

Komunikacja i Media



Przemysław Szews

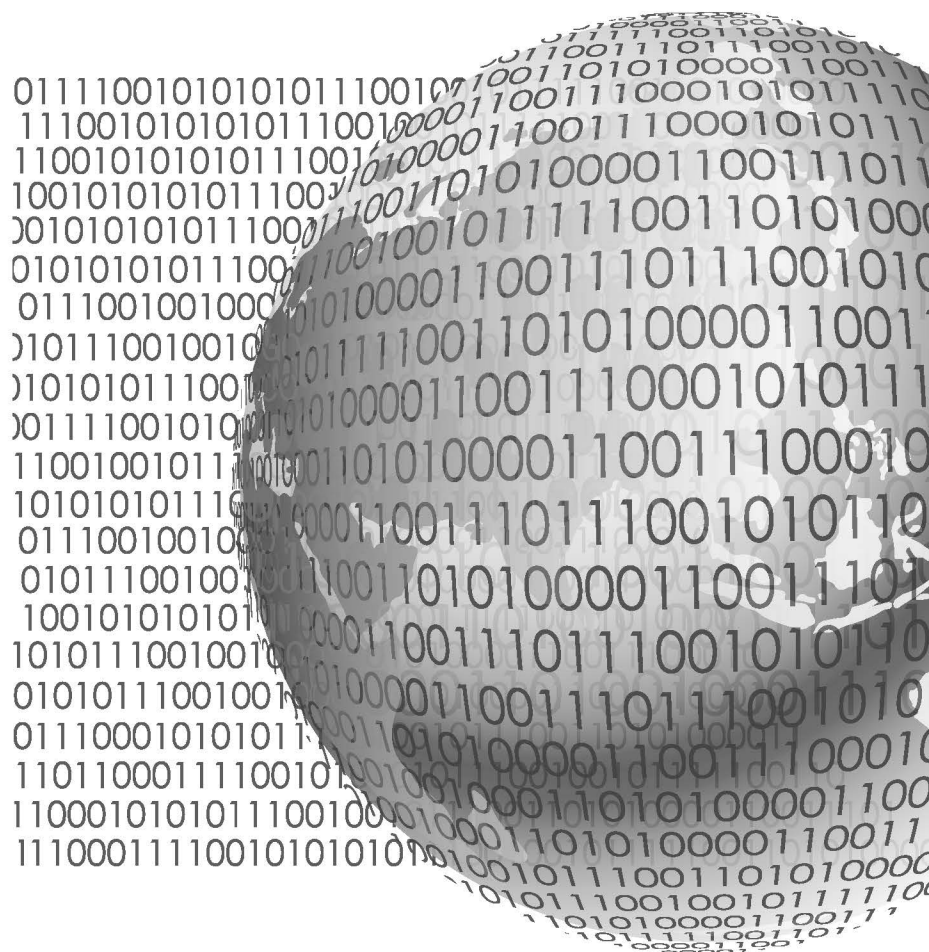
Wizualizacja i dziennikarstwo danych w mediach

Wizualizacja i dziennikarstwo danych w mediach



WYDAWNICTWO
UNIwersytetu
ŁÓDZKIEGO

Komunikacja i Media



Przemysław Szews

Wizualizacja i dziennikarstwo danych w mediach

**WYDAWNICTWO
UNIwersytetu
ŁÓDZKIEGO**
Łódź 2021

Przemysław Szews – Uniwersytet Łódzki, Wydział Filologiczny
Katedra Dziennikarstwa i Komunikacji Społecznej
90-236 Łódź, ul. Pomorska 171/173

RECENZENT
Andrzej Adamski

REDAKTOR INICJUJĄCY
Urszula Dzieciatkowska

OPRACOWANIE REDAKCYJNE
Katarzyna Soszyńska

SKŁAD I ŁAMANIE
AGENT PR

KOREKTA TECHNICZNA
Wojciech Grzegorzcyk

PROJEKT OKŁADKI
krzysztof de mianiuł

Zdjęcie wykorzystane na okładce: © Depositphotos.com/GoodLuckWithUs

© Copyright by Przemysław Szews, Łódź 2021

© Copyright for this edition by Uniwersytet Łódzki, Łódź 2021

Publikacja jest udostępniona na licencji Creative Commons

Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 4.0 (CC BY-NC-ND)

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego

Wydanie I. W.08966.18.0.M

Ark. wyd. 5,6; ark. druk. 7,125

ISBN 978-83-8220-489-6

e-ISBN 978-83-8220-490-2

<https://doi.org/10.18778/8220-490-2>

Wydawnictwo Uniwersytetu Łódzkiego

90-131 Łódź, ul. Lindleya 8

www.wydawnictwo.uni.lodz.pl

e-mail: ksiegarnia@uni.lodz.pl

tel. 42 665 58 63

SPIS TREŚCI

Wstęp	7
-------------	---

ROZDZIAŁ 1

1.1. Definicja danych	11
1.1.1. Dane jako nowe możliwości dla dziennikarza	12
1.1.2. Dane jako część nowych mediów	14
1.2. Wizualizacja danych	15
1.3. Grafika statyczna	17
1.4. Grafika interaktywna	19
1.5. Problemy percepcji danych	20
1.6. Wizualizacja a manipulacja	25
1.6.1. Manipulacja wizualizacją danych – studium przypadku	29
1.6.2. Wybrane sposoby przekłamań na wykresach i grafikach informacyjnych	39
1.7. Zasady poprawnego konstruowania wykresów i prezentacji danych	53
1.8. Funkcja koloru w wizualizacji danych	54

ROZDZIAŁ 2

2.1. Liczby	59
2.2. Bazy danych	62
2.3. Dziennikarstwo danych – definicja pojęcia	66
2.4. Dziennikarstwo danych w Polsce	69
2.5. Dziennikarstwo danych jako część nowych mediów	70
2.6. <i>Computer-Assisted Reporting</i>	75
2.6.1. <i>Computer-Assisted Reporting</i> a dziennikarstwo precyzyjne	75
2.6.2. Związki między dziennikarstwem danych i CAR	76
2.6.3. Przyszłość <i>Computer-Assisted Reporting</i>	77
2.7. Algorytmizacja danych i <i>big data</i>	79
2.8. Źródła informacji w dziennikarstwie danych	81
2.8.1. Systematyka źródeł informacji w dziennikarstwie danych	86
2.8.2. Białe wywiad i źródła <i>open source</i>	87
2.8.3. Otwarte źródła informacji	88
2.8.4. Analiza wybranych portali danych jako przykładów źródeł informacji	90

2.9. <i>Data blog</i> i <i>data blogging</i>	95
2.9.1. Definicja blogu jako tło teoretyczne	95
2.9.2. <i>Data blog</i> – definicja pojęcia	100
2.10. Dane jako źródło rozrywki	102
2.10.1. Infotainment jako punkt wyjścia	102
2.10.2. Datatainment	103
Zakończenie	107
Bibliografia	109

WSTĘP

Nowe media, przyzwyczajenia odbiorców, tempo komunikacji prowadzą do popularyzacji procesów sprzyjających jej skracaniu, dążeniu do efemeryczności, natychmiastowości przekazu. Wpisują się w to zwłaszcza media społecznościowe, w zakresie obecności zarówno mikrokomunikatów tekstowych, jak i memów, demotywatorów lub tzw. *stories*, czyli komunikatów opartych na wizualności. Trend ten występuje również w dziennikarstwie – nie tylko internetowym, w którym wizualizacja informacji była obecna od dawna (teraz wydaje się mieć jeszcze większe znaczenie niż kiedyś), ale również prasowym i telewizyjnym.

Szczególnie chętnie wizualizowane są informacje oparte na danych liczbowych. Pomaga to w efektywnym odwołaniu się do wyobraźni odbiorcy i maksymalizuje skuteczność przekazu. Informacje bazujące na samych liczbach potrafią być skomplikowane i nieprzejrzyste dla przeciętnego odbiorcy mediów masowych. Dlatego chętnie sięga się po wizualizacje w postaci wykresów, diagramów, kolumn – zarówno w prasie, jak i w Internecie czy telewizji (gdzie coraz częściej stanowi to wręcz nieodłączną część serwisów informacyjnych, wprowadzając – poza informacyjną – wartość estetyczną).

Pierwsza część książki poświęcona jest procesowi wizualizacji danych, a także właściwej percepcji grafik obecnych w mediach. Mimo wszechobecności operowania obrazem oraz używania piktogramów, map czy infografik w niemal każdej dziedzinie, w różnych mediach i na różnych poziomach, odbiorcy takich komunikatów nie zawsze posiadają odpowiednie kompetencje do ich prawidłowego odczytania. Wyjaśniona zostaje istota rozumienia i poznawania zależności opisywanych przez dane przedstawione w wizualizacjach. Podobnie jak lektura tekstu przebiega w jego warstwie podstawowej i ukrytej (interpretacyjnej), graficzna prezentacja informacji często jest dwuwarstwowa, wielopoziomowa, pozostawiająca pole do dalszego poznawania w miarę łączenia poszczególnych części.

Punkt wyjścia w pierwszym rozdziale stanowi sama definicja danych, które są pojęciem złożonym i wbrew pozorom nie obejmują tylko liczb. W obecnej rzeczywistości, gdzie dostęp do informacji upowszechnił się dzięki Internetowi – dostarczającemu nie tylko dane, ale również sposoby ich agregowania, katalogowania i przetwarzania – dane stwarzają nowe możliwości dla mediów, zarówno tych tradycyjnych, jak i internetowych. W obu przypadkach

publikowane są nie same liczby, a skutecznie trafiające do odbiorcy ich zestawienia w formie różnego rodzaju wizualizacji (statycznych i dynamicznych, w zależności od medium).

Wizualizacja danych definiowana jest jako metoda prezentowania informacji, uwzględniająca badania dotyczące percepcji i sposobu odbioru komunikatów przez człowieka, ale również jako zestaw technik umożliwiających przedstawianie danych w sposób klarowny, uporządkowany i efektowny. Popularyzacja i rozwój wizualizacji informacyjnej wiąże się bezpośrednio z jej nagromadzeniem w mediach i łatwym dostępem, możliwymi także dzięki mediom społecznościowym. Informacyjny przesyt obecny w mediach, a szczególnie w Internecie, powoduje, że powstaje zapotrzebowanie na przedstawienie danych w formie skondensowanej i zrozumiałej dla odbiorcy. By wskazać trendy i zjawiska niezauważalne przy pobieżnej lekturze, przedstawiono grafiki zróżnicowane pod względem rozmiaru, charakteru i pochodzenia. Ważnym elementem rozdziału jest zwrócenie uwagi na problem manipulacji obrazem wraz ze wskazaniem jej wybranych sposobów na przykładach wykresów i grafik informacyjnych. W tej części rozważań przywołane zostało studium przypadku, aby ukazać te nieprawidłowości na przykładzie istniejących, opublikowanych w mediach wizualizacji.

Nieocenioną pomocą w powstaniu tej książki, szczególnie jej pierwszej części, była publikacja Przemysława Biecka *Odkrywać! Ujawniać! Objaśniać! Zbiór esejów o sztuce prezentowania danych*¹, która stanowi prawdziwe kompendium wiedzy nt. prezentowania danych, zarówno w zakresie sposobów najbardziej efektywnych, jak i takich, które noszą znamiona manipulacji i przekłamania. W pierwszym rozdziale wiele przykładów i odwołań definicyjnych zostało przywołanych właśnie za autorem powyższego tytułu.

Celem drugiej części książki jest zdefiniowanie *data journalism* – specjalizacji dziennikarskiej, która na polskim rynku medialnym dopiero debiutuje, co wiąże się z ograniczoną liczbą materiałów do analizy, a także skromnym zakresem naukowych dociekań. Nieopracowana dotąd na gruncie polskim, lokuje się na pograniczu różnych mediów i typów dziennikarstwa. W obrębie dziennikarstwa danych często dochodzi do wzajemnego współdziałania prasy i Internetu, jednak specyfika procesu tworzenia i przekazu informacji w *data journalism* jest właściwa tylko tej specjalizacji.

Prezentowana w książce analiza obejmuje nie tylko sam konstrukt i sposób realizacji tej metody dziennikarstwa, ale również procesy towarzyszące: zbieranie informacji, wizualizowanie, współpracę między poszczególnymi działami

¹ P. Biecek, *Odkrywać! Ujawniać! Objaśniać! Zbiór esejów o sztuce prezentowania danych*, Warszawa 2014.

redakcji. Wyszczególnić należy też próbę odpowiedzi na pytanie, jakie nowe umiejętności i kompetencje powinien mieć dziennikarz reprezentujący tę odmianę, a także jak może się układać współpraca w już istniejących działach redakcji, by dziennikarstwo danych mogło się rozwijać.

W zdefiniowaniu omawianej specjalizacji, będącej do niedawna zjawiskiem *in statu nascendi*, pomogły zagraniczne publikacje i realizacje. Jako punkt wyjścia do jej opisu i charakterystyki użyto pojęcia liczb, używanych w komunikacji społecznej, wraz z ich oddziaływaniem i funkcjami. Za kolejne teoretyczne umocowanie obrano bazy danych (dziennikarstwo danych było tłumaczone początkowo jako dziennikarstwo „bazodanowe”), które należy pojmować szerzej aniżeli jako jedno z dziennikarskich źródeł informacji. Paradigmat rewolucji komunikacyjnej, mający podłoże w kodzie cyfrowym, stanowi podstawę nowej, nielinarnej formy komunikacji i całkowicie innego sposobu narracji, którymi odróżniają się nowe media, czerpiące często z baz danych, od mediów tradycyjnych. W rozdziale opisano odmienną narrację typową dla mediów tradycyjnych od nienarracyjnej bazy danych, charakterystycznej dla nowych mediów.

Zaczerpnięte z istniejących opracowań definicje dziennikarstwa danych stały się fundamentem dla jego charakterystyki jako specjalizacji reprezentującej nowe media. Rozdział zawiera ponadto rys historyczny zagadnienia oraz wskazuje na związki z innymi technikami zbierania i analizowania danych – czy to jako sposobu rozszerzania kontekstu, czy opowiadania całych historii.

Druga część książki porusza również ważne dla jej tematyki terminy, jak algorytmizacja danych i *big data*, zwracając uwagę na ich istotność w rozwoju specjalizacji dziennikarstwa danych, a także na zmiany podejścia do danych i informacji przedstawianych w sieci. Nie mogło też zabraknąć wskazania głównych źródeł informacji dla dziennikarstwa danych, które zostały scharakteryzowane i uporządkowane według rodzajów i odmian, ze szczególnym zaznaczeniem informacji pochodzących z otwartych źródeł. Kontekst dla źródeł informacji stanowi zagadnienie białego wywiadu, wspomnianego krótko w tej części.

Celem książki jest omówienie procesów wizualizacji danych w mediach, łączących się ściśle z dynamicznie rozwijającą się dziennikarską specjalizacją *data journalism*. Zbiór dostępnych informacji dotyczących obu tych pojęć stanowi punkt wyjścia do dalszych szczegółowych badań nad wizualnością mediów, ale również nad samym dziennikarstwem danych, które jako reprezentant tzw. dziennikarstwa jakościowego może się rozwijać wraz ze swoimi odmianami i ewolucjami (np. datatainmentem) na wyspecjalizowanych serwisach internetowych lub w dedykowanych sobie mediach.

ROZDZIAŁ 1

1.1. Definicja danych

Próbie definicji samych danych podjęto podczas kursu przygotowanego przez School of Data, członka projektu Open Knowledge Foundation. Ustalono wtedy, że dane są wartościami, pojmowanymi w szerokim znaczeniu, przypisanymi do jakiejś rzeczy¹, którą należy usytuować w odpowiedniej sferze zależności. Jako że nie wszystkie przedmioty, osoby czy obiekty można przypisać do tych samych grup, dane dzielą się na dwie główne kategorie: dane jakościowe i ilościowe. Dane jakościowe odnoszą się np. do doświadczeń, tekstur, kolorów, natomiast dane ilościowe opierają się w głównej mierze na liczbach (takich jak cena, wielkość, odległość).

Praca z danymi pozwala wyodrębnić inne kategorie: dyskretne, kateryczne i ciągłe. „Dane dyskretne to dane numeryczne, w których ciągu istnieją pewne luki, puste przestrzenie”². Luki nie występują w kategorii danych ciągłych. Z kolei dane kateryczne służą przypisaniu obiektu do konkretnej kategorii (np. samochód przypisany może być do kategorii „nowe” lub „używane”) i informowaniu o rodzaju przedmiotu.

Kolejnym podziałem, oddającym ideę *data journalism*, jest rozróżnienie danych na nieustrukturyzowane i ustrukturyzowane. Należy pamiętać, że przy tego rodzaju dziennikarstwie mamy do czynienia ze ścisłym związkiem człowieka, zbierającego informacje, i komputera, przetwarzającego je w związki i zależności. Podział na ustrukturyzowane i nieustrukturyzowane dane implikuje kolejne etapy pracy dziennikarza i maszyny. Pierwszy etap dotyczy danych nieustrukturyzowanych, odnoszących się do poszczególnych wartości. W trakcie szkolenia School of Data posłużono się następującym przykładem: „Mamy 5 białych, używanych pilek golfowych o średnicy 43 mm w cenie 0,5 \$ za sztukę”³. To zdanie

¹ Por. P. Kozłowski, *Kurs dziennikarstwa danych #1: Czym są dane?*, <http://datablog.pl/kurs-dziennikarstwa-danych-1-czym-sa-dane/> (dostęp: 29.08.2019).

² Tamże.

³ Tamże.

jest łatwe do zrozumienia dla człowieka, ale nieinterpretowalne dla programów komputerowych. Czynniki ludzkie okazują się więc niezbędne do interpretacji danych, ich zrozumienia, a następnie przetworzenia w formę zrozumiałą dla komputera. Taki proces, często jednorazowy, polega na automatyzacji przetwarzania danych przez program komputerowy, do czego niezbędne są dane ustrukturyzowane. Mają one najczęściej postać tabel i wartości oddzielonych przecinkami (tzw. CSV, tj. *comma separated values*). Przedstawione powyżej zdanie po ustrukturyzowaniu i przygotowaniu do przetworzenia i przeanalizowania przez komputer oraz po zastosowaniu CSV wyglądałoby następująco: „quantity”, „color”, „condition”, „item”, „category”, „diameter (mm)”, „price per unit (AUD)” 5, „white”, „used”, „ball”, „golf”, 43, 0.5.

Ustrukturyzowanie i parametryzacja informacji mogą mieć różny zakres. Rzeczywistość można bowiem pokazywać nie tylko za pomocą artykułów prasowych:

seans kinowy lub impreza taneczna mają swoje parametry: miejsce, w którym się odbywają, godzina, styl muzyki, tytuł filmu itp. Można oczywiście napisać artykuł, który opíše, jakie imprezy czekają mieszkańców miasta X w najbliższy weekend, można jednak również załadować wszystkie powyższe parametry do bazy danych, dając użytkownikom rozległe możliwości wyszukiwania według gatunku imprezy, miejsca, czasu, ceny wejściówek⁴.

Według autora sposób przedstawienia takich informacji w formie interaktywnej, sprzężonej np. z bazą połączeń komunikacji miejskiej lub z ogromem innych danych, jest dużo bardziej komfortowy dla odbiorcy od lektury zwykłego gazetowego artykułu.

1.1.1. Dane jako nowe możliwości dla dziennikarza

Nowe możliwości, jakie niosą za sobą dane, w połączeniu z umiejętnością dostrzegania w nich historii i opowiedzenia jej w formie cyfrowej informacji sprawiają, że dziennikarstwo danych ma bardzo duży potencjał i wpisuje się w panujące trendy rzeczywistości medialnej. W obliczu coraz to nowych źródeł przekazów – od mediów tradycyjnych, poprzez blogi, aż do komunikatów w serwisach społecznościowych – kiedy sama informacja może być filtrowana przy użyciu rozmaitych agregatorów treści, oceniana, komentowana i wielokrotnie przetwarzana (a przez to ignorowana), dziennikarze stoją przed swego rodzaju wyzwaniem. Chodzi o znalezienie takich historii, które są trudno dostrzegalne, ale zarazem ciekawe dla odbiorcy, dotyczące np. codzienności, w której żyje.

⁴ L. Olszański, *Media i dziennikarstwo internetowe*, Warszawa 2012, s. 267.

Gromadzenie, filtrowanie i wizualizacja tego, co nas otacza, obok rzeczy widocznych na pierwszy rzut oka, ma coraz większą wartość. Sok pomarańczowy, który pijesz rano, kawa, którą parzysz: w dzisiejszej globalnej gospodarce istnieją niewidzialne połączenia między tymi produktami, innymi ludźmi i Tobą. Językiem tej sieci są dane: małe punkty informacji, które często nie są istotne w pojedynczym przypadku, ale mają ogromne znaczenie, gdy patrzy się na nie pod odpowiednim kątem⁵.

Dostrzeżenie związków, a następnie przefiltrowanie otaczających dziennikarza zjawisk w postać binarną, by łatwiej je przedstawić w szczegółowy, głębszy sposób, jest nową umiejętnością, której pracujący w tym zawodzie muszą się nauczyć. Łączy się ona z prymarną częścią warsztatu dziennikarskiego, czyli opowiadaniem historii. Posługiwanie się danymi przesuwając nacisk z szerokiego pokazywania rzeczywistości na szczegółowe przedstawienie wybranych kwestii, skupienie się na detalach, tłumaczenie ich wpływu na życie odbiorcy.

Jak zostało wskazane wcześniej, danymi, które można wykorzystać w procesie przygotowywania materiału, może być w zasadzie wszystko. Są one sprowadzane do systemu zero-jedynkowego, czyli w pewnym stadium stają się liczbami, by łatwiej można je było przetworzyć, zanalizować i poddać właściwym filtrom. „Ułożone w tabelkach, najczęściej w Excelu, który dziś kojarzy się dziennikarzom raczej ze zwalnianiem ich z pracy niż z samą pracą”⁶. Jonathan Stoneman, szkoleniowiec i konsultant mediów, były reporter BBC, tłumaczy, że bez danych często w ogóle nie ma tekstu i już w tradycyjnie pojmowanym dziennikarstwie wiele newsów opierano na liczbach. Dodaje, że dziennikarstwo polega na znalezieniu tej małej, lecz najważniejszej informacji, ukrytej wśród wielu otrzymywanych danych⁷. Pomocne są w tym oczywiście odpowiednie narzędzia, ale także umiejętność zadawania odpowiednich pytań i obrania klucza podczas selekcji informacji.

Dziennikarstwo danych, zdaniem Leszka Olszańskiego, jest także szansą na rehabilitację mediów internetowych, często wtórnych, podejmujących tematy za tradycyjnymi mediami, cechujących się postępującą tabloidyzacją, niskim poziomem merytorycznym⁸. Badacz ponadto zauważa, że dane dostępne były w pracy dziennikarzy na długo przed pojawieniem się omawianej specjalizacji:

⁵ M. Lorenz, *Why Journalist Should Use Data*, [w:] J. Gray, L. Bounegru, L. Chambers (red.), *The Data Journalism Handbook. How Journalists Can Use Data to Improve the News*, Sebastopol (CA) 2012, s. 3, tłumaczenie własne.

⁶ R. Gluza, *Newsy z Excelsa*, „Press” 2013, nr 09, s. 37.

⁷ Tamże.

⁸ L. Olszański, *Media i dziennikarstwo ...*, s. 266.

Zgodnie z utrwalonym przez dekady stylem pracy typowy dziennikarz czuje się powołany do wykonywania dwóch typów czynności: 1. Zbieranie informacji, 2. Pisanie na ich podstawie artykułów. [...] Nie zauważamy faktu, że mnóstwo informacji, które zbierają dziennikarze, wykonując swe codzienne obowiązki, ma swoją strukturę, która pozwala na ich automatyczne gromadzenie, a następnie cięcie i porcjowanie na najrozmaitsze sposoby. Utopione w grubej kłusce dziennikarskiego tekstu tracą szansę na powtórne wykorzystanie⁹.

W tym miejscu można wymienić takie typy danych wykorzystywanych już wcześniej w mediach, jak repertuary kin, bazy kawiarni, restauracji, obiektów użyteczności publicznej. Olszański wskazuje, że nie brakowało także narzędzi do ich analizy, stosowanych chociażby w działach informacyjnych, w tym ekonomicznych i giełdowych¹⁰.

1.1.2. Dane jako część nowych mediów

Dane to nie tylko zbiór liczb zebranych w arkuszu kalkulacyjnym, choć w przeszłości była to jedyna forma danych, z jaką dziennikarze mieli styczność¹¹. W rzeczywistości cyfrowej, w której niemal wszystko można opisać statystyką, liczbami, systemem zero-jedynkowym, dane stały się czymś więcej niż rozmytymi cyframi w tabelach:

Historia Twojej kariery, 300 000 poufnych dokumentów, wszyscy w kręgu Twoich przyjaciół; to wszystko może być (i jest) opisane tylko dwoma liczbami: zerami i jedynkami. Zdjęcia, wideo i audio są opisane tymi samymi dwiema liczbami: zerami i jedynkami. Morderstwa, choroby, głosowania polityczne, korupcja i kłamstwa: zera i jedynki¹².

Sprowadzające rzeczywistość do systemu binarnego, dziennikarstwo danych może być właściwe dla każdego tematu. Bez wątpienia stanowi nie tylko nową formę dziennikarstwa, ale także ważny element nowych mediów, będących przecięciem dwóch odrębnych procesów – technik obliczeniowych oraz historii technologii medialnych¹³. Lev Manovich jako syntezę tych dwóch zjawisk rozumie przekład języka wszystkich istniejących mediów na język danych numerycznych,

⁹ Tamże, s. 267.

¹⁰ Tamże.

¹¹ Por. R. Gluza, *Newsy z Excela*.

¹² P. Bradshaw, L. Rohumaa, *The Online Journalism Handbook: Skills to Survive and Thrive in the Digital Age*, New York 2011, s. 2, tłumaczenie własne.

¹³ Por. L. Manovich, *Język nowych mediów*, tłum. P. Cypryański, Warszawa 2006, s. 82.

zrozumiały dla komputerów. Autor ten opisuje nowe media za pomocą pięciu pojęć, z czego właściwie wszystkie wpisują się w dziennikarstwo danych pojmowane zarówno jako proces, jak i specjalizacja dziennikarska. Są to: reprezentacja numeryczna, modularność, automatyzacja, wariacyjność, transkodowanie kulturowe¹⁴. Binarność w nowych mediach dostrzega również Jan van Dijk, profesor socjologii i nauk o komunikowaniu, definiujący je jako zintegrowane i interaktywne media oparte na kodzie binarnym¹⁵. Jak dodaje Andrzej Adamski, „trzeba pamiętać, że pojęcie nowe media składa się z dwóch warstw: komputerowej (technicznej) i kulturowej. Do warstwy technicznej możemy zaliczyć proces i pakiet, sortowanie i dopasowywanie, język komputerowy, strukturę danych”¹⁶. Warstwa techniczna okazuje się tu szczególnie ważna, gdyż w dziennikarstwie danych istotne jest zbieranie pakietów danych, ich sortowanie i selekcja, przełożenie na odpowiedni język i nadanie wizualnej struktury.

1.2. Wizualizacja danych

Graficzna prezentacja danych, wyników analiz, statystyk pełni bardzo ważną funkcję zarówno w środowiskach naukowych, jaki i biznesu, przemysłu czy mediów, w których coraz więcej wiadomości opiera się na danych. Jest istotna na każdym etapie rozumienia i poznawania zależności opisywanych przez dane oraz na etapach późniejszych – prezentacji tych zależności odbiorcy. „Czytelna i komunikatywna wizualizacja danych wzmacnia siłę wyników, a nieprzemysłana lub nieczytelna podważa rzetelność przeprowadzonych analiz”¹⁷. Wartość wizualizacji zależy przede wszystkim od odpowiedniego zaplanowania i przedstawienia danych, nie decyduje o niej wyłącznie estetyka. Graficzna prezentacja danych to ich wizualizacja w formie grafiki statycznej, dynamicznej czy infografiki. Najczęściej dwuwarstwowa, pozostawia miejsce na dalsze odkrycia w miarę stopniowego poznawania oraz łączenia składowych w spójną i ciekawą historię, stanowiąc wielowątkową opowieść, w której pierwsze odkrycia odczytane z wykresu są wprowadzeniem do kolejnych¹⁸.

¹⁴ Tamże, s. 92–118.

¹⁵ Por. J. van Dijk, *The Network Society: Social Aspects of New Media*, London 2006, s. 19–21.

¹⁶ A. Adamski, *Media w analogowym i cyfrowym świecie*, Warszawa 2012, s. 280.

¹⁷ P. Biecek, *Odkrywać! Ujawniać! Objaśniać! Zbiór esejów o sztuce prezentowania danych*, Warszawa 2014, s. 4.

¹⁸ Tamże.

Każda wizualizacja danych jest opowiadaniem historii. Odpowiednie zilustrowanie pozwala odbiorcy lepiej tę historię zrozumieć. Początków wizualizacji szukać można już w czasach człowieka pierwotnego, który zamiast przekazu słownego zostawiał malowidła skalne („zanim wynaleziono pismo, zanim wynaleziono cyfry, kilkanaście tysięcy lat przed naszą erą ludzkość nauczyła się przekazywać historię za pomocą malowideł”¹⁹). Choć pismo daje dużą precyzję i pozwala na wyrażenie wielu komunikatów, coraz częściej towarzyszy mu obraz – dotyczy to zarówno artykułów w gazetach, jak i blogów czy portali internetowych. Głównym zamierzeniem współwystępowania obrazu z tekstem jest zwiększenie skuteczności przekazu poprzez odwołanie się do systemu wizualnego odbiorcy.

