

Mgr Ewelina Dziwak

Uniwersytet Śląski

e-mail: dziwak1987ewelina@gmail.com

Wielkie bazy danych jako nowe źródła informacji. *Big data* przyszłością dziennikarstwa?

Wprowadzenie

XXI wiek przez wielu badaczy – przedstawicieli zarówno nauk społecznych, jak i innych dyscyplin – nazywany jest trafnie epoką nadmiaru. Współczesny człowiek nie może skosztować wszystkiego, co ślepo gromadzi, kierując się reklamą, a także nie potrafi skupić uwagi na wszelkich informacjach, które udostępnia/przechowuje się. Jednostka egzystuje więc w „cyfrowym buszu”, żyjąc już nie w zgodzie z naturą, lecz z technologią.

W obliczu wielu źródeł informacji – począwszy od mediów tradycyjnych, a skończywszy na komentarzach i postach internautów w sieciach społecznościowych, kiedy sama informacja może być wielokrotnie filtrowana, oceniana, komentowana i przetwarzana – dziennikarze stają przed nowym wyzwaniem. Muszą znaleźć taką historię, do której trudno jest dotrzeć, która nie jest powszechnie dostępna. Dodatkowo, aby wzbudzić zainteresowanie publiczności treścią, muszą przedstawić ją w nieszablonowy, jak najbardziej atrakcyjny dla konsumentów sposób. Prezentowany artykuł dotyczy nowej praktyki dziennikarskiej, jaką jest *data journalism*. Autorka dokonuje jej charakterystyki, wskazując na rosnące znaczenie wizualizacji informacji w obliczu przeciążenia informacyjnego. Głównym celem artykułu jest ukazanie dziennikarstwa danych jako remedium na problemy percepcyjne jednostek. Autorka swoje rozważania rozpoczyna od charakterystyki *big data*, następnie przechodzi do zagadnienia rafinacji informacji. Kolejnym elementem, któremu poświęca uwagę, jest wizualizacja przekazów. W tym fragmencie pracy podaje argumenty przemawiające za traktowaniem dziennikarstwa danych jako środka zaradczego na przeciążenie informacyjne. Autorka przedstawia podstawowe ustalenia na temat *data journalism*, opierając założenia głów-

nie na dostępnych studiach teoretycznych podejmujących tę wciąż mało znaną problematykę.

***Big data* jako nieocenione źródło informacji**

Założone w Stanach Zjednoczonych przedsiębiorstwo analityczno-doradcze Gartner szacuje, że do 2020 roku liczba urządzeń podłączonych do sieci internetowej wyniesie dwadzieścia sześć miliardów. Według ABI Research liczba ta będzie oscylować wokół trzydziestu miliardów. Cisco Internet Business Solutions Group w swoich prognozach wskazuje na pięćdziesiąt miliardów (Mayer-Schönberger, Cukier 2014, 24). Każde z urządzeń działających i komunikujących się ze sobą przy wykorzystaniu internetu wytwarza ogromne wolumeny danych. Codziennie różne źródła generują masowy przyrost danych w postaci cyfrowej. Powoduje to, że konwencjonalne techniki przechowywania i przetwarzania danych zaczynają stawać się nieadekwatne do obecnych potrzeb (Tabakow, Korczak, Franczyk 2014, 138). W czasach cyfrowej rewolucji wszechobecność informacji zaczyna odgrywać kluczową rolę w egzystencji człowieka współczesnego. Człowiek – pojmowany do niedawna przez socjologów jako istota niepotrafiąca żyć w separacji od społeczeństwa – okazuje się jednostką, która podobnie jak za drugim człowiekiem, tęskni za informacją. Informacja (choć ogólnie dostępna) staje się jednak zarazem wszystkim i niczym.

