonych, paradygmaty i rewolucje naukowe*, Wydawnictwo UMCS, Lublin 1990, s. 140–147.

tował na rzecz tezy, że przyjęcie metod badawczych paradygmatu *big data* doprowadziłoby do radykalnej rewizji dotychczasowego rozumienia tego, czym jest nauka.

## 1. CHARAKTERYSTYKA *BIG DATA*

Fenomen *big data* najczęściej opisuje się za pomocą trzech charakterystyk, zwanych 3V. Pierwszą z nich jest duży wolumen danych (*volume*) liczony w petabajtach ( $10^{15}$  bajtów) i exabajtach ( $10^{18}$  bajtów) danych; drugą – dostępność w czasie rzeczywistym intensywnego przyrostu danych, tzn. możliwość przesyłania strumieniowego w czasie rzeczywistym olbrzymich ilości danych (w angielskim nazywany *velocity*). Trzecią charakterystyką *big data* jest duża różnorodność formatów danych (*variety*), np. liczby, teksty, zdjęcia, pliki audio oraz video.<sup>2</sup> Innymi słowy *big data* to „aktywa informacyjne charakteryzujące się dużym wolumenem danych, intensywnym przyrostem w czasie oraz różnorodnością formatów danych, które wymagają adekwatnych metod składowania, przetwarzania i analizowania w celu osiągnięcia określonych celów (ekonomicznych, społecznych, politycznych itp.)”.<sup>3</sup>

Najwymowniejszym przykładem dla *big data* jest firma Google, która składa zbierane przez siebie dane na kilkunastu farmach serwerów rozrzuconych po całym świecie. Firma ta codziennie przetwarza ponad 24 petabajty danych (przeczytanie 1 PB danych zajęłoby około 16 mld lat przy założeniu, że czyta się 100 stron maszynopisu dziennie). Także Facebook operuje w skali *big data*, gdyż co godzinę przetwarza 10 milionów nowych fotografii a każdego dnia obsługuje 3 mld kliknięć w emotikony (np. „lubię to”, „super” itd.) oraz komentarzy pod postami. Innym przykładem *big data* jest serwis YouTube, który co sekundę dodaje godzinę nowych filmów przesłanych przez niektórych z 800 milionów użytkowników tej platformy.<sup>4</sup>

Także w naukach empirycznych znajdziemy przykłady z dziedziny *big data*. Na przykład w ramach programu obserwacji nieba o nazwie *Sloan Digital Sky Survey* teleskopy pracujące w ramach tego przedsięwzięcia w ciągu kilku tygodni 2000 roku zebrały więcej danych niż w historii całej astronomii od jej początku do końca XX wieku. Przez 10 lat w ramach tego projektu zebrano 140 terabajtów danych astronomicznych. Jednakże następca *Sloan*

<sup>2</sup> Por. J. Surma, *Cyfryzacja życia w erze BIG DATA. Człowiek, biznes, społeczeństwo*, Wydawnictwo Naukowe PWN SA, Warszawa 2017, s. 23–29.

<sup>3</sup> Ibidem, s. 24.

<sup>4</sup> Informacje dotyczące ilości przetwarzanych danych przez serwisy Google, Facebook i YouTube pochodzą z roku 2013. Z pewnością współcześnie są znacznie większe, co oznacza, że cały czas podpadają pod problematykę *big data*. Por. V. Mayer-Schöneberger, K. Cukier, *BIG DATA efektywna analiza danych. Rewolucja, która zmieni nasze myślenie, pracę i życie*, MT Biznes, Warszawa 2017, s. 22.

*Digital Sky Survey* o nazwie *Large Synoptic Survey Telescope*, który zaczął działać w 2016 roku, gromadzi taką liczbę danych (140 terabajtów) co 5 dni (tj. około 10 petabajtów na rok). Zatem mamy do czynienia z ogromnym przyspieszeniem w dziedzinie gromadzenia danych w naukach empirycznych. Gromadzenie danych empirycznych w astronomii zajmujące na początku XX wieku 10 lat trwa obecnie 5 dni.<sup>5</sup>