Wizualizacja danych z jednej strony jest metodą prezentowania, uwzględniającą badania dotyczące percepcji, reagowania oraz odbioru informacji przez człowieka, a z drugiej stanowi zestaw technik umożliwiający przedstawienie użytecznych informacji w sposób klarowny i uporządkowany²⁰. To wizualna, często interaktywna eksploracja i graficzne ukazanie danych o różnym rozmiarze, charakterze i pochodzeniu. W łatwy sposób pozwala rozpoznać trendy i zjawiska trudno zauważalne podczas studiowania danych liczbowych przedstawionych np. w tabelach.

Wizualizację danych można oprzeć o trzy zasadnicze czynniki: wizję (dane są prezentowane w postaci graficznej/wizualnej, którą użytkownik może zmieniać i dostosowywać), wiedzę (wizualizacja danych umożliwia natychmiastowe zrozumienie danych), udostępnianie (łatwość zrozumienia danych czyni je przystępniejszymi i chętniej udostępnianymi dalej)²¹. Rozwiązania graficzne dla przedstawienia zebranych informacji ilościowych lub jakościowych nadają danym strukturę klarowniejszą od struktury i układu liczbowego. Jeszcze korzystniejsze są wizualizacje interaktywne, dzięki którym odbiorca może uzyskać całościowy obraz danego zjawiska, wybierając zmienne, które go interesują – „innymi słowy, dobra wizualizacja jest w stanie opowiedzieć odbiorcy odnalezioną wśród liczb historię ciekawiej i skuteczniej, niż udałoby się to zrobić, prezentując surowe dane bądź spisując na ich podstawie skomplikowany wywód”²².

¹⁹ Tamże, s. 5.

²⁰ SAS Institute, *Cztery powody, dla których nie możesz się obyć bez wizualizacji danych*, https://www.sas.com/pl_pl/offers/sas-wizualizacja-danych-ebook/index.html (dostęp: 30.04.2019).

²¹ Tamże.

²² L. Olszański, *Media i dziennikarstwo ...*, s. 271.

W odniesieniu do dziennikarstwa danych wizualizacja jest czymś, bez czego ta specjalizacja nie może istnieć²³. Olszański przywołuje słowa Fernandy Viégas, badaczki metod wizualnych stosowanych w dziennikarstwie danych:

Połowa naszego mózgu jest na sztywno przypisana do obsługi funkcji związanych z widzeniem. Wzrok to najgrubsze łącze naszego mózgu z resztą świata. Rozwijając wizualne metody przekazywania informacji, wykorzystujemy fakt naszego mocnego zaprogramowania, żeby rozumieć świat dzięki obrazom²⁴.

Wykorzystując dostępne środki wyrazu, dziennikarstwo danych wypracowało całe spektrum możliwości, które wciąż się rozszerza – od klasycznych, statycznych form po nowe formy interaktywne. Najprostsze z nich to grafiki statyczne, np. wykresy (kolumnowy, słupkowy, liniowy), za których pomocą zaprezentować można stosunkowo niewiele danych; także zakres porównawczy różnych zestawień jest ograniczony, a chociaż możliwy np. przy wykresach kolumnowych i liniowych, porównanie bywa nieczytelne dla mniej zaawansowanego odbiorcy. Wykresy jako element wizualizacji przydają się do ilustracji pojedynczego, wyizolowanego zjawiska, natomiast istotę dziennikarstwa danych stanowi przedstawianie wydarzeń bardziej złożonych. Wykorzystuje się w tym celu zarówno formy statyczne, jak i dynamiczne, interaktywne. Wśród kombinacji statycznych środków wyrazu znajdują się takie gatunki jak już opisana infografika czy zestawienia różnych wykresów w jednym miejscu. Bardzo często korzysta się również z map (dostępnych w wariacie statycznym i interaktywnym). Takie formy sprzyjają ukazaniu ważnych prawd, które „ogromnie trudno byłoby przekazać słowami”²⁵.

W kolejnych podrozdziałach zaprezentowano różne formy wizualizacji danych, usystematyzowane oraz podzielone na statyczne i interaktywne. Każdą z nich ukazano na konkretnym przykładzie, by dokładniej opisać sposoby oddziaływania i formułowania powiązań między przedstawianymi zagadnieniami.

1.3. Grafika statyczna

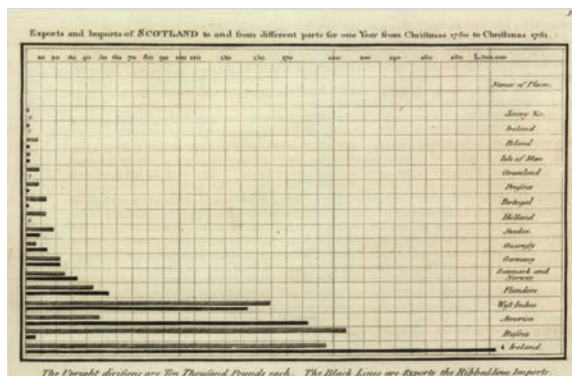
Grafika statyczna jest bardzo przydatna w badaniu różnych zjawisk zachodzących w otaczającym nas świecie. Prezentacja informacji pochodzących z liczb w formie graficznej stanowi w pewnym sensie proces tworzenia

²³ Tamże.

²⁴ Tamże.

²⁵ Tamże, s. 278.

ogólnych, bardziej zrozumiałych pojęć. Za protoplastę dzisiejszych statycznych wykresów paskowych (statycznej wizualizacji danych) uważa się Williama Playfaira, który w 1786 roku opublikował *The Commercial and Political Atlas*²⁶, w którym przedstawił dane o wymianie towarów między państwami – aby zilustrować informacje, stworzył wykresy paskowe.



Rysunek 1. Wykres paskowy autorstwa Williama Playfaira

Źródło: Wikipedia.org, *William Playfair*, http://en.wikipedia.org/wiki/William_Playfair (dostęp: 27.04.2019).

Prezentowany wykres paskowy (rys. 1) stanowi abstrakcję danych bez wykorzystania metafory rzeczywistych obiektów. Dokonania autora, który wynalazł także wykres kołowy²⁷, są efektem jego przekonania, że wykresy przedstawiają dane znacznie lepiej niż tabele. Bez wątpienia myśl ta sprawdza się w dziennikarstwie danych, gdzie zamiast prezentować zebrane i przefiltrowane dane w surowej, tabelarycznej formie, konstruuje się dla nich wizualizacje, w tym również te wynalezione przez Playfaira. Nowe możliwości pozyskiwania danych, ich nowe typy, problemy badawcze i zagadnienia spowodowały, że oprócz statystyków, matematyków i ekonomistów prezentowaniem danych zajęli się również dziennikarze. Rosnąca liczba zbieranych danych pozwala prezentować złożone informacje o badanym zjawisku w formie wizualnej. Przemysław Biecek pisze, że „statystyka pozwala na opisanie świata, ale to grafika statystyczna pomaga pojąć jego sens”²⁸, a sama „wizualizacja danych to przecięcie sztuki, statystyki i projektu”²⁹.

²⁶ W. Playfair, *The Commercial and Political Atlas*, London 1786.

²⁷ Zob. Wikipedia.org, *William Playfair*, http://en.wikipedia.org/wiki/William_Playfair (dostęp: 27.04.2019).

²⁸ P. Biecek, *Odkrywać! Ujawniać! Objawiać! ...*, s. 22–23.

²⁹ Zob. tamże.

Wizualizacja potrafi opowiedzieć o danych więcej niż nawet zaawansowany model statystyczny i pozwala zauważyć współczynniki korelacji, zależności między zmiennymi. Nie może to być oczywiście jałowa estetyzacja. Wyznacznikami dobrej wizualizacji są m.in. *data-ink ratio*³⁰, *lie factor*³¹. Oprócz tego liczą się również takie zmienne jak kolory, kształty, użyte formy.

1.4. Grafika interaktywna

Grafiki interaktywne umożliwiają przedstawienie procesów postępujących w czasie, dynamicznych; zachęcają do interakcji z elementami grafiki lub animują wybrane elementy. Animacja, oprócz tego, że stanowi komponent przyciągający uwagę odbiorcy (np. jako tło), może również przedstawiać dane. Jeśli jest dynamiczna, ale pozbawiona możliwości ingerencji w jej tempo i samodzielne nawigowania, to taki sposób przedstawienia informacji może utrudniać zrozumienie przekazu ze względu na brak czasu na odczytanie wielkości czy niemożność śledzenia wszystkich zmiennych. Sama animacja, choć wpływa na względy estetyczne, nie sprzyja zgłębieniu przekazu. Inaczej sytuacja wygląda w przypadku infografik i wykresów interaktywnych – „te pozwalają na eksplorację danych, a jeżeli pozwolimy czytelnikowi samodzielnie odkrywać historię przedstawioną przez dane, większa jest szansa że historia ta zostanie zapamiętana”³². Taki sposób wizualizowania ciekawych zjawisk stosuje większość redakcji *data journalism*, w tym „The New York Times” czy „Guardian”. Możliwość samodzielnej eksploracji przez odbiorcę oprócz budowania zaangażowania ułatwia też zrozumienie i zapamiętanie informacji, a także wybranie interesujących go zmiennych i samodzielne wygenerowanie danych o zależnościach między nimi. W tym miejscu otwiera się nowa szansa dla dziennikarza i jednocześnie zmienia się jego rola: z przewodnika po świecie informacji w stronę projektanta tej informacji, konstruktora ścieżek poznawczych i wizualizacji.

Tworzenie interaktywnej wizualizacji danych jest procesem jeszcze bardziej złożonym od planowania grafiki statycznej. Pomijając proces pracy z danymi, przy powstawaniu wizualizacji bardzo ważny jest sam projekt. Oprócz układu graficznego zaplanować należy efektywną interakcję, rozważyć

³⁰ Czyli współczynnik liczby danych i informacji przedstawiony na wizualizacji w liczbie elementów.

³¹ Współczynnik opisujący iloraz efektu przedstawianego przez wizualizację i efektu przedstawianego przez dane.

³² P. Biecek, *Odkrywać! Ujawniać! Objaśniać!...*, s. 132.

różne ścieżki poruszania się po materiale. Istnieją również gotowe szablony, które można modyfikować i parametryzować, jednak istnieje wtedy ryzyko zdublowania szablonu dla dwóch zupełnie różnych tematów lub np. przez dwie konkurencyjne redakcje.

Grafiki interaktywne dzięki możliwości dostosowywania elementów wykresu pozwalają najwydajniej przedstawić zachodzące zmiany i zależności. Sama animacja, bez możliwości interakcji z wykresem (np. użyta jako tło), zakłóca obraz, stanowiąc jedynie niepotrzebny element. Samodzielna eksploracja danych przez odbiorcę na wykresach interaktywnych oprócz zatrzymania go na dłużej sprzyja również lepszemu zapamiętaniu zaprezentowanych w ten sposób informacji i historii. Stopień rozbudowania tak przedstawionych materiałów też musi być ograniczony, ponieważ jest to nowa forma i dla osoby nieprzyzwyczajonej do takiego odbioru informacji może stanowić barierę nie do przejścia (np. ukazując zbyt wiele różnych zależności).

1.5. Problemy percepcji danych

Dane są w dzisiejszej rzeczywistości nieodłącznym składnikiem niemal każdej informacji, szczególnie w świecie cyfrowym. Mnogość komunikatów medialnych i rozwój technologiczny powodują, że komunikaty i informacje docierają do odbiorcy wieloma kanałami i za pośrednictwem wielu urządzeń. Wszystko to składa się na problem nazywany przez badaczy przeciążeniem informacyjnym, który „jest złożony i dotyczy wielu aspektów życia społecznego: od codziennych decyzji związanych z wyszukiwaniem informacji, przez ewolucję mechanizmów pamięci, aż po kształtowanie się nowego modelu wiedzy”³³.

Karol Piekarski porusza w tym kontekście termin ekonomii percepcji w Internecie, zwracając uwagę na sposób podejścia odbiorcy do procesu komunikacji. Zaznacza, że większość koncepcji nadmiaru informacyjnego opiera się na dwudziestowiecznych założeniach kultury mediów masowych, zgodnie z którymi odbiorca w jednokierunkowym procesie komunikacji zazwyczaj bezrefleksyjnie chłonie docierające do niego treści³⁴. W przypadku Internetu rola odbiorcy jest aktywna, co zmienia zdecydowanie procesy percepcji. Włączają się w to procesy takie jak podejmowanie decyzji, dokonywanie wyborów, wybieranie kategorii komunikatów i poszczególnych sekcji materiału graficznego, które mogą być wyświetlane. Aktywna rola użytkownika zdecydowanie nie wpływa

³³ K. Piekarski, *Kultura danych. Algorytmy wzmacniające uwagę*, Gdańsk 2017, s. 19.

³⁴ Tamże, s. 20.

na zmniejszenie liczby informacji, które do niego trafiają, natomiast pojawia się pojęcie „użyteczności informacji” – nie ze względu na kategorie obiektywne, a na cechy i profil odbiorcy. Choć liczba danych, które są przesyłane do użytkownika, nie zmniejsza się, to ma on wpływ na ich selekcję, co wiąże się bezpośrednio z personalizacją treści. Okazuje się to szczególnie ważne w kontekście problemu „nadprodukcji danych”, o którym wspomina Piekarski. Według niego ta tendencja nie jest czymś immanentnym i naturalnym dla systemów komputerowych, ale wypadkową możliwości technologicznych: dostępności, niskiej ceny oraz sposobów przechowywania i przetwarzania danych³⁵. Herbert Simon, cytowany przez wspomnianego autora, pisze:

W świecie bogactwa wiedzy postęp nie polega na coraz szybszym odczytywaniu i zapisywaniu informacji lub przechowywaniu coraz większej ich ilości, a raczej na ekstrakcji i wykorzystaniu schematów rządzących światem, prowadzących do ograniczenia ilości informacji, która musi być odczytywana, zapisywana lub przechowywana³⁶.

Komputery same w sobie nie są w stanie wydajnie pełnić funkcji filtrów ograniczających negatywne konsekwencje nadprodukcji informacji, ale mogą wspomagać procesy ekstrakcji danych i ich przedstawienia w bardziej dostępnej formie. Owszem, mogą pomóc również w ich selekcji, kategoryzowaniu, dopasowaniu do profilu odbiorcy, jednak nie przekażą danych w formie zamkniętego zbioru, nie stworzą początku i końca danej kategorii informacyjnej. Kierunkiem rozwoju technologii informacyjnych jest z pewnością dążenie do prezentowania danych w formie wizualnej, z wykorzystaniem narzędzi, które pomogą lepiej zrozumieć zależności między nimi czy przedstawiane przez nie zjawiska.

Udana wizualizacja powinna pokazywać informację zawartą w danych liczbowych w taki sposób, aby można było łatwo odczytać i zrozumieć związki między informacją a danymi. Jej rolą jest obrazowanie, jak duże są pewne wielkości, jak ryzykowne określone rozwiązania, jak wyglądają zależności pomiędzy zjawiskami³⁷. Poprawne zaprojektowanie ułatwia percepcję dużych liczb, rzadkich zdarzeń, interpretację występujących zależności – sumując, zrozumienie informacji. Bieчек pisze, że kluczowa dla dobrej wizualizacji danych jest świadomość, w jaki sposób ludzie myślą o liczbach i zależnościach, wymieniając wśród problemów:

- operowanie na dużych liczbach,
- operowanie na prawdopodobieństwach (szacowanie dużych i małych liczb),

³⁵ Tamże, s. 39.

³⁶ Tamże.

³⁷ Zob. P. Bieчек, *Odkrywać! Ujawniać! Objaśniać!...*, s. 80.

- odczytywanie zależności pomiędzy zdarzeniami,
- wiarę w jedyną prawdziwą historię,
- rozumienie losowości i niepewności³⁸.

Powszechne występowanie dużych liczb (terabajty pamięci dysków twardej, gigabajtowe pliki w postaci filmów wideo czy zdjęć, miliardowe kwoty pojawiające się w mediach) nie wpływa jednak na ich rozumienie przez większość odbiorców. Problemy z rozróżnieniem jednostek takich jak bilion i biliard, nieznamość znaczenia przedrostków „mega”, „tera-” czy „giga-” powodują, że odbiór dużych liczb jest często zniekształcony.

Kolejną trudność stanowi operowanie na prawdopodobieństwach. Nie licząc „pół na pół” czy „1/6” (rzut kostką), również małe liczby są problemem dla percepcji przeciętnego odbiorcy. Wyobrażenie o częstotliwości występowania rzadkich zdarzeń często bywa błędne z powodu sposobów, w jakie się je przedstawia (np. liczba 10 033 zgonów z powodu grypy lub zapalenia płuc w Polsce może wyglądać na dużą, ale to zaledwie 3,42% wszystkich zgonów w kraju³⁹) – „osoby szacujące częstości poszczególnych przyczyn śmierci znacząco przeszacowują częstości zgonów, o których często mówi się w mediach, a nie doszacowują częstości tych zgonów, o których się w mediach nie mówi”⁴⁰.

Z kolei odczytywanie zależności pomiędzy zdarzeniami budzi zastrzeżenia, kiedy odbiorca odnajduje zbyt wiele fałszywych korelacji, doszukując się znaczenia w przypadkowych związkach, co może zacierać i zniekształcać cały przekaz. Problemem jest tutaj nadinterpretacja przypadkowych elementów komunikatu i brak wiedzy, jak wyeliminować przypadkowe zależności. Z tego powodu istotę dobrej wizualizacji danych stanowi rezygnacja z mało znaczących informacji i nieuwzględnianie ich w materiale lub poinformowanie w tekście odbiorcy, na co powinien zwrócić uwagę lub co może być zwykłym przypadkiem (nakreślenie ścieżek nawigacyjnych).

Następnym problemem związanym z percepcją danych jest szukanie jednego prawdziwego przekazu, historii, wytłumaczenia. Mimo złożoności świata i interakcji między wieloma czynnikami człowiek często szuka prostego wytłumaczenia, co skutkuje daleko posuniętymi uproszczeniami. To częsty przypadek dziennikarzy piszących, którzy prezentując jakąś historię, mogą ukazać ją z konkretnej perspektywy w taki sposób, by doprowadzić odbiorcę do określonych wniosków – ukrywając niespójności, akcentując wybrane elementy,

³⁸ Tamże, s. 79–108.

³⁹ World Life Expectancy, *Poland total deaths by cause. 2010*, <http://www.worldlifeexpectancy.com/contry-health-profile/poland> (dostęp: 29.08.2019).

⁴⁰ P. Biecek, *Odkrywać! Ujawniać! Objasniać! ...*, s. 86.

unikając rozwiązań alternatywnych. W dziennikarstwie danych jest mniej miejsca na tego typu zabiegi perswazyjne, mimo to nie sposób uniknąć bierności czytelnika w kwestionowaniu własnych obserwacji czy poszukiwaniu alternatywnych wytłumaczeń wobec przedstawionego problemu, przez co trwa on w przekonaniu o braku „drugiego dna”. Nawet w takich przedstawieniach wizualnych jak wykres na określone wnioski może nakierować długość słupków czy sposób usytuowania wartości bezwzględnych na osi (ma to miejsce szczególnie w wykresach zestawiających lub rankingach).

Z ostatnim problemem wymienionym przez Biecka wiąże się zarządzanie niepewnością⁴¹, które dotyczy oswojenia się odbiorcy z tym zjawiskiem, zrozumienia przez niego prawdopodobieństwa ryzyka. Względna losowość, badanie małych lub dużych grup istotnie wpływają na wyniki. Bardzo łatwo zapomnieć, że im mniejsza badana grupa, im mniej pomiarów, tym mniej dokładna ocena; dotyczy to przede wszystkim badań i wizualizacji mających potwierdzić jakąś tezę przedstawioną w tekście.

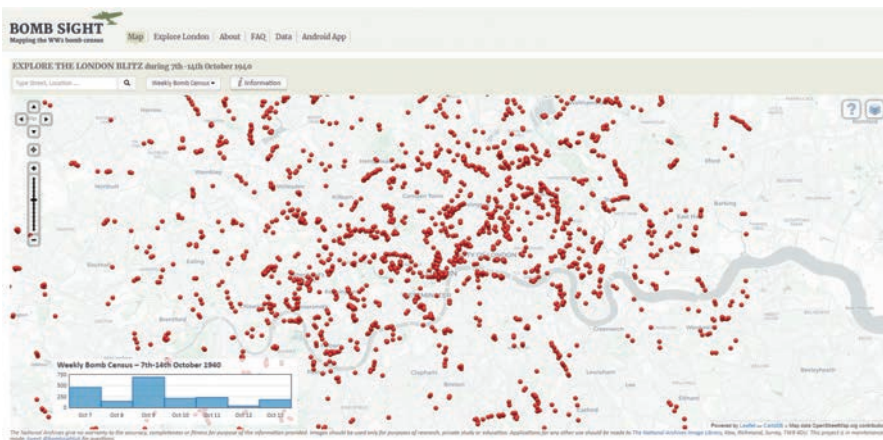
Zadaniem analityka bywa odrzucenie fałszywych zależności, ale także zaakceptowanie zwykłego przypadku. Problem z tym pojawia się przy podejmowaniu określonego tematu ze z góry postawioną tezą, którą, po zebraniu danych, próbuje się udowodnić. Nie zawsze jest to jednak możliwe, co wynika z przypadkowości wielu zjawisk. Za przykład może posłużyć eksploracja narzędzia Bomb Sight – interaktywnej mapy ilustrującej bombardowanie Londynu w czasie II wojny światowej. Są na niej zaznaczone miejsca, na które spadły bomby. Badacze, a także sami mieszkańcy Londynu, próbowali doszukać się wzorców niemieckich ataków, analizując zarówno budynki, na które zrzucono bomby, jak i te, które pozostawiono nietknięte. Niemieckie naloty nie miały jednak żadnej strategii, nie bombardowano jedynie budynków strategicznych, celowo omijając inne. Większe zagęszczenie bomb w pewnych miejscach i mniejsze w innych okazało się losowe⁴².

Ryzyko dostrzegania nieistniejących zależności jest wpisane w prezentowanie danych w formie graficznej i powodowane naturalną skłonnością człowieka do poszukiwania ciągów przyczynowo-skutkowych, które upewniają umysł człowieka „w spójnej, kontrolowanej i zrozumiałej wizji świata. Przypadkowość i brak związku to pojęcia dla naszego mózgu niepożądane”⁴³.

⁴¹ Tamże, s. 103–108.

⁴² Przykład szerzej opisano: tamże, s. 97–98.

⁴³ Tamże, s. 99.



Rysunek 2. Interaktywne mapy Londynu serwisu Bomb Sight

Źródło: <http://bombsight.org/> (dostęp: 29.08.2019).

Przyczyny i skutki są szczególnie często analizowane przy wizualizowanych danych dotyczących poparcia dla partii politycznych. Sondaże publikowane w formie wykresów i grafów w mediach elektronicznych – telewizji i Internecie, jak również te drukowane w prasie, są przedmiotem analiz i dyskusji obserwatorów politycznych, dziennikarzy czy samych polityków. Cykliczność takich badań i publikacji na ich temat rodzi naturalne porównania wyników najnowszego sondażu z ostatnimi. Mówi się wtedy o spadkach lub wzrostach poparcia, prezentowanych w postaci punktów procentowych. Nie bierze się jednak pod uwagę błędów pomiaru ani sposobów, w jaki zestawia się wyniki. Ponadto zauważenie znaczących zmian poparcia jest możliwe jedynie w dłuższej perspektywie, ponieważ zwykle zmiany te następują powoli i dopiero zestawiając wykresy z różnych okresów, da się dostrzec różnice.

Ze wspomnianym na początku tego podrozdziału przeciążeniem informacyjnym nieodłącznie wiąże się uwaga i dystrakcja, ponieważ „w społeczeństwie postindustrialnym uwaga zastąpiła informację w roli deficytowego zasobu”⁴⁴. Proces rozpraszania uwagi mnogością komunikatów jest wprost proporcjonalny do rozwoju środków komunikacji: pisma, druku, filmu, a w końcu mediów elektronicznych i Internetu. Problem informacyjnego przeładowania bądź przeciążenia (ang. *information overload*)⁴⁵ nie stanowi zatem nowości, ma swoje analogie na różnych etapach rozwoju mediów, a pojęcie na stałe weszło do słownika teorii

⁴⁴ K. Piekarski, *Kultura danych...*, s. 57.

⁴⁵ W języku polskim najczęściej stosuje się określenie „nadmiar informacji”, zaś w cytowanej pozycji Karol Piekarski proponuje termin „przeciążenia informacyjnego”.

mediów. Biorąc pod uwagę wzmocnienie przeciążenia zbyt dużą liczbą bodźców, z którymi styka się odbiorca (zaburzających jego percepcję), i szybkość przyrastania informacji (ograniczającą możliwości adaptacyjne), wydajnym remedium wydają się dobrze zaprojektowane wizualizacje informacji.

1.6. Wizualizacja a manipulacja

Jak zostało wcześniej wspomniane, wizualny komunikat ma bardzo dużą siłę przekazu, często większą od zwykłego tekstu, ponieważ jest łatwiej dostrzegany i zapamiętywany przez odbiorcę. Dane wizualizowane w formie prostych infografik czy wykresów bywają jednak różnie interpretowane. Na percepcję można wpływać na wiele sposobów, nierzadko kształtując końcowe wrażenia poprzez zakrzywianie odbioru i rzeczywistych danych. Jednym z zadań graficznej prezentacji danych jest ukazanie istniejących zależności pomiędzy przedstawianymi zjawiskami. Często, szczególnie w tekstach dziennikarskich, obok grafiki występuje również tekst, wyjaśniający te zależności i zgłębiający szczegóły prezentowanej tematyki. Odbiorca zwraca jednak uwagę w pierwszej kolejności na komunikat graficzny. Po zestawieniu z odpowiednio skonstruowanym tytułem może szybko wysnuć błędne wnioski, rezygnując z lektury dalszego tekstu.

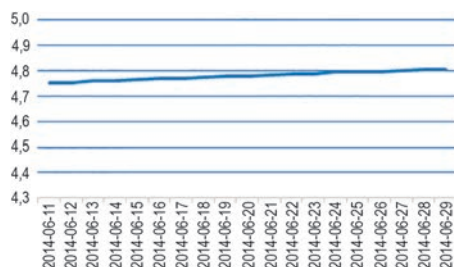
Manipulacja może dotyczyć nawet samej wielkości i grubości tekstu czy użytego fontu, co widać chociażby na przykładzie przedstawionym przez Daniela Kahnemana w książce *Pułapki myślenia. O myśleniu szybkim i wolnym*⁴⁶. Autor podaje dwa zdania, oba zawierające błędną datę urodzin Hitlera; jedno napisane zwykłą czcionką szeryfową, drugie powiększoną, pogrubioną czcionką bezszeryfową. Okazuje się, że więcej osób uznaje za prawdziwą tę drugą wersję, czytelniejszą pod względem graficznym. To dowodzi, że graficzne elementy wizualizacji danych mogą służyć celowemu przekształcaniu odbioru informacji.

Poniżej przedstawiono kilka innych przykładów, jak sposób wizualizowania danych wpływa na odbiór prezentowanych graficznie treści. Większość zaprezentowana została na grafach i wykresach, najpopularniejszych w mediach środkach wizualizacji danych na jej podstawowym poziomie, a jednocześnie najbardziej podatnych na manipulacyjne zabiegi.

Na to, jak będą odczytywane informacje, można wpłynąć np. poprzez efekt skali, dotyczący sposobu przedstawienia danych na wykresach liniowych

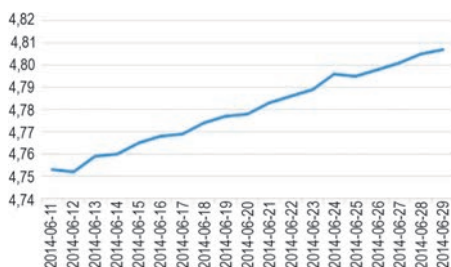
⁴⁶ D. Kahneman, *Pułapki myślenia. O myśleniu szybkim i wolnym*, tłum. P. Szymczak, Poznań 2012.

i słupkowych (czyli wykorzystywanych m.in. w sondażach prezentujących poparcie dla partii politycznych czy konkretnych polityków). Zastosowanie odpowiedniej rozpiętości skali pozwala umniejszyć wzrost danego zjawiska lub wprost przeciwnie, pokazać większą niż w rzeczywistości dynamikę.



Wykres 1. Efekt skali

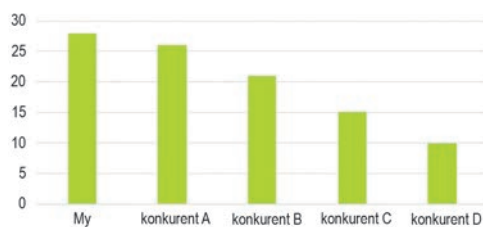
Źródło: opr. własne.