Coraz bardziej powszechny dostęp do internetu oraz upowszechnienie się umiejętności obsługi urządzeń typu *smart* sprawiają, że wgląd do danych i informacji jest łatwiejszy. Użytkownik zostaje uwikłany w niezliczone sieci. Wykorzystuje do nich dostęp, ale równocześnie staje się autorem informacji oraz uczestniczy w ich rozpowszechnianiu. Według Martina Hilberta ilość cyfrowych informacji podwaja się co trzy lata (Mayer-Schönberger, Cukier 2014, 23). Miliony zdjęć, plików muzycznych, plików wideo, dokumentów tekstowych itp. z każdą sekundą zalewają sieć. Dane te, dotyczące niemalże każdej sfery życia, zaczęto określać przymiotnikiem *big*, a całe ich zbiory otrzymały nazwę *big data* (Mach-Król 2015, 41). Nie istnieje jedna, uniwersalna definicja tego terminu. Określenie *big data* odnosi się do zbiorów danych, których rozmiar przekracza możliwości pozyskiwania, przechowywania, zarządzania i analizy oferowane przez typowe narzędzia oprogramowania bazodanowego. Bezpośrednią przyczyną rozwoju *big data* jest proces datafikacji, czyli tendencji do wytwarzania danych dotyczących zjawisk, które uprzednio nie były traktowane jako interesujące źródła informacji. W wielu przypadkach dane te są gromadzone bez konkretnego powodu, poza ogólnym przekonaniem, że w niedalekiej przyszłości będzie można dzięki nim usprawnić jakiś proces lub uzyskać korzyści ekonomiczne. Warto zaznaczyć, że źródłem *big data* jest nie tylko internet, ale są nim także wszelkiego rodzaju wielkie zbiory danych, m.in. archiwa państwo-

we, samorządowe, korporacyjne, firmowe oraz wszystkie te, w których rejestrowane są przejawy ludzkiej działalności (Brosz, Bryda, Siuda 2017, 7).

Zasoby internetu stanowią bez wątpienia znaczącą częścią *big data*. Dane pochodzące z nieustrukturyzowanych źródeł, takich jak: sieci społecznościowe, strumień kliknięć w WWW, blogi, portale, posty, maile, profesjonalne publikacje, zasoby ukryte czy dane produkowane przez urządzenia typu *smart*, tworzą olbrzymie zbiory informacji, których analiza pozwala na uzyskanie nowej wiedzy. *Big data* daje możliwość przewidywania przyszłości i odkrywania nowych wartości na podstawie bieżącej analizy nieustannie zwiększających się zasobów informacji (Graczyk-Kucharska 2015, 267). Analizy danych wykorzystywane są w celu wychwycenia korelacji, wzorców i trendów pomocnych zarówno w prognozowaniu zjawisk oraz tendencji, jak i ludzkich preferencji, zachowań czy nawyków (Stępnik 2015, 150). Ciągłe wzrastająca ilość danych, a także doskonalenie technik ich pozyskiwania, przechowywania i analizowania umniejszają rolę teorii w odkrywaniu praw rządzących społeczeństwami ludzkimi. *Big data* same wytwarzają zdolność do poznania wiedzy dotyczącej skomplikowanych zjawisk. Analizy dokonywane w ramach tego paradygmatu nie wymagają stosowania się do procedur procesu badawczego. Uzyskane dane mają „mówić” same za siebie (Brosz, Bryda, Siuda 2017, 9).

Bardzo istotnym pożytkiem wynikającym z *big data* jest możliwość przewidywania kryzysów społecznych i gospodarczych dzięki wykorzystaniu danych pozyskiwanych z serwisów społecznościowych. Bezbłędnie zastosowane wobec dużych zbiorów danych algorytmy mogą wskazywać na kierunki w myśleniu i postrzeganiu określonych zagadnień, polaryzację poglądów politycznych, gospodarczych, społecznych i religijnych (Polańska, Wassilew 2015, 122). Wnioskowanie z analizy danych zebranych na portalach społecznościowych pozwala poznać najnowsze trendy, preferencje oraz cyfrowe nawyki członków internetowych społeczności. Bazując na zasobach dostępnych w profilach użytkowników, można uzyskać na ich temat zarówno dane demograficzne, jak i informacje pochodzące ze śladów pozostawianych przez nich w rzeczywistości wirtualnej i pozawirtualnej.

Rafinacja informacji w przestrzeni *big data*

Rzeczywistość sieciowa dawno już przestała być uważana za alternatywę dla tradycyjnie rejestrowanej rzeczywistości. Coraz częściej dostrzega się jej immanentny wymiar i traktuje na równi z rzeczywistością realną. Prowadząc badania społeczne w internecie, należało wypracować nowe kategorie opisu, nowe narzędzia oraz nowe paradygmaty – adekwatne do nieszablonowego, „technologizowanego” społeczeństwa. To, o co walczyli zwolennicy teorii zakładających całkowite oderwanie nauk społecznych od nauk przyrodniczych, w obecnych czasach staje się całkowicie bezzasadne. Technologia wymusiła poniekąd na „humanistach” się-

gnięcie do nauk ścisłych. Digitalny wymiar życia mający totalizujący charakter zawłaszcza wszystkie, nawet nieusieciowione wymiary życia społecznego (Szpunar 2016, 355–356). Informacje w internecie zostają pozbawione linearnego charakteru, zostają zredukowane do formatu cyfry i przekształcają się w dane.