Przywoływany wcześniej LHC w CERN od 2011 do 2012 roku zgromadził 30 petabajtów danych. Warto jednak podkreślić, że archiwizowano tylko 0,1% danych generowanych przez LHC, gdyż nie było możliwości gromadzenia wszystkich danych; wymyślono zatem systemy filtrujące dane w czasie rzeczywistym, które kasowały znane i opracowane wcześniej efekty zderzeń cząstek elementarnych. Gdyby zapisywano wszystko, byłoby to 30 eksabajtów (1 eksabajt to 1024 petabajty), tj. 10% wszystkich danych cyfrowych zgromadzonych na świecie pod koniec pierwszej dekady XXI wieku.<sup>6</sup>

Do analizy tak ogromnych baz danych potrzebne są zaawansowane i zautomatyzowane narzędzia informatyczne, które nazwano technologią eksploracji danych (*data mining*). Zadaniem tych metod eksploracji danych jest „odkrywanie nietrywialnych, dotychczas nieznanymi zależności, związków, podobieństw lub trendów – ogólnie nazywanych wzorcami (*patterns*) – w dużych repozytoriach danych”.<sup>7</sup> Odkryty wzorec reprezentować może jakąś regularność; może to być: reguła, charakterystyka zbioru obiektów lub opis powtarzalnego sposobu zachowania. Metody te służą zatem do odkrywania wiedzy w bazach danych (*knowledge discovery in databases*).<sup>8</sup>

Warto podkreślić, że wcześniej stosowana standardowa analiza danych miała na celu uzyskanie odpowiedzi na postawione jednoznacznie pytanie, np. czy w zbiorze danych empirycznych uzyskanych w eksperymencie LHC w CERN istnieją takie zderzenia, w wyniku których mógł powstać bozon Higgsa postulowany w badaniach teoretycznych. Znaczy to, że użytkownik formułował zapytania i dokonywał analizy danych zawartych w bazie. Tak więc standardowa analiza danych metodologicznie związana była z testowaniem hipotez, czyli sytuacją, kiedy weryfikacji podlega jakaś teoria.<sup>9</sup>

Technologia eksploracji danych natomiast pozwala odkryć coś, czego wcześniej nie wiedzieliśmy, czyli niejednokrotnie nie było jakichkolwiek przesłanek, aby dane pytanie w ogóle sformułować (np. pojawienie się epidemii wirusa H1N1 w latach 2007–2008 na podstawie analizy zapytań w internetowej wyszukiwarce Google). W wyniku takiej automatycznej ana-

<sup>5</sup> Por. *ibidem*, s. 21.

<sup>6</sup> Por. *ibidem*, s. 257.

<sup>7</sup> T. Morzy, *Eksploracja danych. Metody i algorytmy*, Wydawnictwo Naukowe PWN, Warszawa 2013, s. 3.

<sup>8</sup> W literaturze przedmiotu termin „odkrywanie wiedzy w bazach danych” używany jest zamiennie z terminem „technologia eksploracji danych” i oznacza on szukanie wzorców. „Odkrywane w procesie eksploracji danych wzorce mają, najczęściej, postać reguł logicznych, klasyfikatorów (np. drzew decyzyjnych), zbiorów skupień, wykresów itp.” *Ibidem*, s. 3.

<sup>9</sup> Por. *ibidem*, s. 3–4.

lize dowiadujemy się czegoś zupełnie nowego o badanym fragmencie świata i w tym kontekście jest to właśnie odkrywanie wiedzy. Najczęściej bowiem w przypadku analiz *big data* podstawa teoretyczna służąca do zadawania konkretnych pytań nie istnieje i działania te mają raczej charakter „ślepej” eksploracji. Takie podejście może być owocne, gdy dysponujemy dużą mocą obliczeniową i posiadamy olbrzymie ilości danych, które możemy analizować.