Wykres 2. Efekt skali

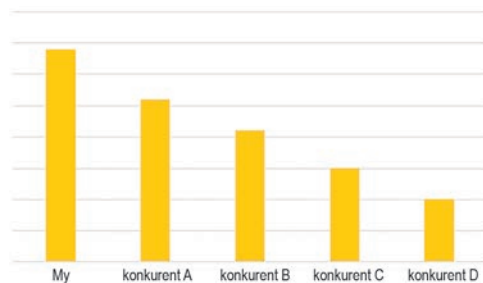
Źródło: opr. własne.

Powyższe wykresy (1 i 2) prezentują wzrost tego samego zjawiska przy użyciu różnych skali. Zakres pierwszej obejmuje linie co 0,1 punktu, z kolei druga skala to zmiany co 0,01 punktu. Zastosowanie skali mniejszej pozwala na ukazanie wzrostu dużo większego niż w rzeczywistości (dolna skala pozostaje bez zmian).



Wykres 3. Efekt skali

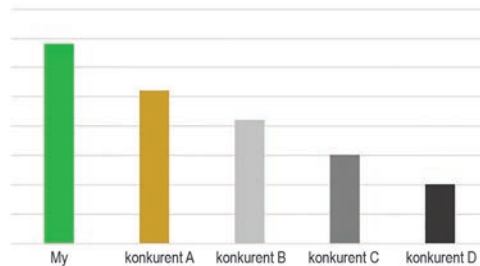
Źródło: opr. własne.



Wykres 4. Efekt skali

Źródło: opr. własne.

Efekt skali występuje także w wykresach słupkowych (kolumnowych), a jego oddziaływanie może być dodatkowo wzmocnione zastosowaną kolorystyką.

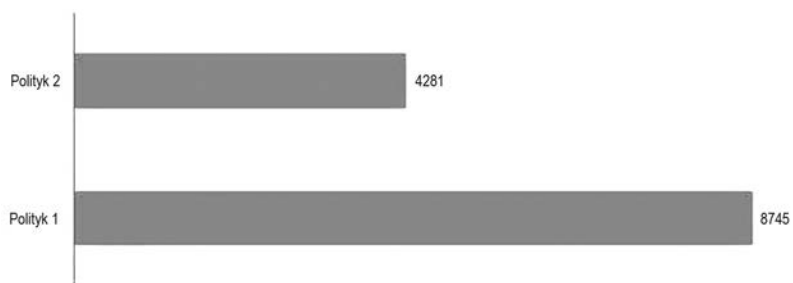


Wykres 5. Efekt skali

Źródło: opr. własne.

W wykresach 3, 4 i 5 użyto takich samych danych, ale przedstawiono je w różnych skalach. Wykres 3 ukazuje rzeczywiste dane, zawierające niewielką różnicę między „my” a głównym konkurentem, co odbiorca może sprawdzić na skali, jak również patrząc na wysokość słupków. Zmiana skali byłaby tutaj bardziej zauważalna niż w przypadku wykresów liniowych. Jednym ze sposobów manipulacji wykresem jest zmniejszenie skali i jej ukrycie (wykres 5). Efekt spotęgowano dodatkowo zastosowaną kolorystyką: szarymi, wyblakłymi odcieniami dla konkurentów i rzucającym się w oczy, żywym kolorem dla danych, które miały zostać ukazane w lepszym świetle.

Wykresy słupkowe bywają przedstawiane również w układzie poziomym. Wtedy manipulować można długością słupków, a także ich początkiem. Jeśli słupki nie rozpoczynają się w punkcie zero, to niewłaściwie porównana zostaje proporcja długości między nimi: kiedy jeden z nich jest dwukrotnie dłuższy, odbiorca traktuje to jako dwukrotną różnicę między poszczególnymi wartościami. Gdy nie umieszcza się punktu początkowego, każdy z pasków wykresu można przesunąć w kierunku wartości ujemnych lub dodatnich, odpowiednio zmniejszając lub zwiększając różnicę między prezentowanymi zjawiskami. Mechanizm działania takich zabiegów ukazano na poniższym przykładzie, zestawiając fikcyjne zarobki dwóch polityków. Tylko wykres 6 pokazuje rzeczywistą proporcję między podanymi kwotami, a punkt początkowy wynosi zero. Dla wykresu 7 zmieniono rzeczywisty obraz, zmniejszając skalę, optycznie redukując także dysproporcję między zarobkami. Ostatni (wykres 8) miał za zadanie uwypuklić różnice, manipulując wielkościami w kierunku dodatnim.



Wykres 6. Manipulacja wykresem słupkowym

Źródło: opr. własne.



Wykres 7. Manipulacja wykresem słupkowym

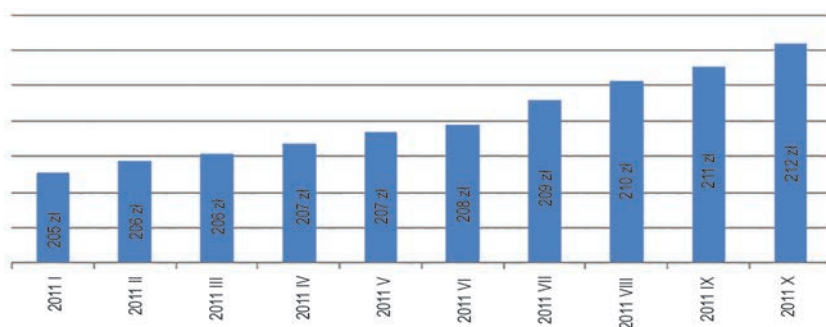
Źródło: opr. własne.



Wykres 8. Manipulacja wykresem słupkowym

Źródło: opr. własne.

Nieprawidłowości w konstrukcji wykresów i ich publikowanie dla poparcia określonej tezy ujawnia również przykład dotyczący kosztów użytkowania nieruchomości, dołączony do artykułu na temat drożejącej energii (wykres 9). Tezę artykułu⁴⁷ stanowi szybki wzrost użytkowania nieruchomości w przeliczeniu na osobę w gospodarstwie domowym. Wykres przedstawia dane od stycznia do października i, patrząc na wielkości słupków, odbiorca ma wrażenie, że podwyżka jest niemalże pięćdziesięcioprocentowa. Tymczasem okazuje się, że przez dziewięć miesięcy wzrost wyniósł 3,4% – mniej nawet od średniej rocznej inflacji⁴⁸. Zamierzeniem autora mogło być zatem wzbudzenie strachu przed szybkim wzrostem cen – poprzez wyznaczenie początku nie w punkcie 0, a od wartości 200 zł. Takie przedstawienie wykresu przyczynia się do nieprawidłowej interpretacji danych.



Wykres 9. Koszt użytkowania nieruchomości na osobę w gospodarstwie domowym

Źródło: K. Siwek, *Drożejąca energia podnosi ceny mieszkań*, <http://abcnieruchomosci.pl/drozejaca-energia-podnosi-ceny-mieszkam> (dostęp: 27.06.2019).

1.6.1. Manipulacja wizualizacją danych – studium przypadku

W 2013 roku Rada Ministrów miała podjąć decyzję w sprawie limitu inwestycji Otwartych Funduszy Emerytalnych w akcje. Początkowy limit miał być ustawiony na poziomie 75% i stopniowo zmniejszany. Był to jeden z pierwszych etapów zmian

⁴⁷ K. Siwek, *Drożejąca energia podnosi ceny mieszkań*, <http://abcnieruchomosci.pl/drozejaca-energia-podnosi-ceny-mieszkam> (dostęp: 27.06.2019).

⁴⁸ Przykład ten został opisany również przez Przemysława Biecka, *Odkrywać! Ujawniać! Objaśniać!...*, s. 145–146 i w artykule Fundacji Naukowej SmarterPoland.pl, *Jeden zły rysunek jest wart 1000 słów... korekty*, <http://smarterpoland.pl/index.php/2012/03/jeden-zly-rysunek-jest-wart-1000-slow-korekty/> (dostęp: 27.06.2019).

w OFE, wobec których Polacy mogli dokonać wyboru, czy swoje środki przenieść do ZUS, czy też zostać w wybranym funduszu. Decyzja dotyczyła składek emerytalnych, które mogły w całości trafiać do ZUS lub w części do OFE, a w części do ZUS. Okres, który wyznaczono na dokonanie wyboru, wynosił cztery miesiące.

W tym czasie próbowano różnymi sposobami ułatwić obywatelom wybór – wykorzystując reklamy, a także powołując się na artykuły ekonomistów, publicystów itd. Ofensywę reklamową prowadziły zarówno komercyjne Otwarte Fundusze Emerytalne, jak i Ministerstwo Pracy i Polityki Społecznej. Co ciekawe, w kluczowym momencie podejmowania decyzji OFE miały ograniczone możliwości reklamowe, by zainteresowane osoby podejmowały decyzję świadomie, a nie pod wpływem agresywnej reklamy⁴⁹. Swoje materiały przed kluczowym okresem przygotowywało również Ministerstwo.

Ministerstwo Pracy i Polityki Społecznej (obecnie Ministerstwo Rodziny, Pracy i Polityki Społecznej) na swojej stronie informowało, że przygotuje aktualne informacje o wysokości waloryzacji konta i subkonta w ZUS oraz o wynikach inwestycyjnych wszystkich funduszy emerytalnych, czym zajmą się minister finansów i minister pracy we współpracy z ekspertami ZUS i Komisji Nadzoru Finansowego⁵⁰. Zgodnie z zapowiedziami z 26 czerwca 2013 roku ministrowie pracy oraz finansów zaprezentowali szczegółowy przegląd funkcjonowania systemu emerytalnego⁵¹. Zespół ekspertów Ministerstwa Pracy i Polityki Społecznej oraz Ministerstwa Finansów przedstawił rekomendacje zmian, a także odrzucił dyskutowane wcześniej warianty. Na stronie znaleźć można było, jak informowali autorzy, „pełną wersję przeglądu funkcjonowania systemu emerytalnego, informację prasową wraz ze streszczeniem raportu [...] także infografiki prezentujące najważniejsze liczby z raportu”⁵².

Właśnie te przygotowane przez ministerstwo infografiki stanowią przedmiot poniższej analizy. Ich funkcją powinno być, zgodnie z charakterystyką gatunku, przekazanie skomplikowanych danych i informacji w możliwie atrakcyjny i przystępny dla odbiorcy sposób. Patrząc jednak na zawarte na infografice wykresy, należy

⁴⁹ Money.pl, *Premier ujawnia, która wersja zmian w OFE wygra*, <http://www.money.pl/gospodarka/wywiady/arttykul/premier;ujawnia;ktora;wersja;zmian;w;ofe;wygra,54,0,1421110.html> (dostęp: 28.06.2019).

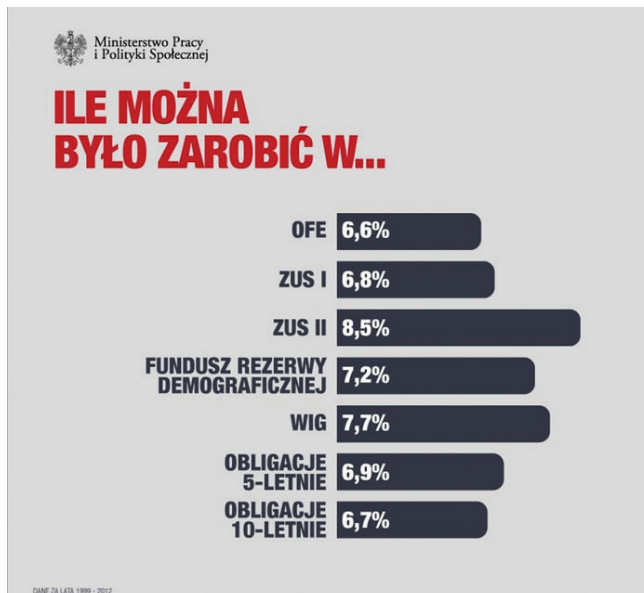
⁵⁰ Ministerstwo Pracy i Polityki Społecznej, *Planowane zmiany w systemie emerytalnym*, <http://www.mpips.gov.pl/aktualnosci-wszystkie/ubezpieczenia-spoleczne/art,6499,planowane-zmiany-w-systemie-emerytalnym.html?serwis=2> (dostęp: 28.06.2019).

⁵¹ Ministerstwo Pracy i Polityki Społecznej, *Przegląd funkcjonowania systemu emerytalnego*, <http://www.mpips.gov.pl/aktualnosci-wszystkie/ubezpieczenia-spoleczne/art,6266,przeglad-funkcjonowania-systemu-emerytalnego.html> (dostęp: 28.06.2019).

⁵² Tamże.

przemyśleć, czy miały spełniać podstawowe kryteria poprawności w ich tworzeniu, czy jedynie stanowić element estetyczny, nie ułatwiający zrozumienia przekazu.

Wizualizowanie danych w poniższych wykresach cechuje mocne uproszczenie, pojawiają się niejasności dotyczące zastosowanej skali, proporcji między przedstawionymi słupkami czy wzmocnienia ukazywanych w materiale różnic (ograniczających się czasami do jednego punktu procentowego). Nasuwa się zasadne pytanie, czy użycie do wizualizacji danych w infografice wykresów – kojarzących się raczej ze szczegółowością i elementami formalnymi jak oś XY, skala, punkty pośrednie – jest prawidłowe.



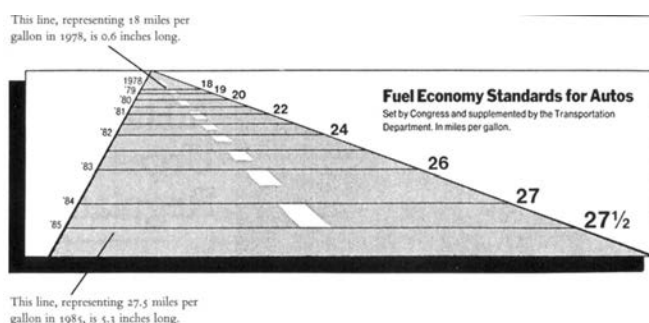
Rysunek 3. Materiały Ministerstwa Pracy i Polityki Społecznej

Źródło: Ministerstwo Pracy i Polityki Społecznej, *Przegląd funkcjonowania systemu emerytalnego*, <http://www.mpips.gov.pl/aktualnosci-wszystkie/ubezpieczenia-spoeczne/art,6266,przeglad-funkcjonowania-systemu-emerytalnego.html> (dostęp: 11.07.2019).

Pierwszy wykres (rys. 3) ukazuje informacje dotyczące wyników inwestycyjnych OFE, ZUS i innych. Jest on oparty na układzie poziomym. Odbiorca zwraca uwagę na długość pasków, koncentrując się na ich porównaniu. Aby uniknąć nieprawidłowości i mylnego wrażenia, każdy z elementów wykresu powinien rozpoczynać się od granicy zera procent, która na powyższym wykresie pozostaje niewidoczna (nie wiadomo, jaka wartość wyznacza początek wykresu). Odnosząc się do samych wartości przedstawionych na wykresie, interesujących potencjalnego odbiorcę (osobę stojącą przed wyborem: OFE czy ZUS), bezpośrednie porównanie ZUS II

i OFE daje stosunek 100:168 (mierząc stosunek długości)⁵³. Dane liczbowe, które wizualizuje wykres, wskazują, że rzeczywisty stosunek to 100:129.

Aby wyjaśnić, z czego wynika ta rozbieżność, należy odwołać się do *lie factor*, sformułowania zdefiniowanego i opisanego przez Edwarda Tufte'a w *The Visual Display of Quantitative Information*⁵⁴. „Współczynnik kłamstwa”, bo tak brzmi bezpośrednie tłumaczenie, oznacza „stosunek efektu widocznego na grafice do efektu wykazywanego przez dane, na podstawie których zbudowana została grafika”⁵⁵. Czynniki ten, mierzony na wizualizacjach, wskazuje stopień błędności w przedstawieniu danych oraz wykorzystania reguł percepcji dla wyolbrzymienia wybranych danych lub ich sztucznego zmniejszenia.



Rysunek 4. Grafika ukazująca *lie factor* na przykładzie⁵⁶

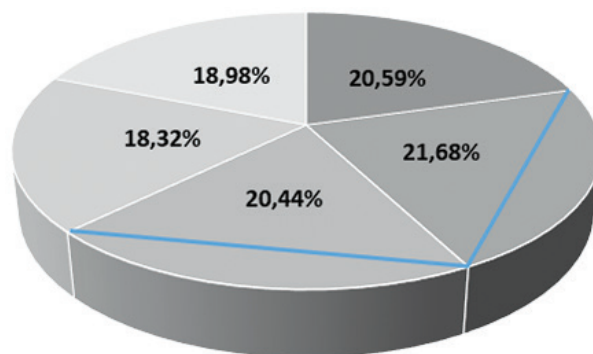
Źródło: E.R. Tufte, *The Visual Display...*, za: K. Schneider, *Visual Display of Information*, <https://technogenii.net/visual-display-of-information/> (dostęp: 11.07.2019).

⁵³ Analiza tego wykresu wraz z odniesieniem do rzeczywistych danych liczbowych wizualizowanych na infografice znajduje się na blogu: Fundacja Naukowa SmarterPoland.pl, *Zmiany w OFE a wykresy Ministerstwa Pracy i Polityki Społecznej*, <http://smarterpoland.pl/index.php/2013/11/zmiany-w-ofe-lie-factor-a-ministerstwo-pracy-i-polityki-spolecznej/> (dostęp: 15.07.2019).

⁵⁴ E.R. Tufte, *The Visual Display of Quantitative Information*, Cheshire (CT) 1991.

⁵⁵ B. Czapiewski, *Co to jest chartjunk (wykres śmieciowy) i lie-factor (współczynnik kłamstwa)?*, <https://skuteczneraporty.pl/blog/co-to-jest-chartjunk-wykres-smieciowy-i-lie-factor-wspolczynnik-klamstwa/> (dostęp: 17.08.2019).

⁵⁶ Rysunek 4 i wykres 10 pokazują przykłady, gdzie współczynnik kłamstwa jest wyjątkowo duży. Obie grafiki pokazują także zaburzoną proporcję, która wpływa na fałszywy obraz przedstawianych danych i informacji. Pierwszy przykład pojawia się najczęściej w kontekście *lie factor*, w którym dane obrazują wzrost na poziomie 53% (od 18 do 27,5 punktu). Według proporcji przedstawionych na grafice wartość wzrostu to ponad 700%. Współczynnik kłamstwa to zatem prawie 15 punktów. Druga ilustracja (wykres 10) to zaburzenie proporcji w wykresie kołowym.



Wykres 10. Wykres ukazujący lie-factor na przykładzie

Źródło: opr. własne.

Dla wykresu 10 *lie factor* wyniósł około 1,3 punktu, co nie jest wynikiem bardzo dużym, ale mimo wszystko wpływającym na rzeczywisty obraz wynikający z danych. Podobna sytuacja występuje w kolejnych dwóch grafikach informacyjnych (rys. 5 i 6).



Rysunek 5. Materiały Ministerstwa Pracy i Polityki Społecznej

Źródło: Ministerstwo Pracy i Polityki Społecznej, *Przegląd funkcjonowania ...*



Rysunek 6. Materiały Ministerstwa Pracy i Polityki Społecznej

Źródło: Ministerstwo Pracy i Polityki Społecznej, *Przegląd funkcjonowania ...*

Grafiki 5 i 6 również oddziałują na odbiorcę wykresami. Dodatkowo przekaz wzmocniono kolorem i wielkością czcionki napisów „10 miliardów złotych – OFE dostały z opłaty od składki” i „17 miliardów złotych – tyle opłat pobrały od przyszłych emerytów od 1999 roku”. To częste zabiegi graficzne stosowane dla podkreślenia jakiejś informacji, czyniące ją mocno dostrzegalnym elementem wizualizacji, najlepiej zapamiętywanym przez odbiorcę. Wróćmy jednak do samego wykresu. Na powyższych grafikach wykresy kolumnowe tworzą monety, nawiązując w ten sposób do tematu z zakresu finansów. Ponownie jednak słupki nie są zbudowane na skali, nie mają punktu 0, a monety nie odpowiadają żadnej jednostce opłat – ich ilość jest przypadkowa. Wykresy przedstawiające opłaty pobrane przez OFE nie pokazują zależności w prezentowanych danych, zestawiając wartości z różnych lat, mimo że wysokość opłat była uzależniona od zmiennych czynników.

Nieścisłości w materiałach infograficznych przygotowanych przez Ministerstwo można znaleźć także na wykresach opartych na osiach (liniowych). Przeciętny odbiorca wykresy z rysunku 7 odczyta jako stworzone na jednej wspólnej osi z porównywanymi dwiema wartościami. Trzeba się przyjrzeć bliżej i zwrócić uwagę na szczegóły, by dostrzec, jak blisko siebie są wartości 0,41% na czerwonej linii wykresu i 0,100% na linii zielonej. Wyrysowane linie pomocnicze nie pomagają, ponieważ nie odnoszą się do żadnych wartości.



Rysunek 7. Materiały Ministerstwa Pracy i Polityki Społecznej

Źródło: Ministerstwo Pracy i Polityki Społecznej, *Przegląd funkcjonowania...*

Podobne rozbieżności dostrzegalne są przy kolejnych wykresach (rys. 8), również liniowych. Także tutaj linie pomocnicze nie spełniają swojej funkcji, a odległości między poszczególnymi punktami są nieproporcjonalne. Wystarczy zwrócić uwagę na rozbieżność różnic między wartościami 48,4 i 52,8 na wykresie czerwonym (różnica 4,4 punktu) a 42,3 i 36,4 na wykresie zielonym (różnica 5,9 punktu). Widać wyraźną dysproporcję między tymi wskaźnikami.

Powyższy przykład pokazuje nie tylko błędy w konstrukcji wizualizacji danych na wykresach, ale również problemy z formułami wyliczeń. Przedstawione w materiałach ministerstwa dane liczbowe okazały się nieporównywalne ze stopami zwrotu – wszystkie średnie, z wyjątkiem stóp zwrotu OFE, liczone są jako średnie geometryczne, przez co stopa zwrotu OFE, liczona w inny sposób, wypada niekorzystnie w porównaniu do ZUS. Po zastosowaniu jednej metodologii dla porównywanych podmiotów – ZUS i OFE – korzystniej wypada OFE.

Powstał na ten temat oddzielny raport, wykorzystujący metodologię zastosowaną przez ministerstwo, który wykazał omawiane nieprawidłowości⁵⁷.



Rysunek 8. Materiały Ministerstwa Pracy i Polityki Społecznej

Źródło: Ministerstwo Pracy i Polityki Społecznej, *Przegląd funkcjonowania...*

Powyższe przykłady potwierdzają łatwość manipulowania danymi wizualizowanymi w formie wykresów. Wizualizacja danych na tym poziomie ma pokazać i uwidocznic różnice między porównywanymi zjawiskami. Problem w tym, że nieznanie przez odbiorcę sposobów konstrukcji takich wykresów i ich specyfiki powoduje błędny odczyt danych i nieprawidłową interpretację rzeczywistej informacji. Zmiana skali, stosowana w celu uwidocznienia różnic, nawet przy bliskich wartościach bezwzględnych powoduje dezinformację i rodzi obawy o manipulację przekazem, co jest łatwo mierzalne m.in. przez *lie factor*.

⁵⁷ Zob. M. Bitner, A. Chłopecki, S. Gomułka, W. Otto, J. Stępień, M. Wiśniewski, *Obywatelski kontrraport – krytyczna analiza rządowego przeglądu systemu emerytalnego*, Warszawa 2013, <http://kobe.org.pl/wp-content/uploads/2013/07/Obywatelski-Kontrraport.pdf> (dostęp: 15.07.2019), s. 17–24.



Rysunek 9. Materiały Ministerstwa Pracy i Polityki Społecznej

Źródło: Ministerstwo Pracy i Polityki Społecznej, *Przegląd funkcjonowania...*

Powyższy przykład (rys. 9) obrazuje manipulację wizualizacją danych w celu wpłynięcia na odbiorcę w określony sposób. Służy temu wykorzystanie naturalnej tendencji człowieka do porównywania względnych długości słupków (wykresy słupkowe mają bowiem za zadanie pokazywanie proporcji). Aby uwypuklić różnice w wykresach punktowych, wystarczy zrezygnować z umieszczenia w wykresie punktu zero.

Z jednej strony „stosowanie wykresów paskowych bez uwzględnienia punktu zero jest błędem. Podobnym do błędów ortograficznych, ale o większym zasięgu. Ponieważ zniekształca postrzeganie liczb”⁵⁸. Z drugiej jednak strony, wykresy skonstruowane w prawidłowy sposób, wiernie odwzorowujące dane z zachowaniem pełnej skali i odpowiednich proporcji, mogłyby się stać dla przeciętnego czytelnika nieczytelne i niezrozumiałe, a główna teza, którą zamierzał przekazać nadawca, byłaby trudniejsza do odczytania. Do pełnego zrozumienia skomplikowanych wykresów, przedstawiających dane liczbowe z zachowaniem matematycznych zasad, odbiorca potrzebuje kompetencji zarówno z zakresu odczytywania danych, jak i z matematyki czy statystyki. Większość przeciętnych czytelników prasy czy portali internetowych nie należy do tej wąskiej, wyspecjalizowanej grupy odbiorców, co może tłumaczyć „upraszczanie” wizualizacji kierowanych do szerszego grona.

Innym zagadnieniem w tym kontekście są infografiki obrazujące dane, ale niebędące wykresami. Prezentacje informacji opartych na liczbach, o formie zbliżonej do wykresu, upraszcza się przede wszystkim po to, by trafić do jak najszerszej grupy odbiorców. Za przykład posłuży tutaj ponownie ministerialna infografika, przygotowana tym razem przez Ministerstwo Edukacji Narodowej.



Rysunek 10. Grafika Ministerstwa Edukacji Narodowej

Źródło: Kancelaria Prezesa Rady Ministrów, *Wsparcie dla dzieci z uboższych rodzin w nowym roku szkolnym*, <https://www.premier.gov.pl/mobile/en/news/news/en-wsparcie-dla-dzieci-z-ubozszych-rodzin-w-nowym-roku-szkolnym.html> (dostęp: 11.07.2019).

⁵⁸ Fundacja Naukowa SmarterPoland.pl, *Zmiany w OFE...*

Grafika (rys. 10) prezentuje poziom dofinansowania zakupu podręczników dla uczniów w trudnej sytuacji materialnej. Przedstawiono na nim kwotę dofinansowania z ostatnich trzech lat za pomocą książek ułożonych jedna na drugiej. Wątpliwości budzą dane z 2012 i z 2010 roku:

ta sterta sprawia wrażenie dwa razy większej w roku 2012 niż w roku 2010, co może sugerować, że wielkość dofinansowania wzrosła znacząco, a może nawet, że liczba podręczników, które udało się kupić, jest znacznie wyższa. Okazuje się jednak, że wzrost nakładów na podręczniki, który wynika z danych to 11% rocznie. I jest to wzrost niższy niż ceny podręczników rosnące szybciej (zgodnie z niektórymi szacunkami drożące w tempie 15% na rok)⁵⁹.

W odróżnieniu od przykładów z Ministerstwa Pracy i Polityki Społecznej, grafika przygotowana przez MEN nie jest wykresem *sensu stricto*. Nie pretenduje do miana profesjonalnej grafiki statystycznej, stanowi raczej graficzną impresję, infografikę stylizowaną na wykres. Czy w takim przypadku dane przedstawione graficznie nie muszą być tożsame z wartościami liczbowymi? Pytanie dotyczy rzetelności prezentowania danych charakterystycznej dla wykresów, a niekoniecznie dla infografik. Jeśli tego typu materiał dziennikarz zestawia z tekstem, w którym omówi przedstawione graficznie zjawiska, to nie będzie problemu. Dezinformacją i do pewnego stopnia manipulacją można jednak nazwać występowanie takiej infografiki samodzielnie, bez wyjaśnienia tekstowego.

1.6.2. Wybrane sposoby przekłamań na wykresach i grafikach informacyjnych

Efekt perspektywy

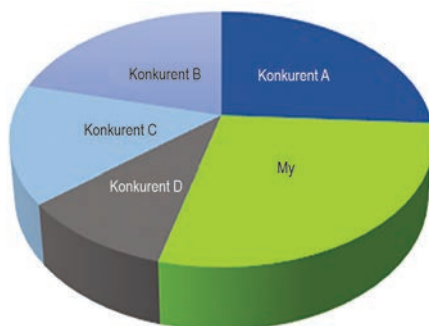
Dane umieszczane na wykresach bardzo często wizualizowane są w formie wykresów (diagramów) kołowych. Mogą wówczas zostać przedstawione „płasko” (w wymiarze 2D), ale również w perspektywie (trójwymiarowo, 3D). Informacje prezentowane w wykresach kołowych oprócz czynnika perspektywy wzmacnia również użyta kolorystyka, dużo bardziej niż w przypadku wykresów kolumnowych i paskowych.

Perspektywa bywa często jedynie zabiegiem wizualnym, mającym wpłynąć na odbiór i estetykę komunikatu. Trzeba natomiast pamiętać, że takie działania nie pomagają w żaden sposób w przekazie, wręcz przeciwnie – dają możliwości do manipulacji informacją i zaciemniania rzeczywistego obrazu. „Trudno znaleźć jakiegokolwiek racjonalne uzasadnienie dla tego typu działań poza chęcią

⁵⁹ P. Biecek, *Odkrywać! Ujawniać! Objaśniać! ...*, s. 146–147.