Dane w społeczeństwie informacyjnym pełnią kluczową funkcję. Wszystko za sprawą kultury algorytmów, która wymusza imperatyw policzalności wszystkiego. Kultura nastawiona na zysk redukuje całą rzeczywistość społeczną do formatu cyfry. Bardziej niż jakość zaczyna liczyć się ilość posiadanych rzeczy, znajomych w sieciach społecznościowych, „lajków” czy wygenerowanych i skonsumowanych informacji. Towarzyszący społeczeństwu informacyjnemu proces makdonaldyzacji sprawia, że najważniejszy okazuje się krótki i zwięzły przekaz, a słowa poboczne, będące wyrazem kunsztu zawodowego dziennikarza, często okazują się zbędne dla przeciętnego czytelnika. Internet jest tym medium, które w największym stopniu sprzyja tworzeniu krótkich, łatwych w odbiorze treści. Tradycyjne dziennikarstwo, oparte na dłuższej narracji, w internecie powoli zanika. Ustępuje miejsca dziennikarstwu danych, które przemawia do czytelnika bardziej niż najlepiej nawet napisany tekst. Dziennikarstwo danych nie ma jednej, formalnej definicji. Pojmowane jest jako umiejętne wykorzystywanie dostępnych w sieci terabajtów do tworzenia prostych i zwięzłych treści opatrzonych wizualizacją. Obok tradycyjnego dziennikarstwa pojawia się dziennikarstwo danych, a obok dziennikarza humanisty wyłania się dziennikarz biegle władający nie tylko „piórem”, lecz także językiem programowania.

Informacje pochodzące z różnych źródeł, poddane digitalizacji, a następnie zewidencjonowane w *big data*, stają się fundamentem do tworzenia nowych treści. Proces ten – nazywany rafinacją informacji – jest wynikiem korelacji oraz analiz danych źródłowych (Gogołek 2012). Podstawą rafinacji jest *Culturomics* – narzędzie, dzięki któremu można monitorować media na całym świecie. Prototyp tego narzędzia narodził się jako owoc współpracy harwardzkich badaczy z pracownikami *Google Books*. W założeniu swoich twórców istotą działania *Culturomicsu* było odzwierciedlenie popularności danego słowa w odniesieniu do częstotliwości pojawiania się go w książkach pochodzących z konkretnego okresu. Koncepcja tego narzędzia została „zapożyczona” i wykorzystana przez *Institute for Computing in Humanities, Arts, and Social Sciences* jako próba użycia codziennych wiadomości i spojrzenie na ich geograficzny rozkład, tak aby można było sprawdzić, w jaki sposób odpowiednie dane się ze sobą łączą (Leetaru 2011). Innym przykładem tego typu projektu jest *The Migrants' Files* skupiający interdyscyplinarne zespoły dziennikarzy, programistów i projektantów mających narzędzia analityczne, które umożliwiają prowadzenie statystycznych śledztw dziennikarskich. Program powstał po tragedii, jaka rozegrała się u wybrzeży włoskiej wyspy Lampedusa i pochłonęła ponad trzysta sześćdziesiąt ofiar wśród uchodźców wojennych z Afry-

ki, Bliskiego Wschodu oraz innych regionów. Po tym wydarzeniu konsorcjum europejskich dziennikarzy zobowiązało się do systematycznego gromadzenia i analizowania danych na temat śmierci potencjalnych migrantów. Pozyskując, przetwarzając i analizując zestawy danych pochodzących z różnych źródeł, *The Migrants' Files* tworzy kompleksową i wiarygodną bazę danych na temat śmierci migrantów. Głównymi źródłami pozyskiwania danych do bazy są: *UNITED for Intercultural Action* – organizacja non profit, której sieć obejmuje ponad pięćset pięćdziesiąt organizacji z całej Europy, *Fortress Europe* – blog założony przez dziennikarza Gabriele'a del Grandego, który monitoruje również zgony i zaginięcia migrantów w Europie, a także *Puls* – projekt prowadzony przez Uniwersytet Helsiński w Finlandii i Wspólne Centrum Badawcze Komisji Europejskiej.

Do analiz wszystkich zewidencjonowanych w bazie danych stosowana jest metodologia oparta na programie OSINT, stworzonym przez służby wywiadowcze, który zbiera dane z publicznie dostępnych źródeł takich jak raporty medialne, publikacje rządowe, a dodatkowo analizuje treści milionów użytkowników sieci, którzy bardzo często okazują się naoczniymi świadkami wydarzeń. Liczba migrantów ponoszących śmierć w trakcie poszukiwania schronienia uzyskiwana jest z globalnych informacji pozyskiwanych w czasie rzeczywistym (*The Migrants' Files*).