Biorąc pod uwagę format analizowanych w ten sposób danych, metody ich eksploracji podzielić można na: standardową eksplorację danych (dla danych ilościowych i jakościowych)<sup>10</sup>, eksplorację danych tekstowych i przetwarzanie języka naturalnego (dla danych typu ciąg znaków, tekst, tekst wygenerowany z pliku audio)<sup>11</sup>, eksplorację sieci społecznych i mediów społecznościowych (dla danych jakościowych, ilościowych, danych relacyjnych reprezentujących powiązania w sieci i dla grafów)<sup>12</sup>, przetwarzanie i analiza multimediów (dla różnorodnych danych multimedialnych reprezentujących: zdjęcia, grafiki, dźwięki, filmy, animacje itd.).<sup>13</sup>

## 2. *BIG DATA* A DOTYCHCZASOWE METODY PRACY BADAWCZEJ

Współcześnie działające ośrodki naukowe generują olbrzymie ilości danych w każdej niemalże dziedzinie naukowej (fizyka, astronomia, biologia, nauki techniczne i społeczne). Peter Galison w monografii *How Experiments End* wyróżnienia dwie kultury badawcze nauki: kulturę teoretyczną i eksperymentalną.<sup>14</sup> W książce *Image and Logic*, w kulturze ekspery-

<sup>10</sup> Zadania standardowej eksploracji danych to: klasyfikacja (przyporządkowywanie obiektu do jednej z wcześniej określonych klas), estymacja (zbliżone do klasyfikacji, tylko klasy mają ciągle zbiór wartości), grupowanie (znajdywanie w zbiorze obiektów podzbiorów obiektów o podobnych własnościach), odkrywanie reguł asocjacyjnych (wyszukiwanie zbiorów obiektów, które występują razem w określonym kontekście). Por. J. Surma, op. cit., s. 111–112.

<sup>11</sup> Eksploracja danych tekstowych polega na wydobywaniu informacji z nieustrukturyzowanych danych tekstowych. Metody te mogą być wykorzystywane do realizacji następujących zadań: klasyfikacji dokumentów tekstowych (przyporządkowanie dokumentu do wcześniej ustalonych klas), grupowania dokumentów tekstowych (identyfikacja zbiorów dokumentów podobnych ze względu na pewną wspólną charakterystykę), ekstrakcji informacji (wyodrębnienie z dokumentu tekstowego obiektów o określonej interpretacji semantycznej), obróbki edytorskiej (automatyczne opracowanie dokumentu, np. usuwanie błędów ortograficznych, interpunkcyjnych itp. oraz opracowanie streszczeń), analizy wydźwięku (automatyczne rozpoznanie opinii autora na poruszany temat), przetwarzania języka naturalnego (dział sztucznej inteligencji, którego zadaniem jest analiza semantyki i udzielanie adekwatnej odpowiedzi na zadane pytanie). Por. ibidem, s. 113.

<sup>12</sup> Podstawowe zadania realizowane przez eksplorację sieci społecznych i mediów społecznościowych to: analiza społeczności (np. identyfikacja przynależności do danej społeczności na podstawie internetowych wpisów), dyfuzja informacji (analiza rozprzestrzeniania się informacji w sieci), analiza wpływu społecznego (identyfikacja przekazywanych informacji oraz zaistniałych pożądanych zachowań adresata informacji), analiza zachowania (modelowanie prawdopodobieństwa wykonania określonego działania w sieci). Por. ibidem, s. 114–115.

<sup>13</sup> Por. ibidem, s. 107–118.

<sup>14</sup> Por. P. Galison, *How Experiments End*, University of Chicago Press, Chicago 1987.

talnej dokonuje wyróżnienia dwóch tradycji generowania danych empirycznych: tradycji wizualnej (inaczej – obrazowej, *image*) oraz tradycji numerycznej, którą nazywa logiczną (*logic*).<sup>15</sup> Według niego te dwie tradycje konstituuje używanie danego typu detektorów. Tradycję wizualną konstituują detektory wizualne (np. komory pęcherzykowe), które dostarczają informacji o poszczególnych, jednostkowych przypadkach danego zjawiska. Tradycję numeryczną natomiast – detektory elektroniczne (np. licznik Geigera), które dostarczają mniej szczegółowych informacji o poszczególnych przypadkach danego zjawiska, ale wykrywają ich znacznie więcej.

Wymienione powyżej tradycje (wizualna i numeryczna) kształtują dwie różne tradycje epistemiczne, które mają swój wyraz np. w różnym sposobie argumentowania. W pierwszym wypadku, tradycji wizualnej, uzasadnia się odwołując do szczegółowych obrazów, a w drugim, tradycji numerycznej – do wyników liczbowych uzyskiwanych z przyrządów pomiarowych; ich opracowanie opiera się na zastosowaniu argumentacji statystycznej wykorzystującej dużą liczbę wyników.