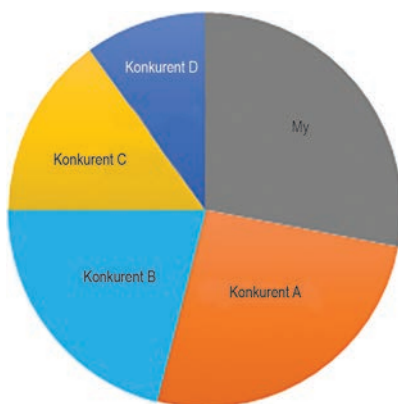
udziwnienia wykresu. Niestety takie zabiegi są wciąż dosyć częste⁶⁰. Zasadny okazuje się tutaj stawiany infografikom zarzut jałowej estetyzacji, aczkolwiek należy zaznaczyć, że takie zabiegi szkodzą jedynie tradycyjnym wykresom.

Użycie prezentacji trójwymiarowej daje możliwości uwypuklenia tych obszarów wykresu, na których zależy nadawcy, i sprawienia, by wydawały się korzystniejsze, nawet jeśli odpowiadają mniejszym wartościom liczbowym. Efekt często zostaje wzmocniony również przez użycie kolorów – jaskrawymi, rzucającymi się w oczy barwami wyróżnia się interesujące nadawcę fragmenty, natomiast dane, które mają być na wykresach ukryte, występują jako te ciemniejsze, zaznaczone szarościami.



Wykres 11. Kolory i nachylenie

Źródło: Cognity.pl, *Manipulacja danymi na wykresach*, <https://www.cognity.pl/manipulacja-danymi-na-wykresach,blog,94.html> (dostęp: 11.07.2019).



Wykres 12. Kolory i nachylenie

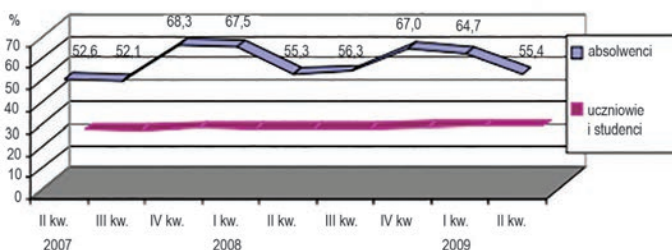
Źródło: Cognity.pl, *Manipulacja danymi...*

⁶⁰ Tamże, s. 150.

Wykresy kołowe 11 i 12 prezentują te same dane, ale na różne sposoby. Na wykresie 11 „Konkurent A” wypada gorzej od kategorii „My”. Efekt ten uzyskano zastosowaniem perspektywy i odpowiedniej kolorystyki. Rzeczywistym danym bardziej odpowiada wykres 12, na którym różnica między „nami” a „konkurentem A” nie jest już tak wyraźna.

Wykresy kołowe stanowią szczególny przypadek, kiedy przedstawienie w tzw. „trzecim wymiarze” zupełnie się nie sprawdza. Ich charakterystyką są bowiem kąty, reprezentujące wartości (im większa wartość, tym większy kąt danego pola). Ulega to zmianie, gdy wykres kołowy zostaje przedstawiony w perspektywie – zamiast koła odbiorca ma do czynienia z elipsą i proporcje pomiędzy kątami, pozwalające ocenić odpowiadające im liczby lub procenty, ulegają zniekształceniom.

Powyżej wyjaśniono stosowanie perspektywy w wykresach kołowych. Zabieg ten pojawia się także w wykresach innego typu. Wykresy liniowe, warstwowe czy kolumnowe przedstawione jako trójwymiarowe grafiki tracą swoją czytelność i przydatność dla odbiorcy. W tym przypadku takie środki graficzne są jedynie sztuczną estetyzacją przekazu, wpływającą niekorzystnie na jego wartość dla czytelnika.



Wykres 13. Efekt głębi

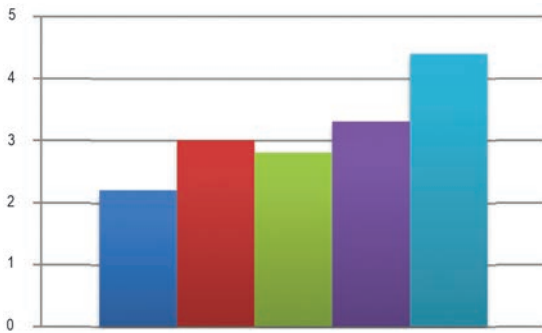
Źródło: Główny Urząd Statystyczny, *Monitoring pracy. Wejście ludzi młodych na rynek pracy*, http://stat.gov.pl/cps/rde/xbcr/gus/PUBL_Wejscie_ludzi_mlodych_na_rynek_pracy.pdf (dostęp: 11.07.2019).

Zastosowanie tzw. głębi w powyższym przykładzie nie ma racjonalnego wytłumaczenia, jedynie utrudnia odczytanie wartości przedstawianych przez wykres. Linia dotycząca uczniów i studentów wydaje się całkowicie płaska, reprezentująca jedną wartość przez cały wykres. Dane z linii absolwentów bardzo trudno odczytać, a pośrednie punkty procentowe zamiast pomagać, wprowadzają jeszcze większy chaos.

Proste wykresy, które samodzielnie wykona przeciętny użytkownik dysponujący choćby kilkoma danymi liczbowymi, bardzo często ulegają niepotrzebnym

zabiegom mającym zwiększyć ich atrakcyjność. Nie wszystkie takie przypadki stanowią bowiem celową manipulację danymi, czasem jest to po prostu jałowa estetyzacja ze zbyt dużą ilością zbędnych i szkodliwych efektów. Obok „głębi” czy „trzeciego wymiaru” można w tym kontekście wymienić również „obroty”.

Pierwszy błąd związany z obrotami dotyczy podstawowego wykresu, jaki można wygenerować w programach typu Excel. Efekt zaprezentowany na wykresie 15 osiąga się poprzez zastosowanie ujemnych wartości na osi Y, a dodatnich na osi X przy zaznaczonej opcji efektu 3W (trzeci wymiar). Obrót w takiej formie utrudnia precyzyjne odczytanie wartości oraz uniemożliwia poprawne ustalenie punktu zero. Porównując te same dane na wykresie 14 i 15, zauważy się brak tzw. ściany i podstawy wykresu w obróconej wersji (wykres 15). Jest to częsty przypadek w tego typu wizualizacjach, utrudniający prawidłowy odbiór informacji.



Wykres 14. Najprostszy wykres kolumnowy

Źródło: opr. własne.

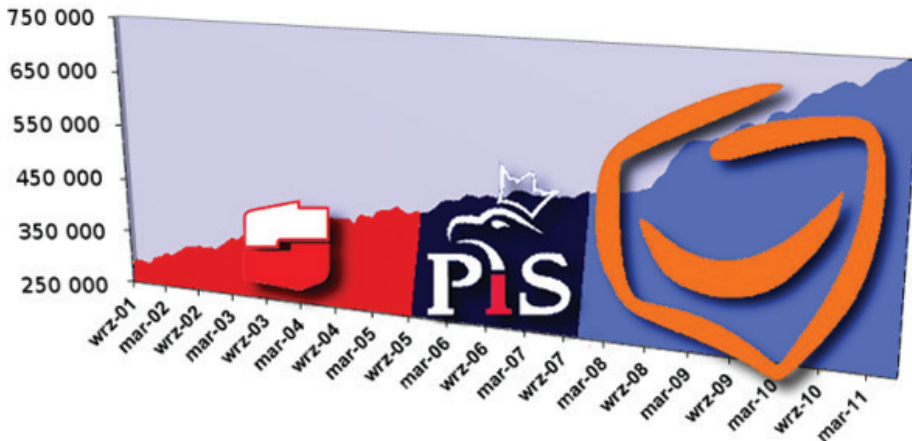


Wykres 15. Wykres kolumnowy z efektem trzeciego wymiaru

Źródło: opr. własne.

Efekty związane z obracaniem wykresów pojawiają się jednak nie tylko w ich prostych odmianach. Przykład stanowi poniższy wykres (16), dotyczący rosnącego zadłużenia państwa w okresie rządów trzech partii politycznych i stanowiący część artykułu *Gospodarcze osiągnięcia partii politycznych*⁶¹, przygotowanego przez serwis Bankier.pl.

⁶¹ M. Świerkot, K. Kolany, *Gospodarcze osiągnięcia partii politycznych*, <http://www.bankier.pl/wiadomosc/Gospodarcze-osiagniecia-partii-politycznych-2417944.html> (dostęp: 11.07.2019).



Wykres 16. Wykres przedstawiający osiągnięcia partii politycznych w „trzecim wymiarze”

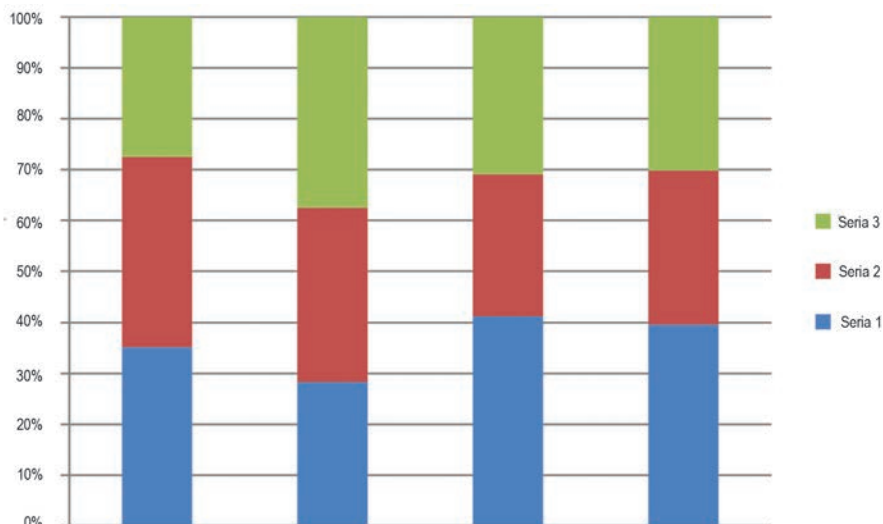
Źródło: M. Świerkot, K. Kolany, *Gospodarcze osiągnięcia partii politycznych*, <http://www.bankier.pl/wiadomosc/Gospodarcze-osiagniecia-partii-politycznych-2417944.html> (dostęp: 11.07.2019).

Przykład ten (wykres 16) analizował w swojej książce Biecek⁶², wskazując, że występuje w nim wiele elementów utrudniających zrozumienie informacji. Chodzi o symbole partii, zakrywające wykres i zaburzające jego postrzeganie, brak punktu zero, komplikujący porównanie wartości bezwzględnych, czy też sam obrót, przez który trudno stwierdzić, kiedy zadłużenie rosło najszybciej i jak wyglądało tempo wzrostów.

Ukrywanie trendów

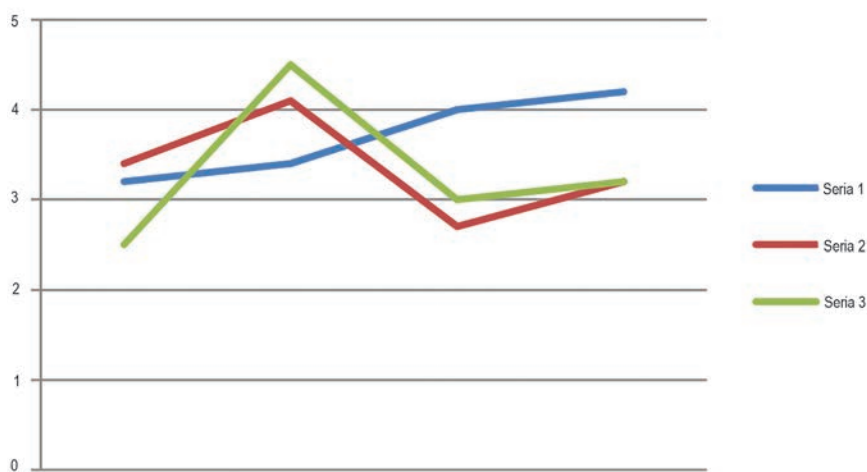
Trzeci przykład nieprawidłowego wizualizowania danych na wykresach już ściślej wiąże się z manipulacją. Polega bowiem na celowym ukrywaniu dwóch powiązanych ze sobą wskaźników, z których jeden pokazuje wyraźny wzrost lub wyraźny spadek. Intencją twórcy takiego wykresu jest zatem zatajenie informacji i wprowadzenie odbiorcy w błąd. Cel osiągany zostaje dzięki zamianie wykresu słupkowego, kolumnowego lub poziomego (paskowego) w tak zwany wykres skumulowany. Zestawienie dwóch pierwszych wykresów opiera się na porównaniu trzech różnych serii (mogą być to np. wpływy trzech różnych firm czy wydatki trzech województw). Poniższy przykład (wykresy 17 i 18) demonstruje, jak zaciera się obraz w wykresie skumulowanym w porównaniu z łatwym do odczytania wykresem liniowym.

⁶² P. Biecek, *Odkrywać! Ujawniać! Objaśniać! ...*, s. 152.



Wykres 17. Ukrywanie trendów na wykresie warstwowym/skumulowanym

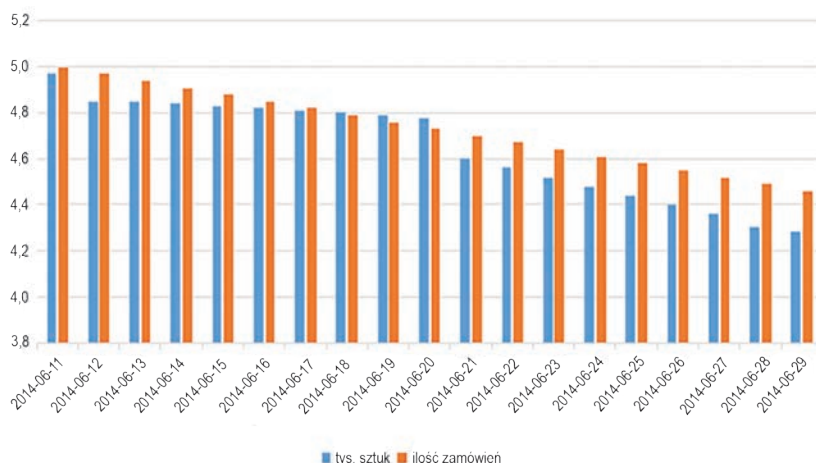
Źródło: opr. własne.



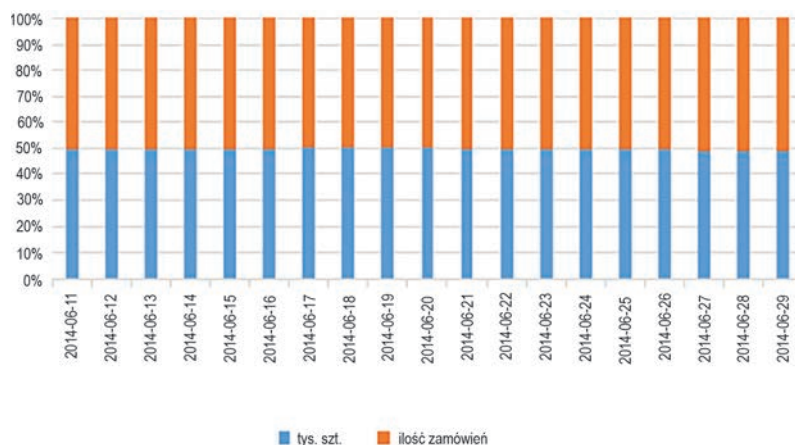
Wykres 18. Te same dane, co w wykresie 17, ukazane w sposób liniowy

Źródło: opr. własne.

Kolejny przykład (wykresy 19 i 20), korzystający z większej ilości zbliżonych informacji, równie dobrze pokazuje efekt „ukrywania trendów”.



Wykres 19. Ukrywanie trendu w wykresach

Źródło: Cognity.pl, *Manipulacja danymi...*

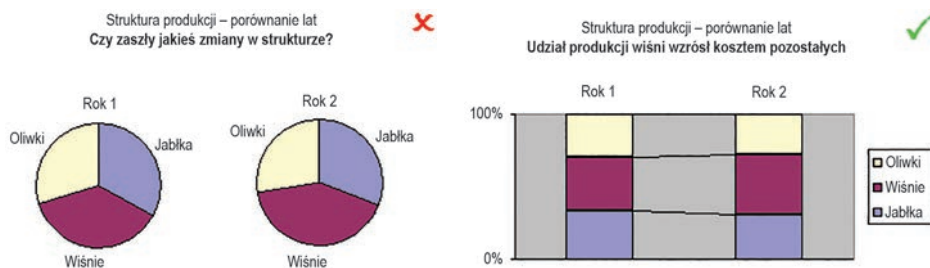
Wykres 20. Ukrywanie trendu w wykresach

Źródło: Cognity.pl, *Manipulacja danymi...*

Zestawienie i kolejność

Zestawienie jest częstym chwytem, mającym za zadanie wykazać różnice między porównywanymi zjawiskami. W przypadku bezpośredniego zestawiania wykresów zaprezentowane zostaje wiele różnych danych, często na zasadzie kontrastu, z użyciem odpowiednich kolorów, wskazaniem skali. Jednak wymóg

poprawnego zestawienia to użycie takich samych wykresów. Aby taki zabieg miał sens i stanowił wartość dla odbiorcy, musi być czytelny i wyraźnie pokazywać różnice między wykresem A i wykresem B, co nie zawsze się udaje.



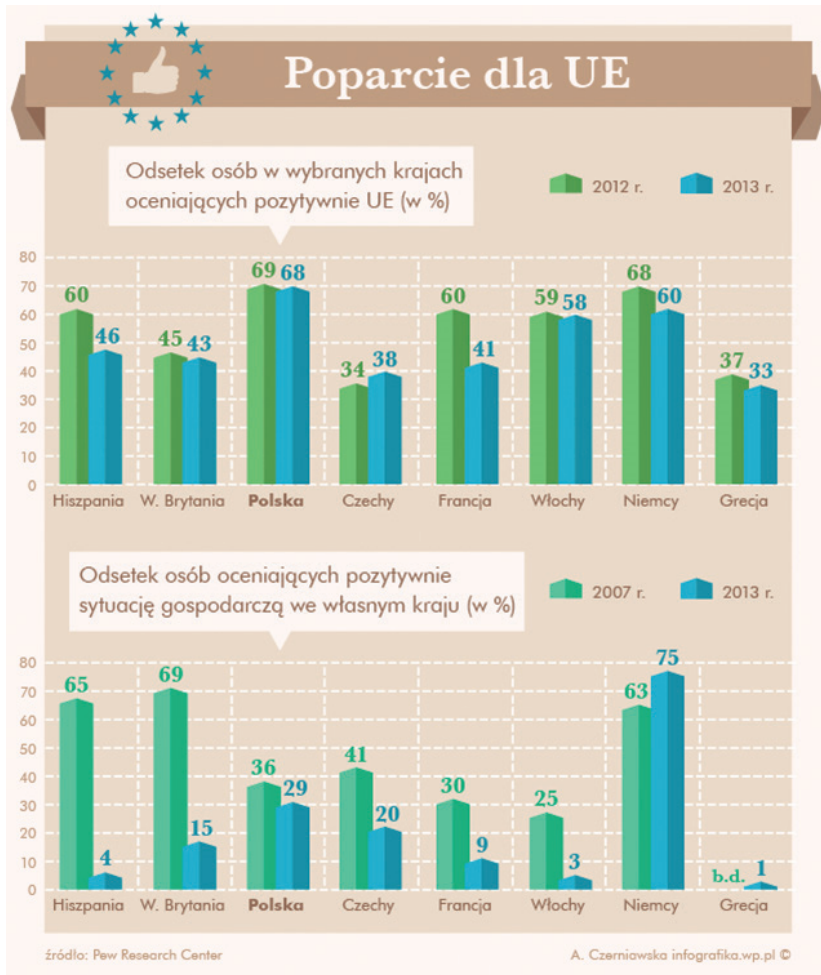
Wykres 21. Zestawienie i kolejność

Źródło: <http://wykresy.blox.pl/html> (dostęp: 11.07.2019).

Na wykresie po lewej stronie zestawiono dane z dwóch lat, dotyczące produkcji owoców, na wykresach kołowych (wykres 21). Bez możliwości zmierzenia kątów zmiany w strukturze produkcji są słabo dostrzegalne. Lepiej widać je na wykresie kolumnowym skumulowanym, który można jeszcze „wspomóc” liniami serii, podkreślającymi wzrost określonych danych kosztem innych (wykres 21 po prawej). Porównywanie wykresów kołowych, szczególnie takich o większej liczbie poziomów, jest problematyczne w związku z bardzo trudnym porównywaniem wycinków.

Przy poprzednich kategoriach omawiano m.in. długość słupków i ich wpływ na percepcję danych. Równie ważne okazuje się ich ustawienie, kolejność na wykresie. W tym przypadku efekt wzmacniany bywa kolorystyką. Czytając wykres powierzchownie, bez dokładnego zapoznania się z legendą (co przy przeglądaniu stron internetowych zdarza się często), z słupków przedstawionych odwrotnie można odebrać mylne wrażenie np. co do lat, z których prezentowane są informacje. W taki sposób zmienia się chociażby percepcję spadków lub wzrostów poparcia w polityce. Zapamiętując wykresy wzrokowo, odbiorca przyswaja obraz sytuacji przeciwny do rzeczywistego.

Błędy w kolejności, a także korzystanie z kolorów dla uwypuklenia konkretnych treści, zauważalne są w wykresie 22. Zestawiono w nim dane dotyczące liczby osób z ośmiu europejskich państw, wypowiadających się pozytywnie o Unii Europejskiej w latach 2012 i 2013. Zielone słupki po lewej stronie błędnie sugerują wzrost; zazwyczaj przedstawia się bowiem jako pierwsze dane bardziej aktualne, zgodnie z ruchem wskazówek zegara, co w tym przypadku zostało zaburzone. Poniższa grafika (wykres 22) składa się z dwóch wykresów, zbudowanych według tej samej zasady.

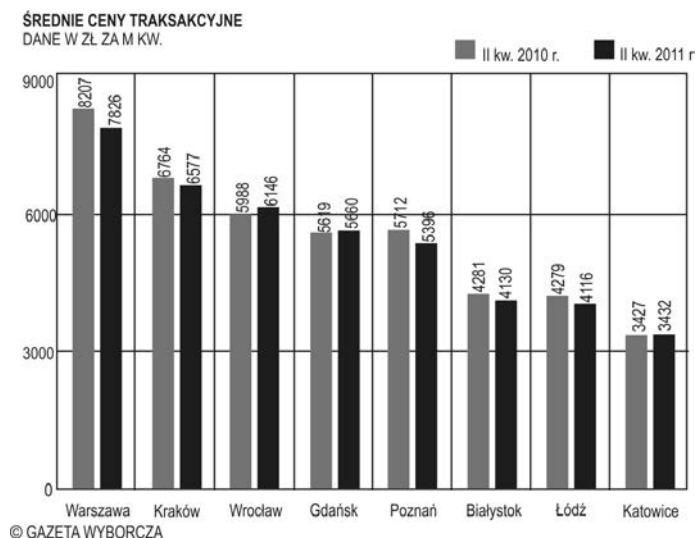


Wykres 22. Poparcie dla UE na infografice

Źródło: <http://infografika.wp.pl> (dostęp: 11.07.2019).

Wykres drugi pokazuje procent osób oceniających pozytywnie sytuację w ich kraju (wykres 22). Ponowne odwrócenie kolejności i zastosowana kolorystyka wskazują na ogromny wzrost w większości państw. Tymczasem legenda mówi co innego: kolumny zielone odpowiadają odpowiedziom udzielonym w 2007 roku, a niebieskie – tym udzielonym sześć lat później.

Inny przypadek dotyczący kolejności słupków widoczny jest na wykresie 23, stanowiącym wizualizację do artykułu o cenach transakcyjnych w latach 2010 i 2011 w największych miastach w Polsce. Na wykresie przedstawiono średnie ceny z drugich kwartałów, wyrażone w złotych za metr kwadratowy.



Wykres 23. Przykład niepoprawnego zestawienia i kolejności

Źródło: P. Biecek, *Odkrywać! Ujawniać! Objaśniać!...*, s. 149.

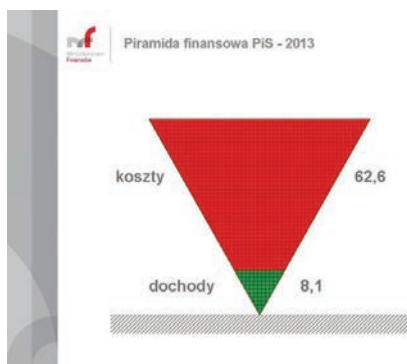
Podobnie jak w poprzednim przykładzie (wykres 22), odwrócono tutaj domyślną kolejność i jako pierwsze pojawiły się informacje dotyczące 2010 roku (wykres 23). Pomijając samą kolejność kolumn i przypisanych im lat, należy dostrzec, że autorzy wykresu ułożyli miasta w taki sposób, by czytelnik odniósł wrażenie znaczącego spadku średnich cen. Najwyższe wartości to jednak dane z Warszawy, a najniższe z Katowic. Wbrew pozorom średnie ceny wcale nie spadają: oprócz nieznacznego wzrostu w Katowicach, podniosły się także we Wrocławiu i Gdańsku. Pierwsze wrażenie zdecydowanie nie jest odzwierciedleniem tego, co działo się w rzeczywistości.

„Punkty” na grafikach

Dla podniesienia atrakcyjności, sprawienia, by infografika wyróżniała się na tle innych materiałów i nie była jedynie „prostym” wykresem, graficy stosują symbole oddające wielkość, wybraną wartość. Może to stanowić problem w prawidłowym odczytaniu i zweryfikowaniu przedstawionych danych. „W sytuacji, gdy nie jest jasne, czy wielkościom odpowiadają pola, czy długości, większość odbiorców intuicyjnie porównuje pola”⁶³, co czasem zaważa o rzetelności przyswojonych informacji.

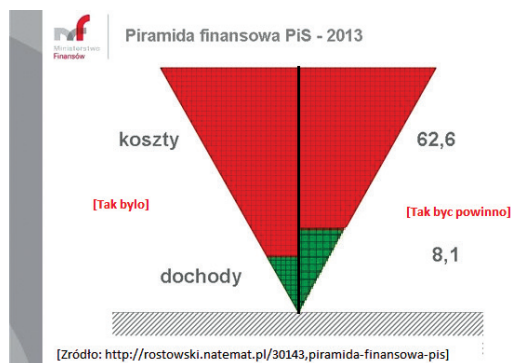
⁶³ Tamże, s. 155.

Pierwsze dwie grafiki zaprezentowane w tym fragmencie zostały szerzej omówione na blogu Fundacji SmarterPoland⁶⁴. Wykres 24 pokazuje materiał przygotowany przez Ministerstwo Finansów, będący przykładem użycia pól. Każde pole wykresu powinno odpowiadać jakiejś wartości. Problem z odbiorciem takiego przedstawienia danych wzrasta, kiedy podzielony na pola kształt jest nierównomierny (inaczej wyglądałoby to w kwadracie czy prostokącie). W takim przypadku istotne proporcje i ewentualne zniekształcenia stają się mniej dostrzegalne. Obok wykresu oryginalnego zamieszczono wersję ilustrującą informacje liczbowe w prawidłowy sposób: „stosunek dochodów do kosztów, zgodnie z podanymi liczbami, wynosi $8,1/62,6 = 0,1293$, podczas gdy po obliczeniu pól figur stosunek tych pól wynosi 2135 punktów/36903 punktów = $0,0578$ ”⁶⁵ (wykres 25).



Wykres 24. Piramida finansowa PiS – grafika oryginalna

Źródło: J. Rostowski, *Piramida finansowa PiS*, <http://rostowski.natemat.pl/30143,piramida-finansowa-pis> (dostęp: 13.08.2019).



Wykres 25. Piramida finansowa PiS – grafika oryginalna porównana z prawidłową wizualizacją

Źródło: P. Biecek, *Odkrywać! Ujawniać! Objawiać! ...*, s. 154.

Powyższy przykład (wykresy 24 i 25) świadczy o sugestywnym zniekształceniu (powiązanym z tytułem wykresu), które powoduje trudność w odbiorze informacji.

Używane w infografikach punktowe sposoby przedstawienia danych nie ograniczają się jedynie do prostych kształtów typu kwadrat czy kółka – w celu

⁶⁴ Fundacja Naukowa SmarterPoland.pl, *Wykres ministerstwa finansów z „lie-factor” przekraczającym 200%*, <http://smarterpoland.pl/index.php/2013/04/wykres-ministerstwa-finansow-z-lie-factor-przekraczajacym-200/> (dostęp: 13.08.2019).

⁶⁵ P. Biecek, *Odkrywać! Ujawniać! Objawiać! ...*, s. 155.

podniesienia atrakcyjności stosuje się także piktogramy, które odpowiadają jednostkom danych. Przykład nieudanej infografiki bazującej na takim koncepcie przedstawiono na wizualizacji z rysunku 11. Ilustracja stanowi uzupełnienie artykułu dotyczącego badania przeprowadzonego przez CBOS po wyborach prezydenckich. Wyniki sondażu zostały omówione szczegółowo w tekście poprzedzającym infografikę, zatytułowaną „Co czuli Polacy po wyborze Andrzeja Dudy?”. Poniżej tego pytania znajduje się zdjęcie prezydenta z rodziną oraz część zasadnicza infografiki – ikony ludzi w różnych kolorach. Każdy z kolorów odpowiada jednej z dziesięciu emocji: od „radości” po odpowiedź „nie mam zdania”. Na tak skonstruowanej grafice bardzo trudno doszukać się proporcji między poszczególnymi odpowiedziami czy stosunku między np. zadowolonymi a rozczarowanymi. Różnicę można by wskazać po policzeniu każdego „ludzika” w danym kolorze. Intuicyjność i kolorystyczny chaos pozwalają sądzić, że jedynym celem tej infografiki jest estetyzacja przekazu.