Wizualizacja informacji

Przykład *The Migrants' Files* pokazuje, jak ważne w epoce nadmiaru informacji jest wejście pod powierzchnię nieuporządkowanego strumienia treści i poszukiwanie w nich wzajemnych zależności, które pozwolą odkryć niedostrzegane wcześniej zjawiska i problemy. Istotne staje się także pozyskiwanie cennych informacji ze źródeł nieustrukturyzowanych, np. portali społecznościowych, blogów itp. Oprócz bazowania na zewidencjonowanych źródłach charakterystyczne dla dziennikarstwa danych jest przedstawianie informacji przy zminimalizowaniu języka pisanego na rzecz języka wizualnego (Piekarski 2017, 207). Istotą dziennikarstwa danych jest zatem synteza słowa pisanego, języka komputerowego oraz obrazu. Obrazy, którymi operują dziennikarze danych, nie odwzorowują bezpośrednio rzeczywistości, ale przynależą do świata obrazów technicznych (Piekarski 2017, 211). Nie są owocem prac fotoreporterów, lecz wytworem interdyscyplinarnych zespołów składających się z dziennikarzy, analityków, programistów i grafików komputerowych.

Otoczający nas świat jest niezwykle złożony, a w warunkach „eksplozji informatycznej i przeciążenia naszego układu zmysłów konieczne jest lepsze przyswajanie, organizowanie i wykorzystywanie płynących z zewnątrz sygnałów” (Vetulani 1998, 37). Ewolucyjnie nasz system zmysłów oraz układ przetwarzania danych są przystosowane do takiego środowiska, które znamy z zeszłego wieku. Nasza psychika „z trudem daje sobie radę z obecnym tempem życia” (Vetulani 1998, 38)

i nadmiarem otaczających nas informacji. Ich ogrom oraz dostępność sprawiają, że „umysł ulega stopniowej degradacji, gdyż w sytuacji braku mobilizujących go do pracy bodźców zaczyna funkcjonować mniej sprawnie i zanikać” (Szpunar 2015, 152). Przekazy internetowe są informatycznym konstruktem, złożonym z linków i tekstów, których hipertekstualna natura sprawia, że należą one do różnych zbiorów lub części zbiorów dających się umieścić i otwierać w oknach o różnych parametrach (Vanderdorpe 2008, 101). Nadmiar tekstów, odsyłaczy i innych elementów w ramach jednego przekazu potęguje stan przeciążenia informacyjnego, sprawiając, że nasz odbiór informacji oraz jej przetworzenie i zapamiętanie są mocno utrudnione.

Przerost wszelkiego rodzaju treści dominuje nad możliwościami ich zapamiętywania i przetwarzania. Spowodowane jest to ograniczeniami mózgu w percepcji świata zewnętrznego. Zdaniem Karola Piekarskiego (2017), aby można było zrozumieć, czym jest percepcja, należy najpierw odróżnić ją od uwagi. Autor *Kultury danych* definiuje uwagę jako zdolność skupienia się na wybranych elementach otoczenia i sądzi, że może być ona traktowana jako wewnętrzna cecha jednostki. Percepcję uznaje natomiast za całościowy proces poznawczy człowieka (spozrzeganie). Łączy ją z postrzeganiem, z selekcją informacji z otoczenia, która powinna odbywać się w sposób automatyczny. Obecnie okazuje się, że proces automatycznego odbioru rzeczywistości zostaje zaburzony na skutek przeciążenia informacyjnego. Powszechny dostęp do internetu sprawił, że sieć stała się jednym z głównych źródeł informacji. Treści w niej zamieszczane są jednak słabo uporządkowane, bywa, że niewłaściwie sklasyfikowane, a do tego występują w wielu formatach. Przez „komercyjnie zaprogramowane” wyszukiwarki i algorytmy poszukiwana informacja staje się często niemożliwa do odnalezienia. Wszystko to sprawia, że proces pozyskiwania informacji z sieci staje się żmudny, pracochłonny i niedający satysfakcji. Dodatkowo wszechobecna superata treści, górując nad możliwościami ludzkiego mózgu, powoduje, że *homo medialis* doświadcza szoku informacyjnego. Każdego dnia człowiek spotyka się z bogactwem przeróżnych treści, równocześnie doznając ubóstwa uwagi. Aby skutecznie radzić sobie z tą „społeczną” bolączką i przezwyciężyć ograniczenia percepcyjne, należało wypracować taki sposób przekazu informacji, który będzie adekwatny do obecnej sytuacji i ułatwi współczesnemu człowiekowi dostęp do jasnych i atrakcyjnych treści, niewywołujących lęku informacyjnego. Przedstawianie treści w postaci informacji zwizualizowanych przezwycięża ograniczenia percepcyjne naszego mózgu. Komputer oraz urządzenia typu *smart* zdają się działać na korzyść wizualności. Dziennikarze danych, tworząc swoje teksty przy wykorzystaniu technik wizualnych, przekazują duże ilości informacji (które w słowie pisanym zajęłyby klika stronnic) w postaci dwu- lub trójwymiarowych obrazów i wykresów. Ludzkie predyspozycje do lepszego odbioru obrazu aniżeli tekstu uwarunkowane są po-