Warto jednak nadmienić, że w gruncie rzeczy współczesne badania z zakresu fizyki cząstek elementarnych prowadzone są łącznie w ramach tych dwóch tradycji, gdyż np. w LHC w CERN używa się łącznie dwóch rodzajów detektorów: detektorów śladowych, które pokazują tor cząstki (tj. tradycja wizualna) oraz kalorymetrów, które wyznaczają energię cząstek (tj. tradycja numeryczna, czy elektroniczna).

Warto zatem zadać następujące pytanie: czy wcześniej omówione metody badawcze zaproponowane w ramach paradygmatu *big data* mogą mieć zastosowanie w naukach eksperymentalnych? Wydaje się, że odpowiedź na to pytanie jest pozytywna, gdyż – omówiona wyżej – standardowa eksploracja danych (klasyfikacja, estymacja, grupowanie, odkrywanie reguł asocjacyjnych w ramach numerycznych danych ilościowych) może być przydatna do analizy danych uzyskanych w ramach proponowanej przez Galisona tradycji numerycznej. Przetwarzanie i analiza multimedialnych danych może być natomiast przydatna przy analizie danych z tradycji wizualnej. Skoro bowiem współczesne eksperymenty, jak na przykład LHC w CERN, generują duże ilości obrazów (zderzeń) i danych liczbowych (energii uzyskanych w zderzeniu cząstek), to analizy obrazów i danych liczbowych z tradycji *big data* mogą być tutaj bardzo przydatne. Zarówno bowiem w paradygmacie *big data*, jak i w naukach eksperymentalnych mamy do czynienia z dużym wolumenem danych (np. poczta elektroniczna każdego roku generuje 400 petabajtów danych a LHC po odfiltrowaniu w dwa lata wygenerował 30 petabajtów danych; bez kasowania danych byłoby to 15 000 petabajtów na rok). Z tym przykładem wiąże się jednak jedna podstawowa różnica pomiędzy standar-

<sup>15</sup> Por. P. Galison, *Image and Logic. A Material Culture of Microphysics*, University of Chicago Press, Chicago 1997.

dowym ujęciem *big data* a *big data* w naukach eksperymentalnych. W paradygmacie *big data* zachowuje się wszystkie napływające dane a w naukach eksperymentalnych często nie da się tego zrobić, gdyż danych jest zbyt dużo (np. w LHC zachowuje się 0,1% danych napływających z detektorów).

Kolejna różnica pomiędzy sposobem dochodzenia do wiedzy w paradygmacie *big data* a dotychczasowymi metodami pracy badawczej jest bardzo znacząca. W paradygmacie *big data* bowiem ważna jest przede wszystkim odpowiedź na pytanie, co się dzieje (w badanym obszarze świata), a nie dlaczego dzieje się tak a nie inaczej.<sup>16</sup> Ilustrują to, wyżej wzmiankowane przewidywania firmy Google dotyczące wirusa H1N1 oraz rekomendacje firmy Amazon. Jedna trzecia sprzedaży tej firmy generowana jest przez systemy spersonalizowanej rekomendacji bazujące na paradygmacie *big data*. Odkryta korelacja: dotychczasowe zakupy – zakupy w przyszłości została znaleziona bez znajomości przyczyn leżących u jej podstaw.<sup>17</sup>

Warto podkreślić, że w naukach eksperymentalnych interesuje nas zarówno odpowiedź na pytanie, co się dzieje z badanym układem (tj. obserwacja, manipulacja, pomiar itd.), jak i dlaczego proces (zjawisko) ma taki a nie inny przebieg (tj. wyjaśnianie poprzez odwołanie się do teorii itp.). W naukach eksperymentalnych nie poprzestajemy zatem tylko i wyłącznie na pytaniu „co”, ale również zadajemy doniosłe pytanie „dlaczego”, gdyż jednym z ważniejszych zadań nauki jest wyjaśnianie.<sup>18</sup>