Powyższe krótkie studium przypadku oraz opisane wcześniej sposoby wpływania na odbiór wykresów i przykłady przekłamań w graficznym przedstawianiu informacji pozwalają wyciągnąć dwa wnioski. Najczęściej pojawiającym się tematem, dla którego projektuje się infografiki, jest polityka. Drugi wniosek dotyczy samego odczytywania tak prezentowanych informacji, bardzo często cechującego się powierzchownością i dającego błędny obraz rzeczywistości. Niewiarygodność przekazu następuje z dwóch powodów. Pierwszym jest nieznanostwo zasad czytania wykresów oraz nieumiejętność interpretacji zależności i proporcji w zestawieniach, a także brak wiedzy o samej charakterystyce poszczególnych rodzajów wykresów (np. o różnicach między wykresem kołowym a słupkowym). Drugi powód wiąże się z celowym działaniem nadawcy, czyli manipulacją. Dotyczy zaplanowanej konstrukcji wykresów i infografik, przemyślanej pod kątem ukrycia wybranych danych lub uwypuklenia tych, na których zależy nadawcy. Manipulować informacjami graficznymi jest znacznie łatwiej niż tekstem. Nadawca ma do dyspozycji bogaty wachlarz możliwości: kolor, perspektywa, wielkość, grubość wykresu czy rodzaj użytych piktogramów. Wciąż niskie kompetencje odbiorców mediów powodują, że manipulacja w infografice pozostaje bardzo częstym zjawiskiem. Co tym bardziej niepokojące, może się pojawić nawet w oficjalnych materiałach rządowych, nie tylko w artykułach dziennikarskich.

Wśród obecnych w mediach wykresów największą podatnością na przekłamanie i występowanie nieprawidłowości cechuje się wykres kołowy, który prezentuje informacje nieefektywnie, a jego forma przysparza trudności przy porównywaniu różnic między przedstawianymi danymi (słabo dostrzegalne są różnice w kątach wycinków). W zasadzie jedyna mocna strona tego typu wykresu to jego sumowanie się do 100%.

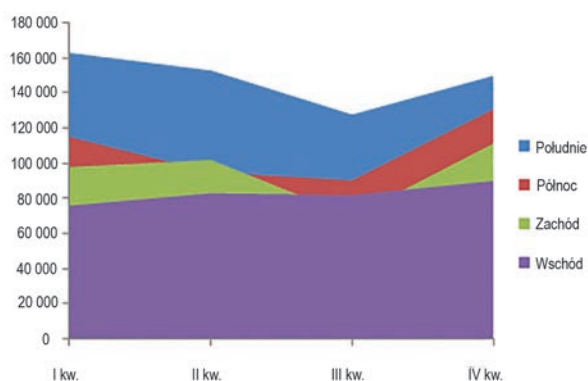
CO CZULI POLACY PO WYBORZE ANDRZEJA DUDY?



Radość Zadowolenie Obojętność Zaskoczenie Rozczarowanie Obawa
Niepewność Zniechęcenie Złość Nie mam zdania

Rysunek 11. Grafika dołączona do artykułu Wyborcza.pl

Źródło: Wyborcza.pl, *Co trzeci Polak zadowolony z wyniku wyborów*, http://wyborcza.pl/1,75248,18131787,Co_trzeci_Polak_zadowolony_z_wyniku_wyborow__Coraz.html (dostęp: 13.08.2019).

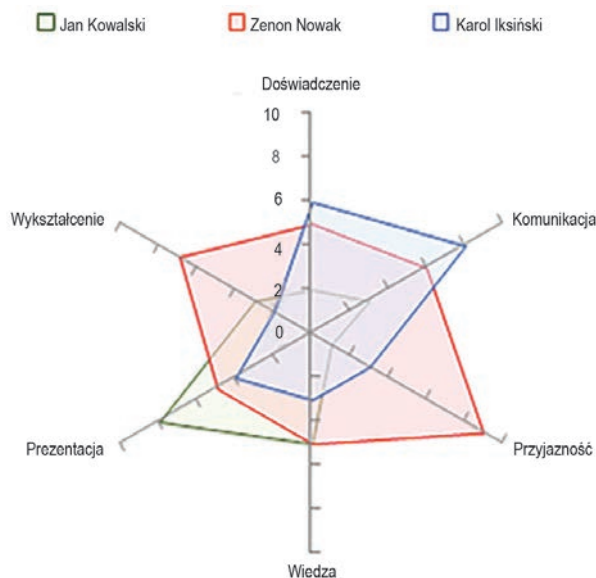


Wykres 26. Sprzedaż w 2010 r. w regionach – wykres powierzchniowy

Źródło: B. Czapiewski, *Wykresy, których należy unikać*, <http://skuteczneraporty.pl/blog/wykresy-ktorych-nalezy-unikac/> (dostęp: 13.08.2019).

Innym typem wykresu, który wprowadzie nie został zanalizowany powyżej, ale na który również trzeba zwrócić uwagę, jest wykres powierzchniowy (wykres 26). Wartości prezentuje się w nim na polach powierzchni, co, podobnie jak w przypadku wykresu kołowego, sprawia trudność w odbiorze i prawidłowej ocenie. Na wykresie 26 ukazano także częsty błąd w tego typu formach, czyli nakładanie się na siebie serii danych.

Ostatni wykres, który należy wymienić w kontekście „ryzykownych” i trudnych w odbiorze, to wykres radarowy (wykres 27). Choć często bardzo efektowny (używany np. w CV), jest wyjątkowo trudny, jeżeli chodzi o porównanie przedstawionych na nim danych, i częściej służy estetyzacji przekazu aniżeli efektywnej prezentacji danych.



Wykres 27. Wykres radarowy

Źródło: B. Czapiewski, *Wykresy, których należy unikać*, <http://skuteczneraporty.pl/blog/wykresy-ktorych-nalezy-unikac/> (dostęp: 13.08.2019).

Powyższa egzemplifikacja nieprawidłowości w wykresach oraz związanego z nimi ryzyka manipulacji przekazem i danymi w żadnym stopniu nie miała na celu ukazania szkodliwości posługiwania się tego typu formą komunikacji. Chodziło o nakreślenie słabych stron i potencjalnych nadużyć w łączeniu danych z wykresami statycznymi, ale wciąż jest to jedna z najwydajniejszych i najskuteczniejszych form przekazu skomplikowanych informacji i danych liczbowych.

1.7. Zasady poprawnego konstruowania wykresów i prezentacji danych

Unikanie opisanych wcześniej nieprawidłowości i błędów w wykresach wydawałoby się przepisem na idealną graficzną prezentację danych. Nie ma jednak szablonu, który sprawdzałby się w przypadku każdej wizualizacji danych czy analiz statystycznych, choć prawidłowe przygotowanie takich grafik opisuje Przemysław Biecek⁶⁶. Przed wymienieniem poszczególnych elementów udanej wizualizacji porządkuje terminologię, definiując wykres jako reprezentację danych, składającą się z elementów (słupki, koła, kolumny). Danymi mogą być zmienne (cena, zarobki, prędkość etc.), opisane w skalach. Sam proces wizualizacji polega zaś na przedstawieniu danych za pomocą cech elementów wykresu. Tymi cechami może być kolor elementów, ich wielkość, kształt, grubość, liczba powtórzeń (np. długość słupka symbolizująca wysokość ceny wybranego produktu). Sposób kodowania cech nie jest stały, więc bazując na tych samych danych, można je przedstawić na różne sposoby.

Autor wymienia osiem istotnych elementów, na które trzeba zwracać uwagę podczas wizualizacji danych. Są to:

- skale pomiarowe – czyli właściwy dobór skali dla odpowiedniego kodowania określonych zmiennych (kolejności, ilorazów);
- długość i szerokość – odnoszące się do często wykorzystywanych wykresów paskowych i słupkowych, mają służyć właściwym ukazaniu zmiennych, np. unikaniu nieregularnych i krzywych kształtów na rzecz proporcjonalnych długości słupków do prezentowanych wartości;
- liczby – choć rzadziej stosowane od długości, również cechują się precyzją w przedstawianiu danych (np. liczba wystąpień danego obiektu na wykresie reprezentuje odpowiadające mu dane);
- pola – element atrakcyjny wizualnie i przykuwający uwagę, jednocześnie mniej dokładny ze względu na trudność trafnej oceny wielkości pól przedstawionych figur np. w porównaniu do długości słupków;
- kąty – obecne w wykresach kołowych, wykorzystywane do przedstawiania zależności między liczbami, z wyraźnie widoczną całością (pełen kąt), ale utrudniające odczytanie dokładnych zależności;
- pozycja i położenie – dotyczą ułożenia elementów na osi, na którym skupia się odbiorca, analizując odległości (należy zwracać uwagę zarówno na jednostki osi, jak i na sam wybór zakresu wartości);

⁶⁶ Tamże, s. 116–142.

- natężenie i nasycenie – są często istotnym elementem na mapach i kartogramach; odnoszą się do kolorystycznego nasycenia i jasności konkretnych danych, które to cechy z jednej strony determinują atrakcyjność wizualną, z drugiej strony mogą ułatwić odczytanie zmiennych i pozwalają na bezpośrednie porównywanie obszarów leżących najbliżej siebie.

1.8. Funkcja koloru w wizualizacji danych

Najważniejszym elementem grafiki statycznej, który ma największy wpływ na odbiorców, jest właściwy wybór palety kolorów⁶⁷. Kolor stanowi czynnik zasadniczy, ponieważ zwraca uwagę i nakierowuje ją; pozwala podkreślić graficznie wybrane fragmenty, które jako pierwsze przykują wzrok podczas wyszukiwania informacji.

Nie brakuje podręczników, słowników i poradników dotyczących znaczenia kolorów, ich symboliki, a także zastosowań (np. jakie kolory stosować w reklamie, jakie w logotypach etc.). Taki stan rzeczy wynika z tego, że ludzie automatycznie i niejako podświadomie przypisują barwom znaczenia (oczywiście wpływ mają na to także uwarunkowania kulturowe, geograficzne czy światopoglądowe).

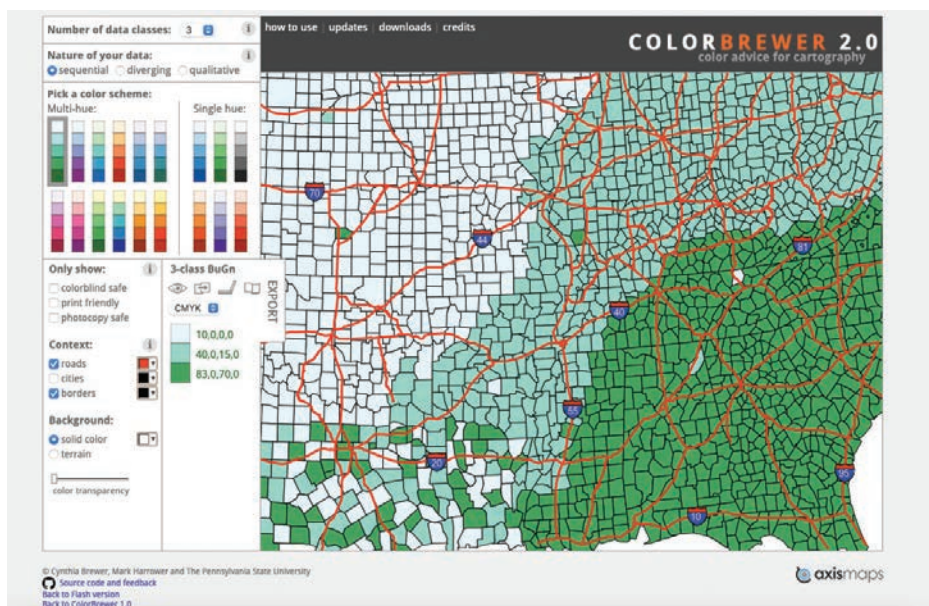
Kolor ma w grafice statycznej bardzo dużą „moc” i siłę perswazyjną. Wizualizacja danych nie może być jednak zdominowana samym kolorem, gdyż stanowi on jedynie środek pomocniczy, sam w sobie nie niesie żadnej informacji. Ponadto może zostać zniekształcony przez takie czynniki jak rodzaj nośnika (rodzaj papieru, rozdzielczość i odmiana monitora lub projektora) czy technika nadruku.

Dobór barw może ułatwić lub utrudnić odbiór, ale ich użycie wydaje się konieczne w obecnych czasach. Wykres czarno-biały, pomijając małą atrakcyjność, mniej efektywnie ukaże różnice między poszczególnymi zjawiskami. Kolorowe grafiki wyglądają nowocześniej i bardziej profesjonalnie, są wizualnie korzystniejsze dla gazetowego czy internetowego layoutu. Wybór kolorów nie powinien być jednak przypadkowy, a cała grafika „przeładowana” kolorami. Czasem wystarczy jeden dla wyróżnienia żądanego elementu wykresu lub kształtu, ale nie może swoją intensywnością lub barwą wpływać na optyczne powiększenie powierzchni obiektów (np. wyjątkowo intensywny kolor na tle delikatnych pasteli). Kolory przydają się również w procesie graficznego grupowania danych, dzieląc wykres na części i różnicując struktury czy podobne do siebie grupy.

Choć kolorom przypisuje się znaczenie i symbolikę, nie ma określonych wytycznych lub stałych palet barw używanych w wykresach czy infografi-
kach, a przynajmniej nie tak kompleksowo opisanych, jak w przypadku map

⁶⁷ Tamże, s. 68.

i infografik na nich opartych. Cynthia Brewer opracowała projekty typografii i kolorystyki do stosowania w mapach, wyróżniając trzy grupy skal: sekwencyjne, rozbieżnych kolorów i jakościowe⁶⁸. Sekwencyjne dotyczą równomiernego rozpięcia skali między dwoma kolorami, rozbieżne umożliwiają określenie skrajnych i pośrednich wartości dzięki kontrastowym barwom, natomiast jakościowe pozwalają na rozróżnienie określonej liczby wartości bez naturalnego porządku (np. skali odcieni)⁶⁹. W zależności od wyboru schematu kolorystycznego twórca infograficznej mapy może przedstawić różną liczbę wartości/kolorów, trzeba jednak pamiętać, że człowiek nie jest w stanie zapamiętać i rozróżnić więcej wartości niż zawierających się w przedziale od 7 do 10⁷⁰. Ponadto dobra wizualizacja danych, opierając się na kolorach, powinna zawierać jedynie ich stonowane odcienie, intensywne męczą bowiem oczy i samego odbiorcę, uniemożliwiając skupienie się na materiale.



Rysunek 12. Widok narzędzia Colorbrewer

Źródło: <http://colorbrewer2.org/> (dostęp: 13.08.2019).

⁶⁸ Wikipedia.org, *Cynthia Brewer*, https://en.wikipedia.org/wiki/Cynthia_Brewer (dostęp: 27.07.2019).

⁶⁹ P. Biecek, *Odkrywać! Ujawniać! Objasniać! ...*, s. 82–84.

⁷⁰ Tamże.

Na powyższym rysunku widoczny jest fragment narzędzia Colorbrewer 2.0 (rys. 12), oferującego do wyboru trzy rodzaje danych, a także schematy kolorów (od odcieni pojedynczych po wielokolorowe). Warto zwrócić również uwagę na takie możliwości edycji mapy jak podstawienie satelitarnego widoku terenu zamiast jednolitego tła czy też zaznaczenie na mapie dróg i miast.

Barwa, obok kształtu, stanowi najczęściej używaną zmienną dla określania skali nominalnej. Pomaga ustalić przynależność obiektów do grupy, różnicując wartości. Na wybranych typach wykresów kolor bywa ważniejszy od kształtu – „kolory na wykresie kropkowym pełnić mogą rolę wyróżnika ułatwiającego szybkie dostrzeżenie skupisk [...] kształty nie przykuwają uwagi tak bardzo, mogą więc sprawdzić się do przedstawiania relacji »drugoplanowych«”⁷¹.

Kolor, podobnie jak pozostałe elementy wykresu/infografiki, umieszczany zostaje w kontekście. Tak jak inne aspekty percepcji wizualnej nie jest postrzegany samodzielnie, a w odniesieniu np. do otaczających go obiektów lub reszty kolorów. Dwa obiekty w tym samym kolorze, ale sąsiadujące z innymi odcieniami, mogą się wydać zbyt niepodobne do siebie nawzajem, co zakłóci przekazywaną informację. Dla uniknięcia tego typu problemów tło wykresu zwykle tworzy się w jednym kolorze, bez używania różnych odcieni.

Najważniejszą zasadą w projektowaniu graficznym i posługiwaniu się kolorami jest unikanie jałowej estetyzacji, a używanie ich tylko dla konkretnych celów komunikacyjnych, dla osiągnięcia jakiegoś celu, nigdy dla udekorowania obrazu. Służy to temu, aby nie odwracać uwagi odbiorcy od najważniejszej rzeczy w graficznym przekazie – danych.

Jak wspomniane zostało wyżej, kolor posiada dużą wartość perswazyjną. Odpowiednie jego natężenie (wysoka intensywność i ciemniejszy odcień) pomaga wyróżnić ważniejsze dane. Aby pomniejszyć znaczenie innych elementów, stosuje się jasną paletę barw⁷². Dodatkowe ozdobniki i elementy wykresów, niezwiązane z danymi, powinny być projektowane w kolorach jasnych, przezroczystych, o niskim znaczeniu wizualnym; elementy takie jak linie osi, obramowania zwykle są koloru jasnoszarego, a tło bardzo często pozostaje białe.

Tak jak w przedsiębiorstwach wnikliwa analiza i interpretacja danych stanowią bardzo istotny element w procesach decyzyjnych, tak w mediach odpowiednie przedstawienie, dobór środków i wybór danych przeważają o sukcesie materiału lub zakwalifikowaniu go do „chartjunk” (wykres śmieciowy). Dzięki wizualizacji można lepiej zrozumieć dane oraz zapoznać się ze złożonymi problemami bez konieczności czytania rozbudowanych opisów tekstowych.

⁷¹ Tamże, s. 137.

⁷² Zob. B. Czapiewski, *Dziewięć zasad skutecznego użycia koloru*, <https://skuteczne-raporty.pl/blog/dziewiec-zasad-skutecznego-uzycia-koloru/> (dostęp: 17.08.2019).

Skuteczna i efektywna graficzna reprezentacja danych podlega pewnym normom i standardom, budowanym w oparciu o takie czynniki jak przyzwyczajenia odbiorcy czy uwarunkowania fizyczne (wzrok czy percepcja kolorów). Praca nad materiałem wizualizującym dane powinna koncentrować się na trafnej selekcji danych w równym stopniu, co na ich przedstawieniu w sposób pozwalający intuicyjny odbiór.

Prezentowanie określonych tematów i zjawisk wymaga użycia wykresów zgodnie z ich przeznaczeniem. Wykres kołowy nadaje się do porównywania elementów całości, ale różnych zestawów danych już nie. Podobnie jest z kolorami – stosowanie niewłaściwych lub nadużywanie ich dezorientuje odbiorcę i może zakłócić prawidłową interpretację. Trzeba mieć na uwadze symbolikę kolorów, a także postrzeganie kształtów i odcieni przez osoby z wadą wzroku. Należy również pamiętać, że nadmiar ozdobników i wskaźników na jednym wykresie powoduje wrażenie chaosu.

Proces wizualizacji danych w formie wykresów musi być dokładnie zaplanowany i zaprojektowany, poczynwszy od samej analizy danych, skończywszy na projekcie diagramów, narracji graficznej i docelowej „opowieści” bazującej na danych. Te ostatnie, podobnie jak każda informacja dziennikarska, powinny zostać poprawnie zebrane z wiarygodnego źródła.

Dziennikarskie teksty informacyjne oceniać można według kryteriów odwróconej piramidy. Jeśli spełniają zasady kolejności poruszanych zagadnień, to odbiorca szybko uzna taki materiał za wartościowy lub nie. W przypadku informacji wizualizowanych graficznie trudniej o określenie parametrów, które pozwalałyby na jednoznaczną klasyfikację materiału. Szczególnie jeśli chodzi o wizualizacje interaktywne, gdzie o tym, ile informacji zostanie zauważonych i zrozumianych przez odbiorcę, decyduje on sam i jego predyspozycje odbiorcze. Niniejszy rozdział miał jednak za zadanie wskazać obszary, którym należy poświęcić uwagę, analizując czy projektując wizualizacje, oraz wyjaśnić, na co warto patrzeć krytycznie z perspektywy odbiorcy mediów.

ROZDZIAŁ 2

2.1. Liczby

Liczby są nieodłącznym elementem komunikacji, występując w mediach w ścisłych związkach z wieloma różnymi tematami i tworząc opis rzeczywistości społecznej. W zależności od sposobu użycia pełnią funkcję ilustrującą lub dopełniającą. „W tych formach liczby współdziałają w procesie konstytuowania komunikacyjnej rzeczywistości społecznej”¹. Z perspektywy nauki o komunikacji i mediach stosowanie liczb i ich współwystępowanie z innymi formami nie zostały ujęte w sposób systematyczny. Anette Siemes w swoim artykule prezentuje wyniki badań nad liczbami w komunikacji społecznej². Analiza kontekstu ogólnych funkcji liczb w medialnym opisie rzeczywistości stanowi właściwy punkt wyjścia do scharakteryzowania specjalizacji dziennikarskiej opartej m.in. na danych liczbowych.

Siemes uważa, że aby móc zrozumieć zastosowania liczb w komunikacji, „ważne są podstawy kognitywne oraz społeczne jako kontekst tych zastosowań, ponieważ te podstawy warunkują komunikacyjne możliwości używania i stosowania liczb w komunikacji”³. Definiuje komunikację jako proces, w którym rzeczywistość społeczna jest konstytuowana i generowana, pojedyncze wypowiedzi są zaś jego manifestem (oferty mediów stanowią natomiast miejsce manifestacji komunikacji i samoopisu społeczeństwa⁴). Dodaje, że w rozważaniach nad rolą liczb w komunikacji medialnej istnieją dwa podejścia teoretyczne. W pierwszym sposoby obchodzenia się z liczbami przez odbiorcę są warunkowane jego umiejętnościami kognitywnymi, ujmowaniem pojedynczych

¹ A. Siemes, *Liczby w mediach: rola i funkcja rzeczywistości liczb w komunikacji medialnej – wyniki badania empirycznego*, [w:] B. Bogułębska, M. Worsowicz (red.), *Styl, dyskurs, media*, Łódź 2010, s. 419.

² Zob. tamże, s. 419–431.

³ Tamże, s. 420.

⁴ Por. S.J. Schmidt, *Blickwechsel. Umriss einer Medienepistemologie*, [w:] G. Rusch, S.J. Schmidt (red.), *Konstruktivismus in der Medien- und Kommunikationswissenschaft*, Frankfurt 1999, s. 119–145.

elementów w zbiory za pomocą liczb. Podejście drugie dotyczy umiejscowienia manifestacji odpowiednich form opisu i konceptów w ramach ich społecznego stosowania: „konteksty kulturowe i czasowo-historyczne warunkują oparte na liczbach formy opisu rzeczywistości społecznej”⁵.

Oddziaływanie liczb w komunikacji jest wielowymiarowe. Mogą odgrywać rolę określającą i wspierającą (np. przy opisywaniu rzeczywistości za ich pomocą), funkcjonują także jako narzędzia o właściwościach autorytetu (ponieważ zawierają obietnicę obowiązywania i trafności)⁶.

Siemes za cel badawczy obrała opis relacji liczb z procesem komunikacji, formułując następujące pytania badawcze:

- Jakie funkcje pełnią liczby w procesie konstruowania rzeczywistości w komunikacjach oraz poprzez komunikacje i jak można na poziomie teoretycznym ustrukturyzować oraz wyjaśnić stosowanie i rolę liczb w komunikacjach?
- Jakie wzory (zabiegi) obchodzenia się z liczbami w komunikacjach, które mogłyby wesprzeć taką teoretyczną strukturyzację, dają się rozpoznać?
- Jak są ustrukturyzowane rzeczywistości generowane przy pomocy liczb w badanych ofertach mediów:
 - co (jakie elementy/obszary rzeczywistości komunikacyjnej) jest opisane poprzez liczby?
 - czy liczby mają różną relewancję w ramach odniesień do różnych wymiarów/obszarów rzeczywistości?⁷

Na podstawie badań jakościowych i materiału badawczego sporządziła otwarty opis rozpoznań. Wprawdzie analizowane teksty pochodziły z prasy (niemieckie czasopismo „Der Spiegel”), jednak z pewnością ich wyniki można by odnieść także do Internetu, przy zachowaniu odpowiedniej skali i proporcji (egzemplifikacje liczb w Internecie z pewnością byłyby większe). We wnioskach po przeprowadzonych badaniach autorka stwierdza, że liczby opisują istotne wymiary rzeczywistości, jak: czas i przestrzeń, pieniądz/gospodarka, osoby/grupy osób oraz inne zbiory (jako heterogenna kategoria zbiorcza)⁸.

Przy całej swej precyzji, określając rzeczywistość i wymienione zjawiska, liczby mogą powodować także wrażenie nieostrości (co bywa wykorzystywane przy perswazjach i manipulacjach), która objawia się chociażby w słowach takich jak „około”, „prawie”, „do...”, ale też zwrotach liczbowych wyrażających zbiorczo konkretne wartości: „setki”, „tysiące”. Duże liczby mocniej wpływają

⁵ A. Siemes, *Liczby w mediach...*, s. 421.

⁶ Tamże, s. 422.

⁷ Tamże, s. 423.

⁸ Tamże, s. 424–425.

na wyobraźnię odbiorców – im większe, tym bardziej emocjonalnie do nich podchodzą⁹. Zachodzi również swoisty paradoks, ponieważ liczby mogą być użyte na dwa różne sposoby – do określenia dokładności i jednoznaczności lub niedokładności i nieostrości, zachowując przy tym formę porównującej perspektywy (tworzącej relacje). W swoich wnioskach Siemes zauważa, że liczby rozmywają się w komunikacji:

Komunikacja może czynić liczby niedokładnymi, a one mogą z kolei być łączone z aproksymatywnym konceptem zbiorów. Mogą więc być w komunikacjach, mówiąc metaforycznie, bardzo elastyczne i niedokładne, a jednocześnie nie tracą swojego połączenia z przeciwbieżnymi kontekstami, związanymi z systemem liczb i liczenia¹⁰.

Biorąc pod uwagę powyższe, dostrzegamy rolę dziennikarstwa danych, w którym fundamentem jest odpowiednia selekcja liczb i przedstawienie ich w formie wielowymiarowej nawiązywalności, ograniczającej pole do dowolnej interpretacji (rodzącej niebezpieczeństwo manipulacji) zastosowanymi środkami wizualnymi lub wprowadzeniem tekstowym.

Z liczbami bezpośrednio wiąże się reprezentacja numeryczna, dotycząca wszystkich obiektów nowych mediów, które „bez względu na to czy tworzone od podstaw na komputerze, czy skonwertowane ze źródeł analogowych, są liczbami zapisanymi w postaci cyfrowej”¹¹. Oznacza to, że obiekty takie jak obraz lub kształt można zapisać jako funkcje matematyczne, a następnie, za pomocą odpowiednich algorytmów, poddać obróbce; nowe media cechuje zatem programowalność. Liczby stanowią więc podstawę procesu cyfryzacji. Przekazy nowomediálne tworzone na komputerze są od razu „cyfrowe”, ale wiele z nich powstaje również poprzez konwersję z mediów analogowych. Sam proces cyfryzacji to przekształcenie danych ciągłych na postać liczbową¹².

⁹ Zob. tamże, s. 427.

¹⁰ Tamże, s. 430.

¹¹ L. Manovich, *Język nowych mediów*, tłum. P. Cypryański, Warszawa 2006, s. 92. Warto w tym miejscu również rozszerzyć definicję procesu cyfryzacji o jego dwuetapowość: „Cyfryzacja składa się z dwu etapów: próbkowania i kwantyzacji. Najpierw dane są próbkowane, najczęściej w regularnych odstępach, tak jak w siatce pikseli używanej do przedstawienia cyfrowego obrazu. Częstotliwość próbkowania nazywa się rozdzielczością. Próbkowanie zamienia dane ciągłe w nieciągłe, czyli dane pojawiające się sekwencyjnie: ludzie, strony książki. Następnie próbki są kwantyzowane, to znaczy przypisywane im jest wartość liczbowa ze zdefiniowanego zakresu” (tamże).

¹² Zob. tamże.

2.2. Bazy danych

Dziennikarstwo „bazodanowe”, jak z początku określano *data journalism*¹³, pierwotną nazwę zawdzięcza właśnie bazom danych, stanowiących dla niego jedno z głównych źródeł informacji. Bazę danych definiuje się również w szerszym ujęciu – jako nowy paradygmat rewolucji komunikacyjnej¹⁴. Jej podstawą jest kod cyfrowy, zawierający nieciągłe, oddzielne sekwencje informacji.