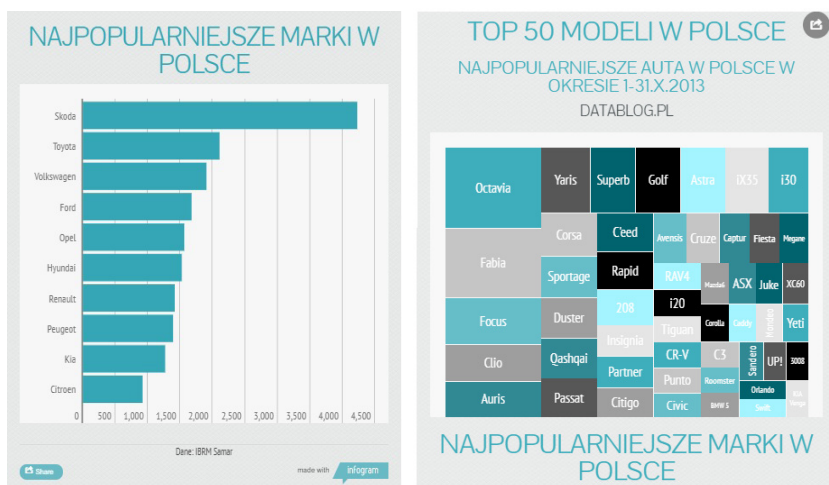
znawczymi i emocjonalnymi predyspozycjami człowieka. Obraz jest bezpośrednio interpretowany przez najstarszą część mózgu, czyli przez układ limbiczny. Przemawia do naszej podświadomości, tworząc wrażenia emocjonalne, zanim jeszcze zdołamy go logicznie przeanalizować. W sferze poznawczej zwizualizowane dane stają się przynętą na współczesnego człowieka, dla którego liczy się szybki dostęp do czystej informacji bez włożenia dodatkowego wysiłku w śledzenie/czytanie większej ilości tekstu. Krakowska uczona Magdalena Szpunar omawiając ludzką skłonność do osiągnięcia celu przy jak najmniejszym wysiłku, określa ją jako zasadę minimaksu (Szpunar 2015, 151).

Syntetyczny charakter obrazu pozwala na przedstawienie konkretnych zjawisk w sposób spójny i całościowy, a także daje możliwość zauważenia pewnej zależności, którą autor chciał w nim przekazać. Prosty rysunek może zatem zastąpić obszerną ilość tekstu, dając czytelnikowi możliwość skupienia się na potrzebnych mu informacjach. Na jednej stronie internetowej mogą nakładać się na siebie elementy tekstowe i wizualne, co modyfikuje ekonomikę tekstu. Obrazowi pozo-
stawiona zostaje strona opisowa i referencyjna, zaś tekst służy wyjaśnianiu elementów abstrakcyjnych lub związków pomiędzy komunikatami. Czytelnik, serfując po tropach ilustracji, „miętko przechodzi” przez proces eksploracji i dokonuje tym samym osobistej, emocjonalnej syntezy detali werbalnych i ikonicznych (Vanderdorpe 2008, 135–138).

Dziennikarstwo danych w Polsce

Przekazy informacyjne bazujące na danych liczbowych, przedstawiane za pomocą wizualizacji, stawiają pierwsze kroki na polskim gruncie dziennikarskim. Z całą pewnością za prekursorów tego innowacyjnego typu „narracji” można uznać twórców portalu informacyjnego Datablog. Jego powstanie dało początek dziennikarstwu danych w Polsce.