W paradygmacie *big data* celem prowadzonych badań jest szukanie korelacji (odpowiedź na pytanie „co?”). W naukach eksperymentalnych natomiast: kreowanie nowych zjawisk (wytwarzanie, doskonalenie, stabilizowanie zjawisk itd. według nowych eksperymentalistów) oraz wyjaśnianie (sprawdzanie, przewidywanie itd. według teoretycyków)<sup>19</sup>. Podstawowym zaś wzorcem wyjaśniania jest model dedukcyjno-nomologiczny,<sup>20</sup> w ramach którego wyjaśnić prawidłowość to wyprowadzić ją z ogólnego prawa lub kilku praw.<sup>21</sup>

<sup>16</sup> W *big data* „ważna jest odpowiedź na pytanie, co się dzieje, a nie dlaczego. Nie zawsze musimy znać przyczyny jakiegoś zjawiska, możemy po prostu pozwolić danym mówić za siebie”. V. Mayer-Schöneberger, K. Cukier, *BIG DATA efektywna analiza danych. Rewolucja, która zmieni nasze myślenie, pracę i życie*, op. cit., s. 30.

<sup>17</sup> „Innowacyjne systemy rekomendacji Amazona znalazły wartościową korelację bez znajomości leżących u jej podstaw przyczyn. Wiedza co, a nie dlaczego jest wystarczająca”. Ibidem, s. 76.

<sup>18</sup> „Pytanie o wyjaśnianie najczęściej zaczyna się od »dlaczego«?”. A. Grobler, *Metodologia nauk*, Wydawnictwo Aureus – Wydawnictwo ZNAK, Kraków 2006, s. 112.

<sup>19</sup> Por. P. Zeidler, *Nowy eksperymentalizm a teoretycyzm. Spór o przedmiot i sposób uprawiania filozofii nauki*, w: *Nowy eksperymentalizm – teoretycyzm – reprezentacja*, D. Sobczyńska, P. Zeidler (red.), Poznań, Wydawnictwo Naukowe IF UAM, s. 87–108.

<sup>20</sup> Istnieją także inne sposoby wyjaśniania: wyjaśnianie dedukcyjno-statystyczne, indukcyjno-statystyczne itd. Por. A. Grobler, *Metodologia nauk*, op. cit., s. 103–112.

<sup>21</sup> „Wyjaśnianie zdarzenia polega na wyprowadzeniu jego opisu z pewnego prawa (lub kilku praw) i zdań opisujących warunki początkowe”. Ibidem, s. 104.

### 3. PARADYGMAT *BIG DATA* JAKO NOWE ROZUMIENIE NAUKOWOŚCI

Warto mocno podkreślić, że w *big data* nie chodzi o wyjaśnianie, ale o informowanie o znalezionych korelacjach. Innymi słowy, w przeciwieństwie do nauk eksperymentalnych, *big data* rezygnuje z wyjaśniania. Entuzjaści takiego podejścia do badań dużych zbiorów danych twierdzą nawet, że „zalew danych sprawił, iż metody naukowe stały się przestarzałe”<sup>22</sup> a ich ilość skumulowała się do takiego poziomu, że nastąpił niemal „koniec teorii”. Według zwolenników paradygmatu *big data* tradycyjny proces odkryć naukowych (tj. np. stawianie hipotez i empiryczne ich testowanie) traci na znaczeniu i jest zastępowany wyszukiwaniem korelacji, za którymi nie stoi żadna teoria. Przekonują także, że w epoce *big data* nie potrzebujemy teorii („wszystkie modele są błędne”<sup>23</sup>), po prostu analizujemy dane.

Czy jednak paradygmat *big data* nie potrzebuje żadnych (teoretycznych) modeli pojęciowych? Cała maszyna związana z *big data* potrzebuje przecież do swojego działania teorii statystycznych oraz matematycznych dotyczących podstaw obliczeń, bez których nie można by opracować skutecznych algorytmów oraz poprawnie działających systemów komputerowych wraz ze stosownym oprogramowaniem implementującym opracowane wcześniej algorytmy. Już choćby to wystarcza do postawienia przeciwnej względem entuzjastów *big data* tezy, że paradygmat ten w rzeczywistości nie oznacza końca teorii.