Powstającą w ten sposób cyfrową bazę danych Lev Manovich, Margot Lovejoy, Christiane Paul, Michael Rush uważają za „nową symboliczną formę nowego wieku, opozycyjną wobec ustalonego, linearnego i hierarchicznego świata form ciążących ku kulturowemu paradygmatowi narracji”¹⁵.

O tym, że dziennikarstwo danych należy traktować jako część nowych mediów, świadczy chociażby stosowany w nim sposób narracji (co zostanie szerzej omówione później) oraz charakterystyka samej bazy danych – fundamentu *data journalism*.

Po tym jak powieść, a potem film, uprzywilejowały narrację jako główną formę kulturowej ekspresji współczesności, epoka komputerowa wprowadza jej korelat – bazę danych. Większość obiektów nowych mediów nie opowiada żadnych historii; nie mają one ani początku, ani końca; nie występuje w nich żaden rozwój, który tematycznie, formalnie albo jeszcze inaczej zorganizowałby ich elementy w sekwencje. Są one raczej zbiorami indywidualnych części składowych, z których każda ma takie samo znaczenie jak pozostałe¹⁶.

Zanik narracyjnej formy może rodzić wątpliwości natury komunikacyjnej i kulturowej, szczególnie w odniesieniu do mediów tradycyjnych.

Narracyjna forma symboliczna jest podstawową formą realistycznej reprezentacji w literaturze, piśmiennictwie, fotografii, filmie i w ogóle w sztukach wizualnych. Ustanawiała ona sieci względnie stałych relacji pomiędzy ludźmi, przedmiotami i wydarzeniami w czasie i przestrzeni, które to relacje są organizowane w linearne

¹³ Por. L. Olszański, *Media i dziennikarstwo internetowe*, Warszawa 2012, s. 266.

¹⁴ Zob. J. Jastrzębski, *Chaos, baza danych i internetowe gatunki dziennikarskie*, [w:] K. Wolny-Zmorzyński, W. Furman (red.), *Internetowe gatunki dziennikarskie*, Warszawa 2010, s. 32.

¹⁵ Tamże.

¹⁶ L. Manovich, *Język nowych mediów*, s. 333.

albo obrazowe struktury, odczytywane lub postrzegane jako pozostające w jakimś związku z rzeczywistością¹⁷.

Andrew Dewdney i Peter Ride¹⁸ porównali narrację (jako symboliczną formę kulturową mediów tradycyjnych) z bazą danych (odpowiadającą nowym mediom), wyróżniając siedem par przeciwstawnych cech (tab. 1):

Tabela 1. Różnice między mediami tradycyjnymi a nowymi

Narracja	Baza danych
(Media tradycyjne)	(Nowe media)
Reprezentacja	Informacja
Linearność	Nielinearność
Fikcja	Rzeczywistość
Iluzja	Kontrola
Stołość	Względność
Przedmiot	Proces
Autor	Użytkownik

Źródło: opr. własne, na podst.: A. Dewdney, P. Ride, *The New Media Handbook*, s. 265.

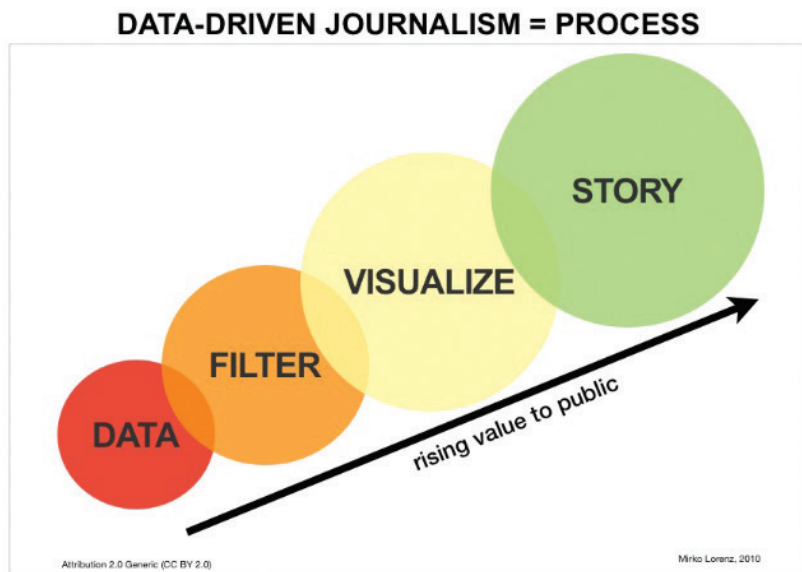
W nawiązaniu do powyższych opozycji można skonstatować, że różnice w sposobach funkcjonowania nowych mediów i tych tradycyjnych są diametralne. Z tego wynika wymóg przystosowania się do nowych form nie tylko odbiorców, ale przede wszystkim dziennikarzy.

Odnosząc opozycję między mediami tradycyjnymi i nowymi do dziennikarstwa, Dan Gillmor twierdzi, że przechodzi ono obecnie „od wykładu do konwersacji”, a pierwszą zasadą konwersacji jest słuchanie. Dotąd dziennikarze słuchali tylko źródeł – teraz muszą nauczyć się słuchać publiczności. Nie chodzi przy tym tylko o dialog, ale o współpracę. Efektem tego współdziałania są nowe rodzaje dziennikarstwa online, takie jak dziennikarstwo bazy danych [...]¹⁹

¹⁷ J. Jastrzębski, *Chaos, baza danych ...*

¹⁸ A. Dewdney, P. Ride, *The New Media Handbook*, London–New York 2010, s. 264.

¹⁹ J. Jastrzębski, *Chaos, baza danych ...*, s. 33.



Rysunek 13. Schemat powstawania materiałów dziennikarstwa danych

Źródło: Wikipedia.org, *Data-driven journalism*, https://en.wikipedia.org/wiki/Data-driven_journalism (dostęp: 30.08.2019).

Aby wytłumaczyć związki i różnice między tradycyjnym pojmowaniem mediów a „bazodanowością” nowych mediów, należy zdefiniować pojęcie bazy danych, a także opisać jej strukturę. Można oprzeć się na spostrzeżeniach Manovicha²⁰, który termin bazy danych przełożył na realia medialne, nie kładąc nacisku na informatyczne punkty widzenia i technologiczne definiowanie tego terminu. W ścisłym znaczeniu jest to bowiem pewien ustrukturalizowany zbiór danych, dostosowany do szybkiego przeszukiwania przez komputer – „różne rodzaje baz – hierarchiczne, sieciowe, relacyjne i zorientowane obiektowo – używają różnych modeli organizowania danych”²¹. Pojęcie bazy danych w odniesieniu do obiektów nowych mediów rozumiane jest bardziej ogólnie – jako pewne modele, zbiory elementów, które użytkownik może nawigować, wyszukiwać, oglądać. Przy porównaniu do oglądania filmów czy czytania utworów fabularnych podstawowa różnica to właśnie sposób doświadczania tego typu „skomputeryzowanych kolekcji”²² – inny model struktury świata. Mimo tej odrębności cyfrowe media

²⁰ L. Manovich, *Język nowych mediów*.

²¹ Tamże, s. 334.

²² Tamże.

okazały się niezwykle podatne na wpływy tradycyjnych gatunków, które już wcześniej miały strukturę bazy danych, takich jak na przykład albumy fotograficzne; zainspirowały one także powstanie nowych gatunków o podobnej strukturze, na przykład biografii opartej na bazie danych. Jednakże miejscem, w którym obserwujemy najbujniejszy rozwój form wykorzystujących bazy danych, jest bez wątpienia internet²³.

Spostrzeżenia autora zostały rozwinięte w dalszej części niniejszego rozdziału. Z powyższego cytatu można domyślać się występowania przenikań, inspiracji i elementów wspólnych (formalnych i strukturalnych) dla nowych i tradycyjnych mediów również w takich gatunkach jak infografika i w samym dziennikarstwie danych. Zasadna wydaje się więc próba odpowiedzi na pytanie, czy w analizowanych zjawiskach dają się odnaleźć warianty gatunków tradycyjnych, czy też są to formy całkowicie niezależne.

Logikę antynarracyjną Manovich dostrzega w otwartości sieci, rozrastających się (a jednocześnie niekompletnych) stronach internetowych, mnogości łącz internetowych. Dodawanie stopniowo nowych elementów ma zatem cechy zbioru, a nie historii, przez co niemożliwe staje się utrzymanie spójnej narracji „ani żadnej innej trajektorii w materiale, który nieustannie się zmienia”²⁴. Mimo wyraźnych przeciwieństw tradycyjnego i nowego pojmowania mediów oraz dziennikarstwa, wydaje się, że narracja w przypadku baz danych (czyli nowych mediów) jest jednak obecna właśnie za sprawą dziennikarzy, którzy z nieliniarnych danych tworzą informację opartą na kontrolowanej narracji (liczby, dane, statystyki). Bazodanowość nowych mediów dała początek omawianemu w tym rozdziale dziennikarstwu danych.

Jak twierdzi Manovich, baza danych sama w sobie nie posiada narracji ani ekspresji i ma charakter nieliniarny. Ideą *data journalism* staje się zatem nadanie jej narracji, wyznaczenie przez dziennikarza początku i końca historii, opowiadanej za pomocą i z wykorzystaniem liczb, oraz zaplanowanie dla niej odpowiedniej struktury. W tych celach konieczna jest selekcja dodawanych elementów, ich kategoryzowanie i porządkowanie – wprowadzenie ładu i rysu tematycznego (narracyjnego) w otwartej strukturze danych. Można to postrzegać jako zatoczenie koła przez historię – w tym przypadku dotyczącą rozwoju mediów; podobnie jak w opisanym przez Manovicha przykładzie albumu fotograficznego, także tutaj nieliniarną formę, charakterystyczną dla Internetu, próbuje się przedstawić w sposób kojarzony z mediami tradycyjnymi.

²³ Tamże, s. 336.

²⁴ Tamże, s. 337.

2.3. Dziennikarstwo danych – definicja pojęcia

Dynamiczny rozwój dziennikarstwa danych łączy się bezpośrednio z epoką cyfrowej informacji i dostępnością danych w przyswajalnej formie. Obecnie jednymi z ważniejszych zadań tej specjalizacji są analizy danych oraz ich efektywne wizualizacje, interaktywne mapy i wykresy. Jej początki sięgają jednak wcześniejszego okresu, sprzed epoki cyfrowej, przed pojawieniem się nowych mediów. Nils Mulvad²⁵ uważa, że ta dziedzina jest dużo starsza i początków upatrywać można już w latach 60.²⁶ Dodaje, że „dziennikarstwo zawsze opierało się na analizie danych”²⁷, zwracając uwagę, że „The Wall Street Journal” narodził się w 1889 roku z codziennego biuletynu, którego podstawę stanowiła analiza notowań giełdowych. Za drugi przykład posłużyć może dziennikarstwo sportowe, w którym „dane zawsze były centralnym punktem analiz”²⁸.

Najważniejsze dla dzisiejszego dziennikarstwa jest odnalezienie się w nowej, cyfrowej rzeczywistości, nadążanie za trendami i zmieniającymi się przyzwyczajeniami odbiorców. Stąd wszechobecne formy konwergencji, powstawanie i rozwój kolejnych hybrydycznych gatunków oraz nacisk na integrację z odbiorcą, multimedialność i interaktywność dziennikarskich materiałów. Wśród tych rozwiązań sytuuje się dziennikarstwo danych – czyli „wykorzystywanie dostępnych w sieci terabajtów do stworzenia wizualizacji, infografik, które w dzisiejszej nastawionej na obraz cywilizacji prawdopodobnie potrafią przemówić lepiej, niż najlepiej nawet napisany tekst”²⁹.

Definicję zjawiska, opierającą się na opisie założeń materiału dziennikarskiego, sformułował Peter Warden, według którego *data journalism* powinien:

- traktować dane jako głównego bohatera artykułu,
- zawierać autorski wkład w interpretację przedstawionych danych,
- opierać się na danych publicznie dostępnych³⁰.

²⁵ Światowej sławy dziennikarz śledczy, wykładowca Duńskiej Szkoły Dziennikarstwa i Mediów, specjalizujący się obecnie w dziennikarstwie danych.

²⁶ P. Pacuła, Nils Mulvad: *Dziennikarstwo danych to współczesny punk*, <https://pl.ejo-online.eu/dziennikarstwo-polityczne/nils-mulvad-dziennikarstwo-danych-wspolczesny-punk> (dostęp: 30.08.2019).

²⁷ Tamże.

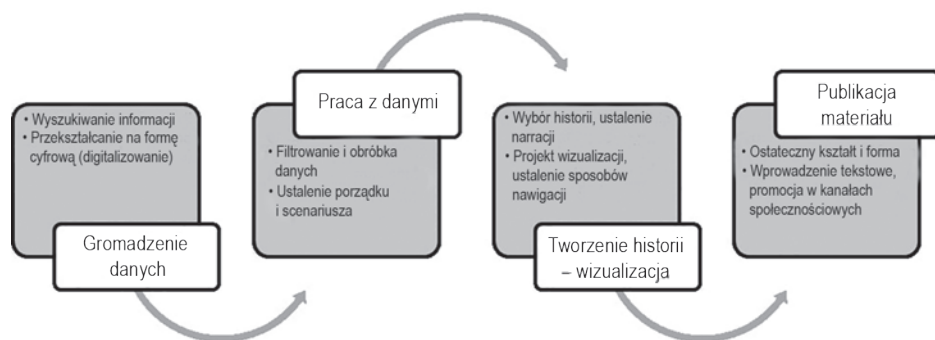
²⁸ Tamże.

²⁹ M. Jędrzyk, *Dziennikarstwo danych – szansa i wyzwanie*, <https://centrumcyfrowe.pl/czytelnia/milada-jedrzyk-o-data-journalism-dla-szkoly-infografiki/> (dostęp: 30.08.2019).

³⁰ P. Warden, *What is data journalism?*, <http://petewarden.typepad.com/searchbrowser/2010/10/page/3/2010> (dostęp: 30.08.2019).

Olszański dodał do tego zestawienia jeszcze jeden punkt, mianowicie że taki materiał powinien „prezentować dane w sposób wygodny i zrozumiały dla użytkownika”³¹. Założenia proponowane przez Wardena budzą jednak wątpliwości, ponieważ dane nie powinny być wartością samą w sobie, a środkiem do opowiedzenia historii, stanowiącej główny walor dziennikarskiej publikacji (i tu ważniejszy wydaje się drugi punkt). Za dyskusyjny można uznać także warunek o dostępności danych, jako że to w zakresie możliwości i kompetencji dziennikarza leży wyszukanie również takich informacji, do których nie dotrze przeciętny użytkownik. Olszański zdaje się bronić koncepcji Wardena, twierdząc, że publiczne dane to warunek przejrzystości i zaufania do autorów, a zainteresowany odbiorca powinien mieć możliwość samodzielnego sprawdzenia postawionej tezy i sposobu rozumowania dziennikarza³².

Dziennikarstwo danych najprościej zdefiniować poprzez sprowadzenie go do procesu zbierania danych i ich wizualizowania w formie infografiki. W rzeczywistości działania w ramach *data journalism* są bardziej złożone, choć często ich efekt faktycznie stanowi wielowarstwowa wizualizacja (interaktywna lub statyczna).



Rysunek 14. Proces publikacji materiałów dziennikarstwa danych

Źródło: opr. własne.

Rozwój dziennikarstwa opartego na analizie danych uważa się za jeden z głównych trendów na rynku medialnym³³. Uprawianie tej specjalizacji wymaga wielu umiejętności, niekojarzonych do tej pory z dziennikarstwem – z zakresu programowania, grafiki czy pracy na arkuszach kalkulacyjnych. Cały wachlarz

³¹ L. Olszański, *Media i dziennikarstwo...*, s. 268.

³² Tamże, s. 269.

³³ P. Pacuła, Nils Mulvad: *Dziennikarstwo danych...*

nie musi jednak zostać opanowany przez jedną osobę – dziennikarstwo danych spowodowało powstanie nowych, okołodziennikarskich profesji dla grafików i programistów, w wyniku czego projekty realizowane są w grupach. Aby zachować efektywną komunikację, każdy członek zespołu powinien mimo wszystko dysponować podstawową wiedzą z każdej z dziedzin (np. dziennikarz wymyśla plan na wizualizację, a programista wie, w jaki sposób przedstawić określoną historię za pomocą interaktywnej mapy).

Niezbędna jest współpraca zespołu ludzi – osoby, która rozumie specyfikę danych, wie, czego potrzebują użytkownicy (a więc dziennikarza), z projektantem aplikacji internetowych, rozumiejącym metody korzystania z nich oraz informatykiem, który potrafi ostatecznie zbudować odpowiednie narzędzie³⁴.

Dziennikarstwo danych jest zatem dziedziną wymagającą, łączącą znajomość rzemiosła dziennikarskiego oraz zasad badawczych z nauk społecznych, takich jak statystyka czy analiza ilościowa i jakościowa. Różnorodna wiedza stanowi konieczność chociażby do prawidłowego zaobserwowania związków między poszczególnymi zjawiskami przedstawianymi w materiale. Błędne wnioski mogą zafałszować rzeczywistość i zaważyć na profesjonalności materiału, a należy pamiętać w tym kontekście o związkach *data journalism* z „dziennikarstwem precyzyjnym”, opartym na dowodach. „Angielskie określenie »data journalism« ma w sobie zakodowaną pewną jakość, bo przecież data, czyli dane, kojarzą się mocno z faktami, z czymś obiektywnym. To sprawia wrażenie, jakby chodziło o »dziennikarstwo faktów«”³⁵. Dane są w tym przypadku narzędziem, sposobem udokumentowania historii, ale rolą dziennikarza wciąż pozostaje umiejętność krytycznego myślenia, opowiadania historii za pomocą danych, analizowania ich pod określonym kątem. Osoby specjalizujące się w *data journalism* same sięgają do danych, zamiast posługiwać się wypowiedziami np. badaczy. Unikają nadinterpretacji, a mając wiedzę na temat danych, mogą samodzielnie analizować materiał, wyciągać z niego wnioski oraz konfrontować swoje odkrycia z eksperckimi.

Dziennikarstwo danych stanowi konsekwencję i kolejny element postępującej cyfryzacji, wpływającej na pojawianie się nowych narzędzi i metod pracy w dziennikarstwie. Obecnie większość informacji podawana jest w formie cyfrowej, w takiej też formie dziennikarz gromadzi je i porządkuje.

³⁴ L. Olszański, *Media i dziennikarstwo ...*, s. 268.

³⁵ P. Pacuła, *Nils Mulvad: Dziennikarstwo danych ...*

2.4. Dziennikarstwo danych w Polsce

Pojęcie *data journalism* można zastąpić polskim, niemal dosłownym tłumaczeniem, jakim jest „dziennikarstwo danych” (*journalism* – dziennikarstwo, *data* – dane). Problem nazewnictwa dotyczy nie samego tłumaczenia, co mnogości pojęć pojawiających się w kontekście tej prężnie się rozwijającej i profesjonalizującej gałęzi dziennikarstwa, opartej na zbieraniu, katalogowaniu i przedstawianiu w formie graficznej dużych ilości danych. W Polsce *data journalism* i *data-driven journalism* stosuje się wymiennie³⁶, natomiast w internetowej encyklopedii pojęcia te są rozróżniane³⁷. Anglojęzyczna Wikipedia definiuje *data journalism* jako dziennikarską specjalizację,

opartą na wzrastającej roli danych liczbowych, które są wykorzystywane w produkcji i dystrybucji informacji w erze cyfrowej. Odzwierciedla zwiększoną interakcję między dziennikarzami (producentami treści) a innymi dziedzinami, takimi jak projektowanie, informatyka czy statystyka³⁸.

Z punktu widzenia dziennikarzy pojęcie to jest tożsame z nakładającymi się kompetencjami z kilku różnych, często bardzo odmiennych dziedzin³⁹. *Data-driven journalism* z kolei odnosi się do samego procesu zbierania, analizowania i filtrowania dużych partii danych (liczbowych i statystycznych) w celu „stworzenia historii” w postaci informacyjnego materiału dziennikarskiego⁴⁰. Trzecim pojęciem, blisko związanym z powyżej opisanymi, jest *database journalism*⁴¹, dotyczący procesu zarządzania informacją, której treść opiera się na ustrukturalizowanych arkuszach

³⁶ R. Gluza, *Newsy z Excels*, „Press” 2013, nr 09, s. 37. W artykule autorka pisze: „tak naprawdę *data journalism* (precyzyjniej: *data-driven journalism*) rozkwitło wraz z Internetem”, sugerując, że obie te nazwy odnoszą się do jednego pojęcia, co, jak wskazano powyżej, nie do końca jest poprawne.

³⁷ Wikipedia.org, *Data journalism*, http://en.wikipedia.org/wiki/Data_journalism (dostęp: 30.08.2019).

³⁸ Tamże, tłumaczenie własne.

³⁹ T. Thibodeaux, *5 Tips for Getting Started in Data Journalism*, <http://www.poynter.org/how-tos/digital-strategies/147734/5-tips-for-getting-started-in-data-journalism/> (dostęp: 11.09.2019).

⁴⁰ P. Szews, *Liczyby, dane i statystyki w dziennikarstwie internetowym*, „Acta Universitatis Lodzensis. Folia Litteraria Polonica” 2014, nr 1(23), s. 247–265.

⁴¹ Wikipedia.org, *Database journalism*, http://en.wikipedia.org/wiki/Database_journalism (dostęp: 29.11.2013).

danych. Oznacza porządkowanie dużych ilości danych i publikowanie ich w jak najbardziej efektywny sposób na różne urządzenia⁴².

W celu usystematyzowania terminów procesy zbierania, analizy i filtrowania obszernych partii danych (*data-driven journalism*) oraz wykorzystywania ich przez dziennikarzy scalono w jedno pojęcie, które łączy zresztą w założeniach definicyjnych wszystkie etapy powstawania materiałów kwalifikowalnych do kategorii *data journalism*, składających się na tę specjalizację. Początkowy brak polskiego odpowiednika dla tej odmiany dziennikarstwa powodował terminologiczne problemy. Próby znalezienia polskiej nazwy dla *data journalism*, podjęte wraz z pojawieniem się w Polsce pierwszych informacji na ten temat, obejmowały takie pojęcia jak „dziennikarstwo liczbowe” czy zbyt dosłowne „dziennikarstwo bazodanowe”. O tym, że najprostsze rozwiązania są często najlepsze, świadczy definicja dziennikarstwa danych autorstwa Phila Bradshawa – określa zjawisko jako dziennikarstwo oparte na danych liczbowych, do których dostęp staje się coraz powszechniejszy⁴³, dlatego też propozycja polskiego terminu „dziennikarstwo danych” jest najbardziej trafna i w pełni oddająca ideę i założenia tej specjalizacji.

2.5. Dziennikarstwo danych jako część nowych mediów

Do pełniejszego scharakteryzowania zjawiska konieczne jest umiejscowienie go w opisach medialnej rzeczywistości (między mediami tradycyjnymi i nowymi mediami) oraz podjęcie próby analizy pod kątem gatunkowym. Bez wątpienia *data journalism* to nowy fenomen, zmieniający narrację medialną, sposoby przekazywania informacji i samo funkcjonowanie mediów. Stanowi ważny składnik nowych mediów przede wszystkim ze względu na interaktywność oraz metody zapisu i przetwarzania rzeczywistości.

Problemy terminologiczne dotyczące nowych mediów związane są z ich bardzo dynamicznym rozwojem i powstawaniem coraz to nowszych form i usług. Sama nazwa „nowe media” pojawiła się na początku lub w połowie lat 90. XX wieku⁴⁴. Najczęściej występowała na zasadzie kontrastu do „starych mediów”, rozumianych jako sekwencyjne, liniowe, analogowe i statyczne,

⁴² Wszystkie przedstawione powyżej definicje pochodzą z anglojęzycznej edycji Wikipedii i są tłumaczeniami autora.

⁴³ P. Bradshaw, L. Rohumaa, *The Online Journalism Handbook: Skills to Survive and Thrive in the Digital Age*, New York 2011, s. 7–9.

⁴⁴ Z. Bauer, *Dziennikarstwo wobec nowych mediów*, Warszawa 2009, s. 210–212.

narzucające konkretny sposób czytania, w którym odbiorca podąża za stałą strukturą narracyjną (wstęp, rozwinięcie, zakończenie)⁴⁵. Nowe media stanowią przeciwieństwo starych niemal w każdym punkcie: są interaktywne, dają odbiorcy możliwość przemieszczania się nielinearnie, umożliwiając wielokierunkową formę przekazu i zmieniając obieg informacji w stronę modeli interakcyjnych⁴⁶. Oprócz tego

charakteryzują się one stosowaniem cyfrowego zapisu i przetwarzania sygnałów [...] integracją różnych sieci i rozproszeniem [...] Informacja w nich jest bardzo łatwo obrabiana, przechowywana, przekształcana, udostępniana, przywracana, a także wyszukiwana i dostępna⁴⁷.

Inną definicję nowych mediów odnajdujemy u van Dijka, który opisuje je jako „zintegrowane i interaktywne media przełomu XX i XXI wieku oparte na kodzie binarnym”⁴⁸. Z kolei Paul Levinson⁴⁹ wprowadza nawet pojęcie „nowych nowych mediów”, określając je jako nowsze niż nowe media, w których:

- każdy konsument jest jednocześnie producentem,
- istnieją duże możliwości personalizacji mediów,
- dostęp do materiałów często jest darmowy,
- istnieje relacja katalityczna mimo rywalizacji.

Natomiast spekulatywne i niesprecyzowane terminy, jak *Web 2.5* czy *Web 3.0* wprowadzone przez Johna Markoffa, dotyczą w większym stopniu sposobu korzystania z sieci, umiejscowienia w czasie konkretnych zjawisk aniżeli ich definicji. Mimo słabej wartości semantycznej liczb (w porównaniu do słów) dla uporządkowania teraźniejszego rozumienia terminu „nowe media” można powołać się na spostrzeżenia Karola Jakubowicza⁵⁰, który wyróżnia trzy generacje nowych mediów:

- 1.0 – czyli nowe media przed okresu konwergencji,
- 2.0 – nowe media powstające za sprawą konwergencji,
- 3.0 – nowe media w pełni zintegrowane i konwergentne.

Kategoria „nowych nowych mediów” cechuje się nieograniczoną elastycznością i swobodą ze strony odbiorcy, dając mu „wolność i władzę, która nie

⁴⁵ Por. A. Adamski, *Media w analogowym i cyfrowym świecie*, Warszawa 2012, s. 92–93.

⁴⁶ Por. T. Goban-Klas, *Pochwała medioznawstwa czyli renesans McLuhana*, „Studia Medioznawcze” 2000, nr 1, s. 22.

⁴⁷ A. Adamski, *Media w analogowym ...*, s. 93.

⁴⁸ J. van Dijk, *The Network Society: Social Aspects of New Media*, London 2006, s. 20.

⁴⁹ P. Levinson, *Nowe nowe media*, tłum. M. Zawadzka, Kraków 2010, s. 11.

⁵⁰ K. Jakubowicz, *Nowa ekologia mediów. Konwergencja a metamorfoza*, Warszawa 2011, s. 60–70.

istniała w rzeczywistości nowych mediów”⁵¹. Levinson postrzega nowe media jako dynamiczne media społecznościowe, przy czym społecznościowy charakter ma mieć kluczowe znaczenie dla ich funkcjonowania. Autor, definiując te media, zwraca uwagę na ich audiowizualność (łączenie pisma, dźwięku, audiowizualności i fotografii), a także na zmianę w postrzeganiu informacji, która tutaj jest celem, a nie formą przekazu⁵². Z kolei R.K. Logan opisuje nowe media przy pomocy czternastu cech i funkcjonalności, które oferują swoim użytkownikom. Są to:

- dwustronna komunikacja,
- łatwość dostępu i upowszechniania informacji,
- ciągłość uczenia się,
- dostosowywanie i integracja,
- społeczność,
- mobilność i asynchroniczność,
- konwergencja,
- interoperacyjność,
- agregacja treści,
- różnorodność i znacznie większy zasięg niż w przypadku mediów tradycyjnych,
- rozmycie podziału na nadawców i odbiorców mediów,
- tworzenie społeczności, których członkowie współpracują ze sobą,
- kultura remiksu, wspomagana przez digitalizację,
- przejście od produktów do usług⁵³.

W nowe media infografika i dziennikarstwo danych wpisują się zarówno warstwą techniczną, obejmującą „proces i pakiet, sortowanie i dopasowywanie, język komputerowy, strukturę danych”⁵⁴, jak i kulturową, na którą składają się „encyklopedia i opowiadanie, fabuła i wątek, kompozycja i punkt widzenia”⁵⁵. Manovich twierdził zresztą, że obie te warstwy się połączą z uwagi na to, że nowe media tworzone są na komputerze i rozpowszechniane przez komputer⁵⁶.

W postrzeganiu przez badaczy dziennikarstwa danych i infografiki jako elementów nowych mediów pojawiają się różnice. Levinson twierdził bowiem, że czynnikiem „wyróżniającym nowe nowe media jest produkcja [...] prawie

⁵¹ P. Levinson, *Nowe nowe media*, s. 15.

⁵² Tamże, s. 17.