Wstępnie Datablog funkcjonował jako jednoosobowy blog prowadzony przez Piotra Kozłowskiego, dziennikarza oraz ówczesnego specjalisty do spraw komunikacji w Fundacji Media 3.0. W 2015 roku portal otrzymał wsparcie Fundacji Batorego, dzięki czemu nastąpił jego dynamiczny rozwój. W jego tworzenie zaangażowane zostały osoby reprezentujące różne dziedziny, a których wspólna praca pozwoliła na utworzenie sprawnie działającego zespołu dziennikarzy danych. Pierwszy artykuł powstały w oparciu o dane, zatytułowany *Najpopularniejsze samochody w Polsce* został udostępniony czytelnikom 7 listopada 2013 roku. Nie ma on szczególnych walorów graficznych, autor zastosował w nim nieskomplikowane diagramy, utrzymane w zimnej kolorystyce. Prostota zastosowanej infografiki pozwala na szybkie odczytanie najważniejszych informacji. Dane wykorzystane przy tworzeniu przekazu pozyskane zostały z serwisu IBRM Samar (ilustracje 1 i 2).



Ilustracja 1 i 2. Pierwsze infografiki pochodzące z portalu Datablog (Kozłowski 2013)

W kolejnych latach działalności portalu zamieszczane w nim infografiki stały się o wiele bardziej barwne i zaawansowane. W artykule Beaty Matuszek pt. *Jak Polacy wspierają OPP 1% podatku* autorka oprócz infografik zamieszcza także obraz przedstawiający dużą grupę ludzi oraz leksykalny opis danych przedstawionych graficznie (ilustracja 3).

Kwota 1% podatku dochodowego przekazywana przez podatników rośnie niezmiennie od kilku lat. W 2015 roku do organizacji pożytku publicznego trafiło 557,6 mln zł, a wsparcie uzyskało aż 7888 organizacji. Ciągle jeszcze najchętniej Polacy wspierają największe organizacje pozarządowe, które najczęściej prowadzą szeroko zakrojone kampanie medialne.

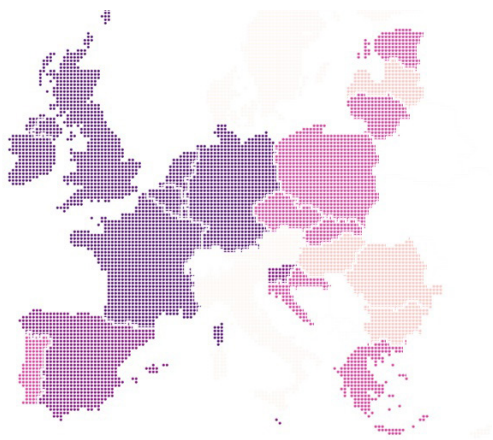
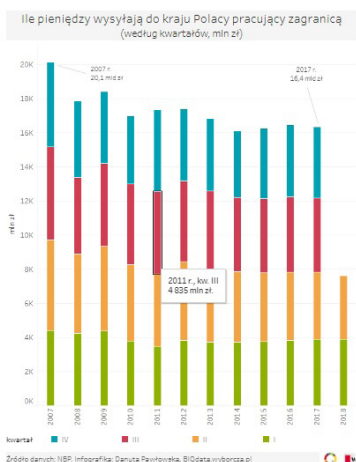
Najwięcej środków z 1% na swoje działania otrzymały w ubiegłym roku Fundacja Pomocy Dzieciom „Zdążyć z pomocą” (136189516,87 zł), Fundacja Pomocy Osobom Niepełnosprawnym „Stoneczko” (20716552,76 zł) oraz Fundacja Avalon (15791702,80 zł).



Ilustracja 3. Przykład infografik z serwisu BIQdata (Matuszek 2016)

Niedługo po uruchomieniu Databloga Piotr Kozłowski rozpoczął pracę nad projektem BIQdata, będącym jednym z serwisów związanych bezpośrednio z „Gazetą Wyborczą”. Zamieszczane na portalu przekazy są chętnie czytane oraz często

komentowane przez internautów. Powstają jako efekt korelacji danych pochodzących z różnych źródeł. Zastosowane infografiki są multimedialne, po najechaniu na nie kursorem można podglądać skrywane dane (ilustracja 4).



Ilustracja 4. Przykład infografik wygenerowanych na łamach BIQdata (Pawłowska 2019)

Jeden z wiodących polskich internetowych serwisów informacyjnych – onet.pl także wychodzi naprzeciw oczekiwaniom swoich czytelników. Pomędzy „tradycyjnymi” reportażami zaczynają pojawiać się przekazy powstałe w oparciu o dane liczbowe i przedstawione w sposób graficzny (ilustracja 5).