Warto także zauważyć, że korelacje i przewidywania stworzone za pomocą algorytmów *big data* są często zbyt skomplikowane, aby zrozumieli je ludzie.<sup>24</sup> *Big data* operuje w skali petabajtów (1 PB = 1024 TB = 1024\*1024 GB), eksabajtów (1 EB = 1024 PB) i zettabajtów (1 ZB = 1024 EB). Jak wspomniałem wcześniej przeczytanie choćby 1 PB danych zajęłoby około 16 mld lat a to stanowczo wykracza poza możliwości każdego człowieka. Korzystając z dobrodziejstw *big data* musimy zadowolić się tylko i wyłącznie uproszczonymi przez systemy komputerowe, a dzięki temu - zrozumiałymi dla ludzi wynikami analiz *big data* bez możliwości skontrolowania poprawności sposobu ich uzyskania.

Jeśli zatem w badaniach naukowych, także w dziedzinie nauk eksperymentalnych, zadowolilibyśmy się tylko i wyłącznie odpowiedziami na pytania typu: co z czym koreluje, wtedy w całości moglibyśmy przejść do opracowywania danych naukowych metody badawcze wypracowane w obrębie paradygmatu *big data*. Jednakże musielibyśmy także zrezygnować z dotych-

<sup>22</sup> Ch. Anderson, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, Wired, 16.07, July 2008.

<sup>23</sup> Ibidem.

<sup>24</sup> „Data science zakłada wykorzystanie komputerów do automatycznego przetwarzania i analizowania dużych zbiorów danych”. M. Szeliga, *DATA SCIENCE i uczenie maszynowe*, Wydawnictwo Naukowe PWN SA, Warszawa 2017, s. XVIII.

czasowych odpowiedzi na pytanie o to, czym są nauki eksperymentalne. W nich bowiem nie tylko ważna jest odpowiedź na pytanie: co z czym koreluje, ale także – a może przede wszystkim – odpowiedź na pytanie: dlaczego dzieje się tak a nie inaczej. Na to ostatnie pytanie w ramach paradygmatu *big data* nie znajdziemy jakiegokolwiek sensownej odpowiedzi.

### BIBLIOGRAFIA

Ch. Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, Wired, 16.07, July 2008.

P. Galison, *How Experiments End*, University of Chicago Press, Chicago 1987.

\_\_\_\_\_, *Image and Logic. A Material Culture of Microphysics*, University of Chicago Press, Chicago 1997.

A. Grobler, *Metodologia nauk*, Wydawnictwo Aureus – Wydawnictwo ZNAK, Kraków 2006.

K. Jodkowski, *Wspólnoty uczonych, paradygmaty i rewolucje naukowe*, Wydawnictwo UMCS, Lublin 1990.

V. Mayer-Schöneberger, K. Cukier, *BIG DATA efektywna analiza danych. Rewolucja, która zmieni nasze myślenie, pracę i życie*, MT Biznes, Warszawa 2017.

T. Morzy, *Eksploracja danych. Metody i algorytmy*, Wydawnictwo Naukowe PWN, Warszawa 2013.

J. Surma, *Cyfryzacja życia w erze BIG DATA. Człowiek, biznes, społeczeństwo*, Wydawnictwo Naukowe PWN SA, Warszawa 2017.

M. Szeliga, *DATA SCIENCE i uczenie maszynowe*, Wydawnictwo Naukowe PWN SA, Warszawa 2017.

P. Zeidler, *Nowy eksperymentalizm a teoretycyzm. Spór o przedmiot i sposób uprawiania filozofii nauki*, w: *Nowy eksperymentalizm – teoretycyzm – reprezentacja*, D. Sobczyńska, P. Zeidler (red.), Poznań, Wydawnictwo Naukowe IF UAM, s. 87–108.

### THE BIG DATA PROBLEM IN EXPERIMENTAL SCIENCES

#### ABSTRACT

In the paper the phenomenon of *big data* is presented. I pay my special attention to the relation of this phenomenon to research work in experimental sciences. I search for answers to two questions. First, do the research methods proposed within the paradigm *big data* can be applied in experimental sciences? Second, does applying the research methods subject to the *big data* paradigm lead, in consequence, to a new understanding of science?

**Keywords:** *big data*, experimental sciences, philosophy of science, methodology of science.

O AUTORZE — dr hab., prof. UAM, Wydział Filozoficzny, Uniwersytet im. Adama Mickiewicza, ul. Szamarzewskiego 69 c, Poznań.

E-mail: slaaw@amu.edu.pl