⁵³ A. Adamski, *Media w analogowym ...*, s. 98. Zob. także: R.K. Logan, *Understanding New Media*, New York 2010, s. 48–49.

⁵⁴ A. Adamski, *Media w analogowym ...*, s. 96.

⁵⁵ Tamże.

⁵⁶ Zob. L. Manovich, *Język nowych mediów*, s. 115.

tak samo łatwa jak konsumpcja”⁵⁷. Natomiast oba zjawiska stanowią pracochłonne i czasochłonne formy dziennikarskiego przekazu, przez co dostęp do nich często bywa częściowo lub całościowo płatny, w odróżnieniu od innych usług internetowych zaliczanych do nowych mediów (jak np. serwisy społecznościowe).

Dziennikarstwo danych jest profesją zorientowaną na Internet, na co wpływ mają oferowane narzędzia, specyfika tej specjalizacji i dostęp do danych (źródła informacji). Przekaz w mniejszym stopniu stanowi narrację, w większym natomiast bazę danych, ale tak przygotowaną, by była przystępna i zrozumiała dla szerszej publiczności. W tym miejscu można postawić pytanie, czy nie mamy do czynienia z transformacją informacji dziennikarskiej, przekształcającej funkcjonowanie dziennikarzy i uruchamiającej nowy sposób przekazu, opartego na danych liczbowych. Maryla Hopfinger w artykule dotyczącym rekonfiguracji komunikacji społecznej wspomina o pierwszym przypadku, kiedy kultura oparta na oralności przekształcona została w kulturę posługującą się pismem⁵⁸. Podobnie tradycyjnie pojmowany tekst dziennikarski zdominowały liczby, stanowiące fundament i punkt wyjścia w tworzeniu materiałów *data journalism*. Jest to nic innego jak efekt zmiany w komunikacji medialnej, mianowicie rekonfiguracji digitalnej⁵⁹. Dotyczy ona technik cyfrowych opartych na systemie zero-jedynkowym, tworzących „fundament dla przekodowania wszystkich znanych dotąd sposobów i form komunikacji”⁶⁰. Wspólny digitalny mianownik, jakim w tym przypadku są dane, pozwala na zapis tekstów wszelkiego rodzaju w nowym systemie. Jak wykazano w dalszej części rozdziału, system zero-jedynkowy obejmuje cały repertuar komunikacji społecznej, wszystkich otaczających nas zjawisk, pełne spektrum doświadczeń i procesów. Każda informacja, nawet bardzo złożona, może zostać przekształcona w postać cyfrową – taka sytuacja redefiniuje międzyznakowe relacje, ułatwiając bezpośrednie istnienie odmiennych znaków w przekazie (na przykład pozornie odrębnych danych statystycznych).

System binarny został przywołany nie bez powodu – zero-jedynkowa metoda zapisu danych jest podstawą przy segregowaniu informacji w dziennikarstwie danych, zbliżając tę specjalizację do procesu cyfryzacji rozumianego jako

⁵⁷ P. Levinson, *Nowe nowe media*, s. 18.

⁵⁸ M. Hopfinger, *Rekonfiguracja komunikacji społecznej*, [w:] K. Wolny-Zmorzyński, W. Furman (red.), *Internetowe gatunki...*, s. 16.

⁵⁹ Tamże, s. 17.

⁶⁰ Tamże.

przekształcenie różnorodnych sposobów komunikacji – druku, mowy, materiałów audiowizualnych, surowych danych – w uniwersalną formę elektronicznych ciągów binarnych, pozwalających na szybkie przekształcenie danej formy w dowolną inną oraz na przechowywanie, odtwarzanie, manipulowanie i dystrybuowanie każdej z nich z bezprecedensową łatwością poprzez połączone światy komputeryzacji i telekomunikacji⁶¹.

Proces digitalizacji (cyfryzacji) łączy w sobie liczby i bazy danych, a także technologię – „dane nie istnieją tak sobie, wprost przeciwnie – muszą zostać utworzone. Trzeba je zebrać i uporządkować według jakiejś zasady albo stworzyć od podstaw [...] zarejestrować trzeba materiał filmowy i dźwiękowy albo trzeba zdigitalizować już istniejące media”⁶². Przekształcanie tradycyjnych materiałów w formę cyfrową miało swoje początki już w latach 90.: „kiedy komputer zaczął pełnić funkcję uniwersalnej maszyny medialnej, skomputeryzowane społeczeństwa oszalały na punkcie digitalizacji”⁶³. Trend ten przetrwał do dziś – wystarczy przywołać chociażby przykłady skanowania w celu przekształcenia w formę cyfrową archiwalnych ksiąg, raportów, historycznych pism i publikacji.

W procesie digitalizacji wszystkie dane muszą być uporządkowane, zorganizowane i zindeksowane. Komputery przyczyniły się w tym względzie do powstania nowego kulturowego algorytmu (rys. 15).



Rysunek 15. „Epoka komputerów stworzyła nowy kulturowy algorytm: rzeczywistość → media → dane → baza danych”

Źródło: opr. własne, na podst.: L. Manovich, *Język nowych mediów*, s. 341.

„Rozwój sieci – gigantycznego i nieustannie zmieniającego się korpusu danych – dał milionom ludzi na świecie nowe hobby i zajęcie – indeksowanie danych”⁶⁴. Dostęp do technologii i łatwość samego procesu cyfryzacji sprawiły, że w postaci

⁶¹ G. Murdock, P. Golding, *Cyfrowe możliwości a realia rynku: sprzeczności konwergencji*, tłum. M. Łuczak, [w:] G. Murdock, *Media, kultura, ekonomia. Krytyczne pytania. Antologia tekstów*, opr. J. Fras, Wrocław 2010, s. 107.

⁶² L. Manovich, *Język nowych mediów ...*, s. 341.

⁶³ Tamże.

⁶⁴ Tamże.

cyfrową można przekształcić niemal każdą rzecz, proces czy wydarzenie. Wystarczy do tego odpowiednie oprogramowanie komputerowe, indeks liczbowy i umiejętność stworzenia bazy danych. W ten sposób – w postaci liczbowej – danymi można przekazywać w *data journalism* informacje o rzeczywistości, co jest niemal dosłownym przeniesieniem rysunku 15 w medialne realia. Rzeczywistość – wydarzenie, okres historyczny, wybrana dziedzina – za pomocą transformacji przy użyciu odpowiednich mediów zostaje przekształcona w dane cyfrowe, które po odpowiednim uporządkowaniu tworzą bazę danych.

2.6. Computer-Assisted Reporting

2.6.1. Computer-Assisted Reporting a dziennikarstwo precyzyjne

Computer-Assisted Reporting (CAR), czyli w wolnym tłumaczeniu dziennikarstwo wspomagane komputerowo (np. specjalnym oprogramowaniem generującym treści), jest uważane za początek dziennikarstwa danych⁶⁵. Posługiwanie się danymi, aby uzupełnić przekaz i dostarczyć ustrukturyzowaną informację odbiorcy, ma długą historię. CAR było pierwszą zorganizowaną, dostępną metodą wykorzystania komputerów do zbierania i analizy danych w celach informacyjnych. Po raz pierwszy użyto jej w 1952 roku w stacji CBS, by przewidzieć wyniki wyborów prezydenckich. W latach 60. dziennikarze śledczy w USA rozpoczęli analizowanie baz danych i rejestrów publicznych z użyciem metod naukowych. Niezależni dziennikarze – tzw. „dziennikarze publiczni” – za pomocą komputera obalali kłamstwa władzy, ujawniali nieprawidłowości i uchybienia popełniane przez władze i prywatne korporacje. Posługiwanie się danymi przy użyciu komputerów pozwoliło Stephenowi Doigowi, dzięki przeprowadzonym analizom, określić skalę zniszczeń spowodowanych przez huragan Andrew. Ponadto informacje zawarte w zebranych przez niego danych i statystykach wykazały, że ogromu strat winna jest niewłaściwa polityka rozwoju obszarów miejskich⁶⁶.

W latach 70. wprowadzono termin „dziennikarstwo precyzyjne”⁶⁷ (ang. *precision journalism*) do opisu sposobu tworzenia informacji polegającego na

⁶⁵ Zob. L. Bounegru, *Computer-Assisted Reporting and Precision Journalism*, [w:] J. Gray, L. Bounegru, L. Chambers (red.), *The Data Journalism Handbook. How Journalists Can Use Data to Improve the News*, Sebastopol (CA) 2012, s. 18–19.

⁶⁶ Tamże, s. 18.

⁶⁷ Zob. więcej na ten temat w publikacji elektronicznej autorstwa P. Meyer, *Paper Route: Finding My Way to Precision Journalism*, <http://www.unc.edu/~pmeyer/book/Chapter1.htm> (dostęp: 28.08.2019).

korzystaniu ze społecznych i behawioralnych metod badawczych w praktyce dziennikarskiej⁶⁸. Ten typ dziennikarstwa wszedł do „mainstreamowych” mediów stosowany przez specjalistów nauk społecznych i dziennikarzy. Dał początek nowej formie, w której technik wspomaganych komputerowo zaczęto używać przy pisaniu artykułów. Phillip Meyer sugeruje, że to właśnie techniki naukowe dotyczące gromadzenia i analizy danych, a nie techniki literackie, pozwalają na dziennikarską obiektywność i że tylko dzięki nim można osiągnąć skuteczność w dążeniu do prawdy⁶⁹. Dziennikarstwo precyzyjne stanowi niejako reakcję na powszechnie rozumiane słabości dziennikarskie – uzależnienie od informacji agencyjnych, brak obiektywizmu, autorytatywne źródła itd.; ma rozszerzać zestaw narzędzi reporterów, aby tematy, które wcześniej były niedostępne lub dostępne w niewielkim stopniu, mogły stać się przedmiotem dziennikarskiej kontroli, co jest szczególnie przydatne do opisu marginalizowanych grup i ich historii czy ujawnienia pewnych skrywanych nieprawidłowości w danym obszarze. Założeniem tej specjalności jest rozszerzenie tego, co stanowi jakąś wiadomość – od „aktualności wydarzenia” po „sprawozdawczość sytuacyjną” (lub raportowanie trendów społecznych). Niezbędne przy tym okazuje się zastosowanie danych spisowych, ankietowych, dzięki którym dziennikarze są w stanie wyjść poza pojedyncze zdarzenie, zapewnić informacji szerszy kontekst.

2.6.2. Związki między dziennikarstwem danych i CAR

Istnieją bardzo bliskie związki dziennikarstwa danych z dziennikarstwem precyzyjnym. Obie specjalizacje opierają się bowiem na CAR. Podstawową różnicą między *Computer-Assisted Reporting* a dziennikarstwem danych jest jednak sposób funkcjonowania i idea. CAR należy rozumieć jako technikę zbierania i analizowania danych, jako sposób rozszerzania kontekstu (np. reportaży śledczych). Z kolei dziennikarstwo danych obejmuje więcej procesów i dotyczy zarówno zbierania danych i ich późniejszej analizy, jak i opowiadania za ich pomocą całych historii. O ile w pierwszym mamy do czynienia z ukonstytuowaniem pewnych faktów, uzupełnieniem brakujących wątków, o tyle drugie to pełny proces dziennikarski: od danych po wizualizację. Dlatego same dane w przypadku *data journalism* są ważniejsze aniżeli proces ich komputerowego przetwarzania.

W latach, w których pojawiła się metoda CAR, niewielka dostępność danych ograniczała pracę dziennikarzy, która przypominała bardziej polowanie, zbieractwo, szukanie ukrytych informacji w liczbach. Teraz ich zdobycie

⁶⁸ Tamże.

⁶⁹ Tamże.

nie stanowi już problemu, istotniejsze jest filtrowanie (agregowanie) i forma przedstawienia. Technologie cyfrowe i internetowe zmieniają sposób publikowania informacji, często prezentowanych właśnie w postaci wielowarstwowych wizualizacji – efektu pracy dziennikarzy danych. Sama profesja *data journalism* należy do praktyk i narzędzi, które powstały wokół miejsc i usług związanych z danymi, a cytowanie, udostępnianie materiałów źródłowych wpisuje się w hipertekstową strukturę sieci i sposób nawigacji, do jakiego został przyzwyczajony odbiorca dzisiejszych mediów. Cytowanie i wskazywanie źródeł danych są ponadto elementami, za których pomocą ta specjalizacja wzbogaca dziennikarstwo⁷⁰.

Spółeczność dziennikarzy skupionych wokół dziennikarstwa danych w znacznym stopniu różni się od bardziej dojrzałego środowiska CAR. Liliana Bounegru wyraża nadzieję, że te dwie grupy mogą ze sobą współpracować i że więzi pomiędzy nimi będą silniejsze – innowacyjne sposoby prezentacji danych charakterystyczne dla *data journalism* i umiejętność opowiadania za ich pomocą historii połączą się z głębszym, analitycznym i krytycznym podejściem, reprezentowanym przez *Computer-Assisted Reporting*⁷¹.

2.6.3. Przyszłość *Computer-Assisted Reporting*

Uzasadnione byłoby stwierdzenie, że znaczenie terminu CAR wykracza poza posługiwanie się komputerami podczas pracy z danymi. Rozwój technologii i procesy algorytmizacji⁷² mogą wszakże doprowadzić do rozszerzenia kontekstu współpracy człowieka z komputerem, do ograniczenia roli dziennikarza. Według Tadeusza Kononiuka⁷³ funkcja dziennikarza stojącego na straży informacji zanika, co stanowi pośrednią przyczynę rozbudowywania się sieci medialnej oraz antyinstytucjonalnych i antyprofesjonalnych tendencji Internetu. Zasadniczym problemem jest pytanie, czy w świecie cyfrowym i sieciowym XXI w., kiedy w Internecie funkcjonuje coraz więcej darmowej informacji, można utrzymać jakość oraz zachować tradycyjne funkcje dziennikarstwa.

Możliwe, że w przyszłości w ogóle rola dziennikarza zostanie ograniczona do moderowania dyskusji lub wprowadzania haseł do algorytmów zarządzających informacjami – „rozwój sztucznej inteligencji może oznaczać, że już

⁷⁰ L. Bounegru, *Computer-Assisted Reporting...*, s. 22.

⁷¹ Tamże.

⁷² O algorytmizacji przestrzeni internetowej i zmieniającej się roli dziennikarza szerzej w artykule: P. Szews, *Algorytmizacja przestrzeni internetowej*, „Acta Universitatis Lodzensis. Folia Litteraria Polonica” 2015, nr 2(28).

⁷³ Tamże.

w niezbyt odległej przyszłości algorytmy komputerowe zastąpią człowieka na różnych stanowiskach. Jednym z nich, jak się okazuje, może być właśnie dziennikarstwo⁷⁴. Jak pokazały badania przeprowadzone przez Christera Clerwalla⁷⁵, algorytm komputerowy może cechować większą dokładność, wiarygodność i obiektywność niż prawdziwego dziennikarza. Wyniki eksperymentu były zaskakujące: okazało się, że artykuł napisany przez dziennikarza oceniono przede wszystkim jako „dobrze napisany”, „jasny” i „przyjemny do czytania”, a treść wygenerowaną przez oprogramowanie jako „opisową”, „informacyjną”, „dokładną”, „wiarygodną” i „obiektywną”⁷⁶.

Gazeta „Los Angeles Times” poszła jeszcze dalej i jako pierwszy dziennik na świecie wykorzystwała automatycznie wygenerowany tekst, informujący o trzęsieniu ziemi. Dziennikarz i programista Ken Schwencke opracował algorytm, który samodzielnie stworzył ten krótki artykuł. Już po około trzech minutach od opisywanego w nim wydarzenia wygenerowany raport pojawił się na łamach gazety⁷⁷. Był to pierwszy, poza tematyką sportową, przypadek, kiedy sztuczna inteligencja została wykorzystana do generowania artykułów. Stworzone narzędzie nazwano Quakebotem. Pobiera dane z US Geological Survey i wybiera z nich istotne informacje, a wygenerowany tekst trafia do redaktora, który decyduje o publikacji artykułu. Przypadek „Los Angeles Times” nie jest odosobniony. Z podobnego rozwiązania zaczęła korzystać agencja Associated Press⁷⁸, która do tej pory w ciągu trzech miesięcy publikowała 300 raportów,

⁷⁴ J. Moll, *Sztuczna inteligencja może w przyszłości zastąpić dziennikarzy*, <http://tylkonauka.pl/wiadomosc/sztuczna-inteligencja-moze-w-przyszlosci-zastapic-dziennikarzy> (dostęp: 06.09.2019).

⁷⁵ „Clerwall poprosił 45 studentów, aby wybrali i przeczytali jedno z dwóch podsumowań, dotyczących meczu Narodowej Ligi Futbolowej (NFL) – jedno z nich zostało napisane przez dziennikarza LA Times, drugie natomiast – stworzone przez oprogramowanie generatywne. Po przeczytaniu, studenci mieli ocenić artykuł na podstawie treści i wiarygodności oraz przypisać mu jedno z 12 słów: obiektywne, wiarygodne, dokładne, nudne, ciekawe, przyjemne do czytania, jasne, informacyjne, dobrze napisane, użyteczne, opisowe lub spójne”, J. Moll, *Sztuczna inteligencja...*

⁷⁶ Tamże.

⁷⁷ Zob. BBC.com, *Robot Writes LA Times Earthquake Breaking News Article*, <http://www.bbc.com/news/technology-26614051> (dostęp: 06.09.2019).

⁷⁸ F. Diep, *Associated Press Will Use Robots to Write Articles*, <http://www.popsci.com/article/technology/associated-press-will-use-robots-write-articles?dom=PS&loc=recent&lnk=4&con=associated-press-will-use-robots-to-write-articles> (dostęp: 04.09.2019). Uwagę zwraca lead artykułu – „Algorithms may have written more of what you read than you think”.

natomiast dzięki zaawansowanym algorytmom liczba ta może wzrosnąć nawet piętnastokrotnie. Sztuczna inteligencja zajmie się generowaniem analiz wyników finansowych⁷⁹.

Choć obawy o jakość dziennikarstwa i samą rolę dziennikarzy są zasadne, gdy weźmie się pod uwagę powyższe przykłady, to nadal człowiek pełni ważną funkcję „gatekeepera”, rozstrzygającego, jaki materiał wyprodukowany przez algorytmy może zostać opublikowany. Również o tym, jakie dane mają być zbierane i dołączane do raportu, decyduje dziennikarz prowadzący, który dzięki pomocy algorytmów oszczędza czas. Warto jednak mieć na uwadze, że istnieją już całe działy prowadzone z powodzeniem przez algorytmy i specjalne oprogramowanie (np. dział sportowy Yahoo), w związku z czym dziennikarz jako pośrednik informacji zaczyna być odsuwany – często także dlatego, że technologie są wydajniejsze, ich *research* sprawniejszy, a informacje szybciej publikowane. Zagrożeniem w tej sytuacji jest z pewnością wzrost treści internetowych bez pogłębionych analiz, dominacja tzw. treści miękkich kosztem artykułów problemowych oraz pojawienie się „kultury argumentu zamiast weryfikacji faktów”⁸⁰.

Zagrożenie wykluczeniem dziennikarza z procesu twórczego dotyczy treści miękkich, płytkich, prostych w odbiorze i edycji, takich jak wspomniane krótkie informacje o trzęsieniu ziemi czy podsumowanie meczu piłkarskiego. *Data journalism* to materiały zbyt rozbudowane i złożone, by mogły powstać bez uczestnictwa człowieka. Technologia może w tym przypadku jedynie ułatwić proces twórczy, a dziennikarz pozostanie projektantem i narratorem historii wizualnej, opowiadanej za pomocą danych, które pomógł mu przefiltrować komputer.

2.7. Algorytmizacja danych i *big data*

Opisana wyżej rola komputerów jako urządzeń wspierających, zmieniających formy współpracy i relacje na linii komputer-człowiek, ma swoje rozwinięcie w procesach algorytmizacji. W ich ramach dane są wyszukiwane i przetwarzane automatycznie przez systemy komputerowe. Choć jest to oczywiście proces programowalny, stworzony przede wszystkim jako odpowiedź na problem przeciążenia informacyjnego i usprawniający wyszukiwanie informacji, to jednak nie w pełni kontrolowalny. Sposób korzystania użytkowników z treści internetowych powoduje, że istnieje wiele różnych sieci:

⁷⁹ Tamże.

⁸⁰ T. Kononiuk, *Profesjonalizacja w dziennikarstwie. Między modernizmem a ponowoczesnością*, Warszawa 2013, s. 201.

Z punktu widzenia doświadczeń poszczególnych internautów nie istnieje jedna Sieć, ale wiele różnych sieci, których kształt jest wynikiem interakcji użytkowników ze środowiskiem informacyjnym. Nie da się w sposób obiektywny doświadczyć sieci, bowiem korzystając z jej zasobów, każdorazowo tworzymy unikatową sytuację komunikacyjną⁸¹.

Tworzone są w ten sposób pozornie bezwartościowe, nieuporządkowane, początkowo pozbawione sensu treści, przypominające chaos informacyjny, jednak po odpowiedniej analizie mogą dostarczyć wielu informacji na temat użytkowników: ich zainteresowań, kupowanych produktów, relacji społecznych, nastrojów, popularności marek. Algorytmy w tym kontekście nie tylko sugerują użytkownikowi odpowiednie dla niego treści/produkty, ale także przetwarzają informacje o samym użytkowniku i jego działaniach. Jest to zatem proces dwustronny. Algorytmy przetwarzają dane generowane przez użytkownika w sieci, by następnie przetworzyć obecne w nich informacje, przefiltrować je i dostarczyć odbiorcy. Odgrywają zatem szczególną rolę w procesach *big data*, rozumianych szeroko jako przetwarzanie i analiza danych w Internecie wspomagane technologicznie.

Według autorów *Big Data: A Revolution That Will Transform How We Live, Work, and Think*⁸², *big data* wiąże się nie tylko z analizą dużych zbiorów danych, ale również ze sposobem myślenia i postrzegania relacji społecznych, politycznych czy gospodarczych. Dzięki dostępowi do wielu informacji cyfrowych z różnych źródeł i dziedzin może uporządkować i scalić te dane w spójny obraz. O tym, w jaki sposób będą one łączone, jaki sens zostanie im nadany i jakie powiązania uruchomione, bardzo często decydują algorytmy, wpływając zatem na kształt i wizerunek tworzonej w ten sposób rzeczywistości.

Rozwojowi *big data* sprzyja proces datafikacji (*datafication*), definiowany przez Piekarskiego jako tendencja do wytwarzania danych, których uprzednio nie postrzegano jako interesującego źródła informacji, a które często gromadzi się na wszelki wypadek lub w nadziei na usprawnienie konkretnych procesów czy uzyskanie korzyści ekonomicznych⁸³. Jest to poniekąd pochodną rozwoju korporacyjnych systemów zarządzania, w których raportowane i digitalizowane są niemal wszystkie procesy zachodzące w firmie, w tym również rozmowy, wymiany e-maili czy, oczywiście, dane finansowe. Warto w tym miejscu

⁸¹ K. Piekarski, *Kultura danych. Algorytmy wzmacniające uwagę*, Gdańsk 2017, s. 121–122.

⁸² V. Mayer-Schonberger, K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Boston 2013.

⁸³ K. Piekarski, *Kultura danych...*, s. 177.

jednak rozdzielić datafikację od digitalizacji; w wyniku tego drugiego procesu dane zostają przekształcone z analogowych w cyfrowe, natomiast pierwszy dotyczy nadawania wartości liczbowych procesom i zjawiskom, by następnie wykorzystać je w systemach bądź w odpowiednim oprogramowaniu komputerowym. Algorytmy bardzo często urzeczywistniają potencjał zebranych danych, ułatwiając wiele procesów, choć wiążą się też z pewnymi zagrożeniami, m.in. z zamykaniem horyzontów poznawczych, ustalaniem agendy informacyjnej za użytkownika czy swego rodzaju niekontrolowanym „gatekeepingiem”⁸⁴.

2.8. Źródła informacji w dziennikarstwie danych

Tworzenie informacji w dziennikarstwie danych zaczyna się od identyfikacji problemu badawczego, co daje asumpt do typowo dziennikarskich poszukiwań konkretnych danych i odpowiedzi na pytanie „kto monitoruje i archiwizuje dane na temat dziedziny życia, która nas interesuje?”⁸⁵. *Data journalism*, jako specjalistyczna dziedzina dziennikarstwa, wykształcił swoje źródła informacji.

Spośród wielu podziałów źródeł informacji dziennikarskich i różnej ich systematyki Kazimierz Wolny-Zmorzyński⁸⁶ jako powszechny proponuje podział na zastane, czyli źródła gotowe (przedruki z prasy, serwisów radiowych, telewizyjnych, Internetu, informacje od agencji prasowych i *public relations*), oraz wywołane, czyli takie, które dziennikarz zdobywa sam (rozmowy, wiadomości od informatorów, z sond ulicznych i telefonicznych etc.). Ta druga kategoria w głównej mierze kształtowana jest przez umiejętności i zdolności samego dziennikarza, o ich formie i jakości decyduje spostrzegawczość, łatwość nawiązywania kontaktów, wyczucie i intuicja.

Inny podział dziennikarskich źródeł informacji wprowadzili Bogdan Fischer i Weronika Świerczyńska-Głownia⁸⁷, jako kryterium obierając wiarygodność oraz sprawdzalność uzyskanych informacji. To pozwoliło podzielić je na formalne i nieformalne z wewnętrzną systematyką. Na formalne źródła informacji składają się źródła zewnętrzne i wewnętrzne. Do zewnętrznych zaliczono źródła osobowe, instytucjonalne, dokumenty, natomiast do wewnętrznych „formalne zasoby

⁸⁴ Temat zagrożeń rozwijam w artykule *Algorytmizacja przestrzeni...*

⁸⁵ L. Olszański, *Media i dziennikarstwo...*, s. 288.

⁸⁶ K. Wolny-Zmorzyński, A. Kaliszewski, W. Furman, K. Pokorna-Ignatowicz, *Źródła informacji dla dziennikarza*, Warszawa 2008, s. 44.

⁸⁷ B. Fisher, W. Świerczyńska-Głownia, *Dostęp do informacji ustawowo chronionych, zarządzanie informacją. Zagadnienia podstawowe dla dziennikarzy*, Kraków 2006.

archiwalne poszczególnych redakcji oraz wewnętrzną bazę danych”⁸⁸. Wśród źródeł nieformalnych znalazły się źródła osobowe i poznawcze oraz normy, zasady i reguły postępowania dziennikarza, jego wiedza i wykształcenie⁸⁹.

Jak już zaznaczano, dziennikarstwo danych jest specjalizacją dominującą w Internecie, natomiast powyższe źródła informacji wymieniono raczej w odniesieniu do mediów tradycyjnych – przede wszystkim prasy, ale i radia czy telewizji. Dziennikarze danych, jak moglibyśmy nazwać osoby specjalizujące się w *data journalism*, działają w podobnym „środowisku” co dziennikarze internetowi, dlatego warto też przytoczyć źródła informacji dla dziennikarzy pracujących w sieci. Pisząc o pracy dziennikarza internetowego, Leszek Olszański wymienia współpracę z redakcją klasycznego medium oraz agencje informacyjne jako jedno z zewnętrznych źródeł informacji⁹⁰. Anna Barańska w artykule *Źródła informacji dziennikarza internetowego* wprowadza jeszcze jedną kategorię – źródła internetowe, do których zalicza dziennikarstwo obywatelskie, weblogi, fora internetowe i sondaże⁹¹.

Dziennikarstwo danych, specjalizacja rozwijająca się bardzo dynamicznie w Internecie, wykształciło swoje własne źródła informacji, charakterystyczne tylko dla tej formy. Opierając się na podziałach proponowanych dla mediów tradycyjnych, także tutaj można wskazać zastane źródła gotowe oraz te wywołane przez dziennikarza. Większość z dostępnych źródeł to źródła internetowe, choć wyróżnić można także tradycyjne formy „papierowe”, np. archiwa. Przed omówieniem szczegółowego podziału na kategorie i rodzaje źródeł informacji dla dziennikarstwa danych warto wymienić te najpopularniejsze.

Dziennikarstwo danych rozwinęło się w odpowiedzi na coraz większą dostępność danych, publikowanych przede wszystkim w Internecie. Pierwsze katalogi danych, które przekształcały się w portale, powstały (jako rejestry) w Stanach Zjednoczonych, a następne w Wielkiej Brytanii. Początkowo były to przede wszystkim oficjalne kanały informacji, gdzie publikowano dane urzędowe. Obecnie staje się to standardem – coraz więcej państw publikuje dane na dedykowanych portalach, różniących się tylko obszernością i spektrum tematycznym. Wśród nich wymienić można:

- katalogi rządowe,
- katalogi pozarządowe,

⁸⁸ Tamże, s. 45.

⁸⁹ Tamże, s. 57–59.

⁹⁰ L. Olszański, *Dziennikarstwo internetowe*, Warszawa 2006, s. 196–203.

⁹¹ A. Barańska, *Źródła informacji dziennikarza internetowego*, [w:] G. Habrajska, J. Mikosz (red.), *Dziennikarstwo internetowe. Teoria i praktyka*, Łódź 2010, s. 81–90.

- tradycyjne archiwa,
- fora i listy mailingowe/grupy dyskusyjne,
- specjalistyczne portale, serwisy i usługi,
- źródła osobowe,
- partnerzy komercyjni.