Ilustracja 5. Dziennikarstwo danych na łamach portalu onet.pl (Co oznaczają więzienne tatuaże 2019)

Zakończenie

Czas biegnie nieubłaganie, a wraz z nim zmienia się cała rzeczywistość społeczna. Nieograniczony dostęp do informacji dał „kluczyk do wiedzy”, ale nie wskazał sposobu na jej otwarcie. Umiejętnie zaprogramowane wyszukiwarki nie rzadko kierują człowieka do treści, których wcale nie oczekiwał. W „cyfrowym buszu” coraz trudniej znaleźć mu to, czego naprawdę szuka. Lawina przeróżnych treści potęguje przeciążenie informacyjne, które jeszcze bardziej komplikuje poszukiwania. Ogromne zbiory *big data* wykorzystywane przez współczesnych dziennikarzy, umiejętnie przez nich przeanalizowane i przetworzone, dają dostęp do niedostrzegalnych bezpośrednio treści. Dziennikarstwo danych jawi się zatem jako przyszłość dziennikarstwa internetowego, ponieważ swoją formą idealnie wpisuje się w strukturę sieci. Staje się pomostem łączącym to, co ścisłe, cyfrowe z tym, co narracyjne, humanistyczne. Krótkość i zwięzłość informacji prezentowanych przez dziennikarzy danych oraz wizualizacja treści wychodzą naprzeciw potrzebom współczesnego człowieka. Oferują ciekawe wiadomości, opatrzone dodatkowo atrakcyjną infografiką, która przewyższa ograniczenia percepcyjne, pozwalając czytelnikowi na dokładne przetworzenie przekazu. Czytelnik nie ma możliwości nadinterpretacji infografiki. Może dekodować ją jedynie zgodnie z założeniem jej twórcy.

Bibliografia

- Brosz Maciej, Bryda Grzegorz, Siuda Piotr (2017), *Od redaktorów: Big Data i CAQDAS a procedury badawcze w polu socjologii jakościowej*, „Przegląd Socjologii Jakościowej”, nr 2, s. 6–23, [online:] http://www.qualitativesociologyreview.org/PL/Volume38/PSJ_13_2_Brosz_Bryda_Siuda.pdf [dostęp: 11.11.2018].
- Co oznaczają więzienne tatuaże? [infografika]* (2019), [w:] *Onet Facet*, [online:] <https://facet.onet.pl/warto-wiedziec/co-oznacza-wiezienne-tatuaże-infografika> [dostęp: 25.02.2019].
- Gogołek Włodzimierz (2012), *Nowy wymiar zasobów informacyjnych www*, [online:] http://www.ktime.up.krakow.pl/symp2012/referaty_2012_10/gogolek.pdf [dostęp: 23.03.2019].
- Graczyk-Kucharska Magdalena (2015), *Big data koniecznością współczesnego marketingu*, „Zeszyty Naukowe Uniwersytetu Szczecińskiego. Problemy Zarządzania, Finansów i Marketingu”, nr 41, t. 2, DOI: 10.18276/pzfm.2015.41/2-22, s. 265–277, [online:] <http://www.wzieu.pl/zn/875/22Graczyk-Kucharska.pdf> [dostęp: 19.10.2018].
- Kozłowski Piotr (2013), *Najpopularniejsze samochody w Polsce*, [online:] <http://datablog.pl/infografika-najpopularniejsze-samochody-w-polsce/> [dostęp: 23.09.2018].

- Leetaru Kalev Hannes (2011), *Monitorujemy media aby przewidywać przyszłość (Culturomics 2.0)*, rozm. przepr. Barbara Jakubek, [online:] <http://www.mediafun.pl/monitorujemy-media-aby-przewidywac-przyszlosc-culturomics-2-0/> [dostęp: 19.10.2018].
- Mach-Król Maria (2015), *Analiza i strategia big data w organizacjach*, „Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą”, t. 74, s. 41–53, [online:] http://www.pszw.edu.pl/images/publikacje/t074_pszw_2015_mach-krol-analiza-i-strategia-big-data-w-organizacjach.pdf [dostęp: 19.10.2018].
- Matuszek Beata (2016), *Jak Polacy wspierają OPP 1% podatku?*, [online:] <http://datablog.pl/jak-polacy-wspieraja-opp-1-podatku/> [dostęp: 23.09.2018].
- Mayer-Schöberger Viktor, Cukier Kenneth (2014), *Big data. Rewolucja, która zmieni nasze myślenie, pracę i życie*, przeł. Michał Głatki, Warszawa.
- Pawłowska Danuta (2019), *Polacy mieli wracać z emigracji, ale wyjeżdża ich coraz więcej. Ile pieniędzy wysyłają do Polski?*, [online:] <http://biqdata.wyborcza.pl/biqdata/7,159116,24452247,polacy-na-emigracji.html> [dostęp: 18.02.2019].
- Piekarski Karol (2017), *Kultura danych. Algorytmy wzmacniające uwagę*, Gdańsk.
- Polańska Krystyna, Wassilew Aleksander (2015), *Analizy big data w serwisach społecznościowych*, „Nierówności Społeczne a Wzrost Gospodarczy”, nr 4 (44), cz. II, DOI: 10.15584/nsawg.2015.4.2.11, s. 117–128, [online:] <https://repozytorium.ur.edu.pl/bitstream/handle/item/1555/2.11%20pola%5c5%84ska-analizy.pdf> [dostęp: 19.10.2018].
- Szpunar Magdalena (2015), *Internet i jego wpływ na procesy pamięciowe*, „Teraźniejszość – Człowiek – Edukacja” t. 18, nr 2 (70), s. 149–156, [online:] http://www.magdalenaszpunar.com/_publikacje/2015/2015_70_10.pdf [dostęp: 19.10.2018].
- Szpunar Magdalena (2016), *Humanistyka cyfrowa a socjologia cyfrowa. Nowy paradygmat badań naukowych*, „Zarządzanie w Kulturze”, t. 17, z. 4, s. 355–369, [online:] http://www.magdalenaszpunar.com/_publikacje/2016/ZwK_17_4_355-369.pdf [dostęp: 19.10.2018].
- Tabakow Marta, Korczak Jerzy, Franczyk Bogdan (2014), *Big data – definicje, wyzwania i technologie informatyczne*, „Informatyka Ekonomiczna”, nr 1 (31), DOI: 10.15611/ie.2014.1.12, s. 138–153.
- The Migrants' Files*, [online:] <http://www.themigrantsfiles.com/> [dostęp: 20.10.2018].
- Vanderdorpe Christian (2008), *Od papirusu do hipertekstu. Esej o przemianach tekstu i lektury*, przeł. Anna Sawisz, Warszawa.
- Vetulani Jerzy (1998), *Jak usprawnić pamięć. Przewodnik po najnowszych teoriach pamięci i osiągnięciach farmakologii na polu walki z niedostatkami naszych czynności intelektualnych*, wyd. 3, Kraków.

Summary

Big Databases as New Sources of Information. Are “Big Data” the Future of Journalism?

The essence of any device coupled to an IT network is the collection, processing and transmission of data. Along with technological progress, the amount of information and, above all, the possibilities of their collection and interpretation allowed exploration of the big data in the fields of science, which until recently were outside the margin of the so-called ‘data revolution’. Big data sets were initially used only by financial institutions. On their basis, mathematical economic and economic forecasts were determined. Nowadays, when the humanities are in the heat of criticism and the science is leading the science, the humanities begin to defend themselves and, like the exact sciences, begin to use new technologies. Within the humanities, data play a large role, which, if skilfully analyzed, allows predicting social phenomena. Data correlated with each other by data journalists become the beginning of new, unknown stories. The author in her text shows the role of data in the process of creating journalistic messages. The article not only presents data as a remedy for information overload, but also points to data journalism as the future in acquiring and generating information.

Keywords: big data; data journalism; visualization; information

Streszczenie

Istotą każdego urządzenia sprzęgniętego w sieć informatyczną jest gromadzenie, przetwarzanie oraz przesyłanie danych. Wraz z postępem technologicznym ilość informacji, a przede wszystkim możliwości ich gromadzenia oraz interpretacji, pozwoliły na eksploatację tak zwanego big data w tych dziedzinach nauki, które do niedawna znajdowały się poza marginesem tzw. rewolucji danych. Zbiory *big data* początkowo wykorzystywane były jedynie przez instytucje finansowe. Na ich podstawie wyznaczano matematyczne prognozy ekonomiczno-gospodarcze. Obecnie, kiedy humanistyka stoi w ogniu krytyki, a prym w nauce wiodą nauki ścisłe, humanistyka zaczyna bronić się i – podobnie jak nauki ścisłe – korzystać z nowych technologii. W obrębie nauk humanistycznych dużą rolę zaczynają odgrywać dane, które umiejętnie przeanalizowane, pozwalają przewidywać zjawiska społeczne. Dane skorelowane ze sobą przez dziennikarzy danych stają się początkiem nowych, nieznanych historii. Autorka w swoim tekście wykazuje funkcje danych w procesie tworzenia przekazów dziennikarskich. Przedstawia je jako remedium na przeciążenie informacyjne. Upatruje w dziennikarstwie danych przyszłości w pozyskiwaniu i generowaniu informacji.

Słowa kluczowe: big data; dziennikarstwo danych; wizualizacja; informacja