Jeśli chodzi o **katalogi rządowe**, to podstawowym założeniem tego typu stron jest promowanie i rozpowszechnianie zamieszczanych tam informacji, także w celach handlowych. Istnieje również metawyszukiwarka – Guardian World Government Data, zawierająca wiele międzynarodowych katalogów danych rządowych, z wydajnym „silnikiem wyszukiwania” wspieranym przez Google. Usługa Guardian dysponuje API (ang. *Application User Interface*), czyli ściśle określonym zestawem reguł i ich opisów, umożliwiającym, w uproszczeniu, sprawną komunikację pomiędzy różnymi systemami lub programami. Dla dziennikarzy pracujących z danymi jest to duże ułatwienie, ponieważ korzystanie z wielu różnych katalogów, udostępniających treści w odmiennych formatach, może być kłopotliwe przy zestawianiu lub porównywaniu danych. Nadanie temu jednolitej formy i opcja wyboru jednego z dwóch formatów umożliwiają porównanie danych z wybranego sektora dla dwóch krajów, miast czy województw.

Oprócz oficjalnych katalogów prowadzonych przez jednostki państwowe istnieje też wiele **katalogów pozarządowych**, za które odpowiedzialne są organizacje pozarządowe oraz sami użytkownicy – społeczności skupione wokół określonej tematyki. Stanowiąc źródła danych, niektóre narzędzia umożliwiają również pracę z nimi, ich obróbkę, a także wymianę informacji i dyskusję dla osób zrzeszonych w grupach specjalizacyjnych. Wiodącym językiem większości serwisów poświęconych danym pozostaje język angielski. Największymi pozarządowymi portalami danych są dedykowane strony Organizacji Narodów Zjednoczonych oraz Banku Światowego, które w następnej części podrozdziału poddano analizie pod względem budowy i wykorzystania.

Zjawisko gromadzenia różnego rodzaju danych liczbowych istnieje od wielu lat – proces ten nasilił się w miarę cyfryzacji i digitalizacji rzeczywistości, jednak **tradycyjne archiwa** papierowe funkcjonują już od wieków. Odpowiednio zachowane dane z archiwów mogą zostać przetworzone w postać cyfrową, nie tylko za pomocą fotografii i ręcznego przepisywania ich treści, co byłoby procesem bardzo pracochłonnym, ale i z użyciem specjalistycznych skanerów lub oprogramowania rozpoznającego tekst.

Fora internetowe zajmujące się wymianą informacji na temat danych i dziennikarstwa danych należy przyporządkować do kategorii forów tematycznych, na których

uczestnicy znacznie rzadziej nawiązują kontakty osobiste [...] przede wszystkim skupiają się na wymianie praktycznych doświadczeń. Wątki przeważnie rozpoczynają się od pytań typu „Jak można to zrobić?”, „Czy to możliwe, żeby...?”, „Gdzie mogę znaleźć...?”, po których następuje seria rzeczowych porad⁹².

Drugą, równoważną odmianą są fora eksperckie⁹³. Aktywni na nich doświadczeni użytkownicy pracują z danymi na co dzień. Specjalistyczne fora internetowe, które zrzeszają osoby zbierające dane i zajmujące się ich wizualizowaniem, to nie tylko miejsca wymiany poglądów, ale także źródła wielu przydatnych informacji: skąd czerpać dane, w jaki sposób ich szukać, jak je katalogować i filtrować. Autorzy wątków często szukają odpowiedzi na nurtujące ich pytania lub rozwiązań np. problemów technicznych. Wartościową bazę wiedzy stanowią np. wątki archiwalne, które w przypadku tego typu usług przechowywane są na serwerach⁹⁴. Jednym z największych forów internetowych jest Get The Data⁹⁵ – forum typu Q&A (pytanie – odpowiedź). Wbudowana wyszukiwarka ułatwia znalezienie odpowiedniego tematu, a także zlokalizowanie tych najbardziej popularnych lub takich zapytań, na które jeszcze nie udzielono odpowiedzi. Każdy wątek jest „tagowany”, czyli oznaczany słowami kluczowymi w celu przyporządkowania kategorii i ułatwienia znajdowania odpowiedzi.

Oprócz forów internetowych platformami wymiany doświadczeń, spostrzeżeń czy informacji są również **listy dyskusyjne**, mające swoje początki w Usenecie – pierwszej, działającej do dziś sieci, założonej w roku 1979⁹⁶. Dyskusja wewnątrz takiej listy przypomina wymianę wiadomości e-mail, tyle że między wieloma internautami. „Aby wziąć udział w zdalnej dyskusji, należy połączyć się ze specjalnym serwerem przechowującym wpisy użytkowników. Wpisy grupowane są w wątki, zaś te w grupy tematyczne”⁹⁷. Listy dyskusyjne (ang. *mailing list*) zbierają specjalistyczną wiedzę środowiska skupionego wokół konkretnego tematu – dla dziennikarzy danych takimi miejscami są m.in. Data-Driven Journalist List czy NICAR-L⁹⁸. Wśród aktywnych użytkowników obu wymienionych list figurują dziennikarze danych, a także

⁹² L. Olszański, *Dziennikarstwo ...*, s. 230.

⁹³ Zob. tamże.

⁹⁴ Tamże.

⁹⁵ Zob. <http://getthedata.org> (dostęp: 28.09.2019).

⁹⁶ L. Olszański, *Dziennikarstwo ...*, s. 33.

⁹⁷ Tamże.

⁹⁸ M. Linksvayer, *Getting Data*, [w:] J. Gray, L. Bounegru, L. Chambers (red.), *The Data Journalism Handbook ...*, s. 113.

osoby związane ze specjalnością CAR, czyli publikowaniem materiałów dziennikarskich częściowo generowanych przez komputery. Jak zauważył Mike Linksvayer, istnieje szansa, że ktoś poruszał już interesujący dziennikarza temat i może podzielić się wskazówkami bądź nawet linkiem do samych danych⁹⁹. Przeglądając wątki innych znaczących list mailingowych, można również odnaleźć takie poświęcone zbieraniu, opracowywaniu czy nawet wizualizowaniu danych.

Pozyskiwanie danych z sieci jest procesem bardzo złożonym, ponieważ Internet składa się z danych przetworzonych binarnie oraz takich, które mogą dopiero zostać przetworzone. Wśród **specjalistycznych postali, serwisów i usług** nie brak przydatnych w dziennikarstwie danych narzędzi, które są zarówno źródłem informacji, jak i metodą na ich przetwarzanie, edycję:

- The Data Hub¹⁰⁰ – serwis powstały dzięki społeczności skupionej wokół Open Knowledge Foundation. Umożliwia bezpłatny dostęp do danych, publikację gotowych zestawów, tworzenie i zarządzanie zbiorami danych, a także aktualizację swoich informacji z grup i zestawów innych użytkowników;
- ScraperWiki¹⁰¹ – specjalistyczna platforma poświęcona danym, tworzona przez wykwalifikowanych specjalistów w tej dziedzinie. Jej narzędzia umożliwiają m.in. eksportowanie danych tabelarycznych z plików PDF.

Pośród **źródeł osobowych** można wymienić kategorię ekspertów – osób związanych z danymi zawodowo i profesjonalnie (statystycy, profesorowie, ludzie z branży). Drugą grupę stanowią osoby potocznie nazywane hakerami (ang. *hacks/hackers*¹⁰²), które w tej kategorii mogłyby reprezentować dziennikarstwo oddolne (ang. *grassroots journalism*). Grupują się w organizacjach zrzeszających członków na czterech kontynentach, a ich misją jest stworzenie sieci dziennikarzy (*hacks*) i technologów (*hackers*), którzy zredefiniują przyszłość informacji. Tak szeroka sieć to szansa na dotarcie do wybranych osób i uzyskanie poszukiwanych danych¹⁰³.

Olszański prócz wspomnianych wyżej źródeł, które nazywa w podobny sposób (tj. instytucje samorządowe, branżowi „regulatorzy”, parlamentarne i samorządowe komisje), wskazuje również **partnerów komercyjnych**, czyli np. banki, w kontekście instytucji dysponujących kompletnymi informacjami

⁹⁹ Tamże.

¹⁰⁰ <http://datahub.io> (dostęp: 29.09.2019).

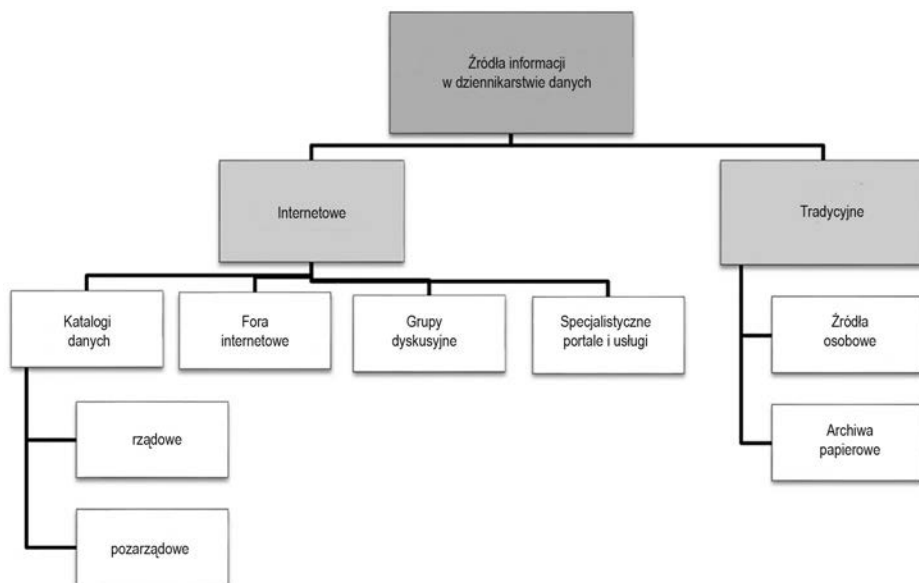
¹⁰¹ <https://scraperwiki.com> (dostęp: 29.09.2019).

¹⁰² M. Linksvayer, *Getting Data*, s. 114.

¹⁰³ Tamże.

o cenach transakcyjnych nieruchomości¹⁰⁴. Takimi podmiotami mogą być też firmy maklerskie, finansowe, analityczne, które takimi danymi „chętnie się dzielą, by promować w mediach swój ekspercki wizerunek”¹⁰⁵.

2.8.1. Systematyka źródeł informacji w dziennikarstwie danych



Rysunek 16. Systematyka źródeł informacji w dziennikarstwie danych

Źródło: opr. własne.

W tabeli (tab. 2) wymieniono największe, oficjalne katalogi danych udostępnianych przez rządy kilku państw: Wielkiej Brytanii, Stanów Zjednoczonych, Australii czy Nowej Zelandii. Warto w tym miejscu wspomnieć także o istnieniu katalogów danych dla wybranych Stanów w USA: Massachusetts, Utah, Michigan, Rhode Island, a także dla miast: Nowy Jork, San Francisco (USA), Vancouver, Toronto (Kanada), Londyn, Lichfield (Wielka Brytania). Rozbudowane bazy danych tworzone są również przez organizacje międzynarodowe, takie jak Bank Światowy czy Organizacja Narodów Zjednoczonych.

¹⁰⁴ L. Olszański, *Media i dziennikarstwo ...*, s. 288.

¹⁰⁵ Tamże.

Tabela 2. Oficjalne źródła danych

Oficjalne kanały rządowe – państwa	
Wielka Brytania	Data.gov.uk
Stany Zjednoczone	Data.gov
Australia	Data.australia.gov.au
Nowa Zelandia	Data.govt.nz
Oficjalne kanały rządowe – regiony, stany, miasta	
Stan Kalifornia	ca.gov/data
Waszyngton (Dystrykt Kolumbii)	data.octo.dc.gov
Stan Massachusetts	mass.gov
Stan Nowy Jork	nystate.gov/opendata
Nowy Jork (miasto)	nyc.gov/data
Londyn	data.london.gov.uk
Edmonton (Kanada)	data.edmonton.ca
Organizacje Międzynarodowe	
World Bank (Bank Światowy)	http://data.worldbank.org
Organizacja Narodów Zjednoczonych (UNdata)	https://data.un.org/

Źródło: opr. własne.

2.8.2. Biały wywiad i źródła *open source*

Wiele z wymienionych w poprzednich podpunktach źródeł informacji można zaklasyfikować do kategorii otwartych, co mogłoby wskazywać, że łącznie należałoby przypisać je do „białego wywiadu” w klasycznym rozumieniu tego pojęcia, tj. jako działalności opierającej się na zdobywaniu informacji ze źródeł jawnych i ogólnie dostępnych. W kontekście danych bardziej odpowiedni wydaje się anglojęzyczny termin „open source intelligence”, znany również jako akronim OSINT, „którego rozwinięcie można przetłumaczyć jako wywiad ze źródeł otwartych lub też wywiad jawnoźródłowy”¹⁰⁶.

Rozwój metod i sposobów pozyskiwania informacji z otwartych źródeł zawdzięczamy amerykańskim służbom specjalnym, które stworzyły pełną

¹⁰⁶ B. Saramak, *Wykorzystanie otwartych źródeł informacji w działalności wywiadowczej. Historia, praktyka, perspektywy*, Warszawa 2015, s. 15.

siatkę pojęciową dotyczącą tego zagadnienia¹⁰⁷. Bazę otwartych źródeł stanowią jawne dane, które mogą pochodzić np. z radiowych, telewizyjnych czy drukowanych nośników informacji, takich jak zdjęcia, mapy czy listy. Dane z tej grupy zazwyczaj są surowe, niedopracowane, nieposegregowane. Drugą grupę stanowią informacje bardziej uporządkowane, znajdujące się m.in. w książkach, raportach, artykułach z mediów drukowanych.

Wracając do OSINF, tłumaczonego jako „wywiad jawnoźródłowy”, warto przywołać definicję zawartą w książce Bartosza Saramaka *Wykorzystanie otwartych źródeł informacji w działalności wywiadowczej*, która opisuje to pojęcie jako „zaplanowane poszukiwanie, selekcjonowanie, wnioskowanie i dystrybucja informacji do określonej grup odbiorczej [...] zgodnie ze zgłoszonym wcześniej, ściśle określonym zapotrzebowaniem (ang. *request for information*)”¹⁰⁸.

Należy podkreślić, że w cytowanej publikacji biały wywiad omówiono pod kątem pozyskiwania informacji wywiadowczych, natomiast metoda ta ma swoje odniesienie również do dziennikarstwa danych. Jest to szczególnie widoczne przy przywołaniu sposobów weryfikowania informacji z wywiadu jawnoźródłowego za pomocą analiz krzyżowych.

2.8.3. Otwarte źródła informacji

Jak zostało wspomniane, biały wywiad korzysta z otwartych źródeł informacji, tzw. *open source*, za które można uznać przykładowo:

- źródła publikowane lub nadawane odbiorcom publicznym (np. treści multimedialne),
- dostępne publicznie na żądanie (np. dane spisowe),
- dostępne publicznie w ramach subskrypcji lub zakupu (np. czasopism branżowych),
- możliwe do zobaczenia lub usłyszenia przez każdego przypadkowego obserwatora,
- udostępnione na spotkaniu otwartym dla publiczności,
- uzyskane poprzez odwiedzenie dowolnego miejsca lub udział w jakimkolwiek wydarzeniu, które jest otwarte dla publiczności¹⁰⁹.

¹⁰⁷ Tamże.

¹⁰⁸ Tamże, s. 20.

¹⁰⁹ „Published or broadcast for a public audience (for example, news media content). Available to the public by request (for example, census data); Available to the public by subscription or purchase (for example, industry journals); Could be seen or heard by any casual observer; Made available at a meeting open to the public; Obtained by visiting any place or attending any event that is open to the public”.

Saramak proponuje inny podział tych źródeł, wyodrębniając dwie grupy: źródła pierwotne i wtórne. „Te pierwsze zawierają oryginalny opis obserwacji, spostrzeżeń czy badań, drugie zaś gromadzą, przetwarzają i prezentują informacje pochodzące ze źródeł pierwotnych”¹¹⁰. Ponadto dzieli kanały informacyjne na naukowe, administracyjne (do których zaliczają się informacje publiczne, rządowe i pozarządowe czy organizacji międzynarodowych), komercyjne (np. raporty publikowane przez firmy) i prezentowane przez media. Jako piąty rodzaj autor wskazuje osoby fizyczne, dostarczające informacje w niesformalizowanych źródłach¹¹¹.

Wśród źródeł informacji właściwych dla białego wywiadu, a tym samym dla źródeł danych mogących być wykorzystanymi w materiałach *data journalism*, Saramak wymienia:

- prasę, jako najistotniejszą podwalinę białego wywiadu, zaliczając do tej kategorii także agencje prasowe;
- szarą literaturę, rozumianą jako publikacje książkowe zawierające np. wiedzę specjalistyczną czy naukową, ze szczególnym zwróceniem uwagi na pozycje i opracowania trudno dostępne, spoza rejestracji bibliograficznej, włączając w to również dokumenty rządowe, akademickie i biznesowe funkcjonujące poza komercyjnym obiegiem wydawniczym;
- radio i telewizję, w kontekście zwiększającego się udziału programów informacyjnych i publicystycznych, a także połączenie redakcji tych mediów z ich internetowymi odpowiednikami, umożliwiające szybszą i pełniejszą analizę przekazu;
- ogólnodostępne rejestry i dokumenty, takie jak Ewidencja Działalności Gospodarczej, Krajowy Rejestr Sądowy czy Monitor Sądowy i Gospodarczy;
- geoinformację, czyli dziedzinę geografii, która gromadzi, przetwarza, przesyła oraz analizuje i interpretuje dane geoprzestrzenne. Narzędzia dostępne w ramach tej kategorii umożliwiają przetwarzanie danych geograficznych i wizualizację ich, co jest ważnym aspektem w kontekście dziennikarstwa danych¹¹².

Należy zauważyć, że liczba otwartych źródeł informacji dostępnych w mediach, szczególnie w Internecie, bardzo dynamicznie wzrasta i zmienia się, na co wpływa rozwój zarówno technologii cyfrowych, jak i samego dziennikarstwa

The Recorded Future Team, *What Is Open Source Intelligence and How Is it Used?*, <https://www.recordedfuture.com/open-source-intelligence-definition/> (dostęp: 19.03.2019), tłumaczenie własne.

¹¹⁰ B. Saramak, *Wykorzystanie otwartych źródeł...*, s. 63.

¹¹¹ Tamże, s. 63–64.

¹¹² Tamże, s. 63–91.

danych. Dziennikarze tej specjalizacji wyszukują coraz to nowsze sposoby legalnego i otwartego pozyskiwania informacji, a wiele analogowych źródeł informacji jest przenoszonych do Internetu, uzyskując formę cyfrową.

Wraz ze wzrostem dostępności otwartych źródeł rośnie także liczba serwisów i portali, które je zbierają, agregują i przetwarzają. W kolejnym podpunkcie analizie poddano dwa przykłady portali internetowych z ogromem otwartych informacji, opartych na danych z całego świata.

2.8.4. Analiza wybranych portali danych jako przykładów źródeł informacji

World Bank

Na oficjalnym portalu Banku Światowego znajduje się „subportal” poświęcony danym, który umożliwia darmowy i otwarty dostęp do informacji na temat rozwoju krajów na całym świecie („World Bank Open Data: free and open access to data about development in countries around the globe”¹¹³). To niezwykle rozbudowany i funkcjonalny serwis, dostarczający wielu bardzo szczegółowych danych, z których większość jest już opracowana i zwizualizowana. Dzięki narzędziom do agregowania informacji można wyodrębnić konkretną dziedzinę, wybrany sektor czy kraj.

Z uwagi na rozbudowaną formę oraz liczbę zintegrowanych podstron i narzędzi mapę portalu zaprezentowano w postaci schematu (tab. 3), by dokładnie przedstawić funkcjonalność i możliwości wykorzystania portalu np. przez dziennikarzy.

World Databank zapewnia wgląd do siedemnastu najbardziej popularnych wskaźników rozwoju na świecie, dostępnych w wielu językach, w tym polskim¹¹⁴. Rozbudowana baza danych obejmuje wskaźniki gospodarcze, społeczne, ekonomiczne i ekologiczne, korzystając ze zbiorów Banku Światowego oraz ponad 30 agencji partnerskich. Informacje, dostępne w formie statystyk i danych liczbowych, sięgają 1960 roku (łącznie ponad 900 różnych wskaźników). Wśród agregujących je tematów znajdują się m.in. zdrowie, gospodarka, zmiany klimatu, edukacja, sektor publiczny i finansowy, handel, ekonomia i wzrost, rozwój społeczny czy nauka i technologie (w sumie 20 kategorii tematycznych).

¹¹³ <http://data.worldbank.org> (dostęp: 29.09.2019).

¹¹⁴ Zob. <http://data.worldbank.org/polish> (dostęp: 29.09.2019).

The image shows the World Bank Open Data portal. At the top, there's a navigation bar with links: By Country, By Topic, Indicators, Data Catalog, Microdata, Initiatives, What's New, Support, and Products. Below this is a language selector: This page in English | Español | Français | العربية | 中文.

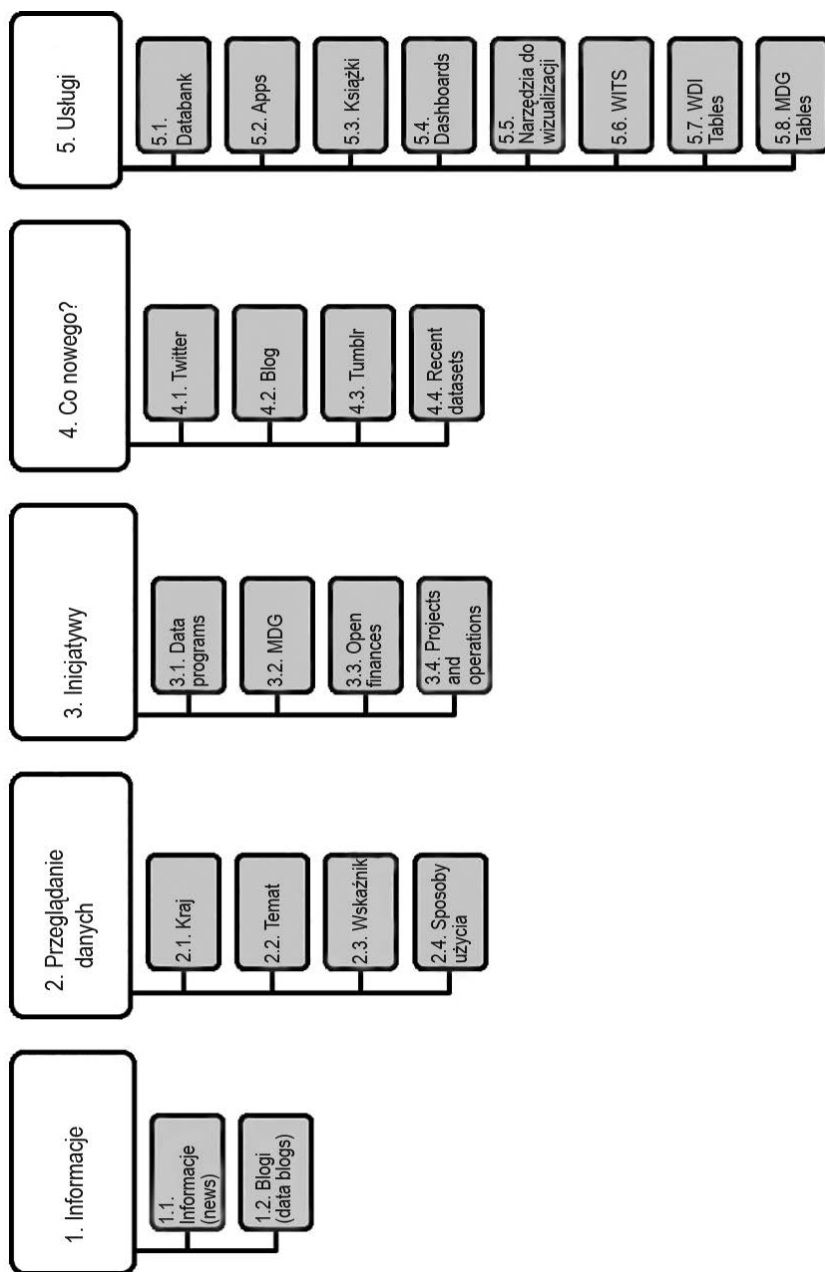
The main content area is divided into several sections:

- World Bank Open Data:** A banner stating "free and open access to data about development in countries around the globe."
- Find an Indicator:** A search bar with the text "GNI per capita, Atlas method (current US\$)" and a "Go" button.
- BROWSE DATA:** A grid of icons for "By Country", "By Topic", "Indicators", and "Use our Data".
- FEATURED:** A list of featured content including "World Development Indicators", "Open Finances", "Projects & Operations", "Open Government Data", and "Toolkit".
- RECENTLY UPDATED:** A list of recently updated content including "IBRD Statement of Loans - Latest ...", "IDA Statement of Credits and Grants ...", and "View data catalog >".
- The World at a Glance:** A section with various indicators and charts:
 - Economy & Growth:** "GDP (current US\$)" with a value of "\$75.59 trillion" for 2013 and a line chart showing growth over time.
 - Health:** "Life expectancy at birth, total (years)" with a value of "71" for 2012 and a line chart showing life expectancy over time.
 - Education:** "Primary completion rate, female (% of relevant age group)" with a value of "91%" for 2012 and a line chart showing completion rate over time.
 - Environment:** "CO2 emissions (kt)" with a value of "33,615,389" for 2010 and a line chart showing emissions over time.
- Visualizing the structure of the world economy and population in one chart:** A large chart titled "Visualizing the Structure of the World" showing GDP per capita and population over time.
- Footer:** A row of three icons: "AidFlows", "Global Consumption Database", and "DataBank".

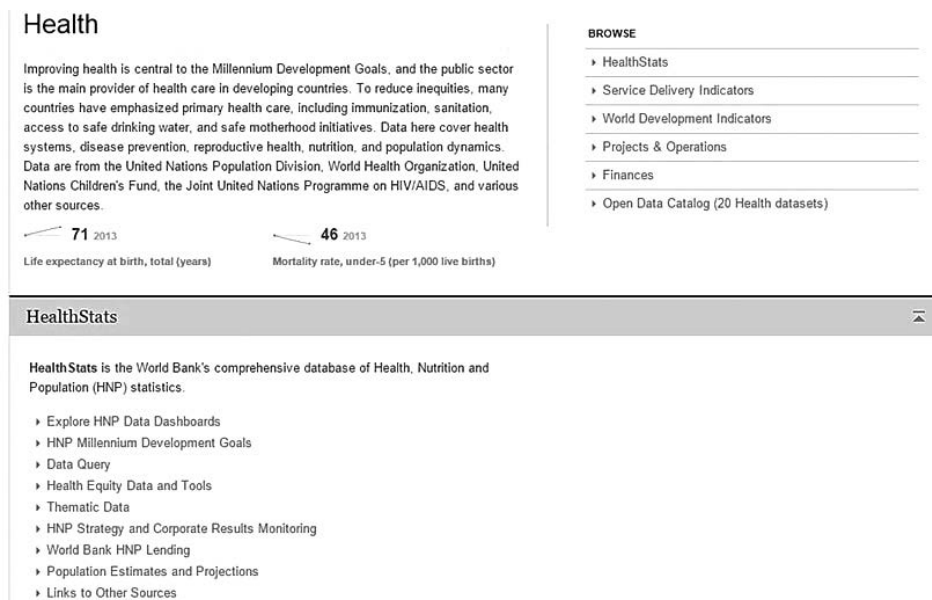
Rysunek 17. Strona główna i wyszukiwarka portalu

Źródło: <http://data.worldbank.org> (dostęp: 29.09.2019).

Tabela 3. Struktura portalu data.worldbank.org



Źródło: opr. własne.



Rysunek 18. Podstrona tematyczna portalu World Bank dotycząca zdrowia

Źródło: <http://data.worldbank.org/topic/health> (dostęp: 29.09.2019).

Po wybraniu kategorii użytkownik jest przenoszony do rozbudowanej podstrony ze statystykami i danymi, które można przeglądać według wybranych krajów, tematów i wskaźników. Interesujące dane użytkownik może pobrać na swój komputer w dostosowanym do potrzeb formacie.

UNdata

UNdata to portal danych uruchomiony przez Departament Spraw Gospodarczych i Społecznych (DESA) Organizacji Narodów Zjednoczonych. Informacje dostarczane są przez globalną społeczność użytkowników. W jednym miejscu znajdują się bazy danych, spośród których korzystający z portalu mogą wyszukiwać i pobierać różne zasoby statystyczne znajdujące się w systemie ONZ. Przy okazji uruchomienia tej usługi podsekretarz generalny Sha Zukang stwierdził:

W ciągu ostatnich 60 lat System Narodów Zjednoczonych zgromadził imponującą ilość informacji. UNdata, opracowane przez Wydział Statystyki DESA jest nowym, potężnym narzędziem, dzięki któremu ten wyjątkowy i wiarygodny zbiór

danych trafi nie tylko na biurka decydentów i analityków, ale także do dziennikarzy, studentów i wszystkich obywateli świata¹¹⁵.

Otwartość tych danych w połączeniu z tym, że pochodzą z całego świata, daje potężną bazę informacji dla dziennikarzy, studentów i wszystkich obywateli zainteresowanych działalnością ONZ. Gromadzone dane i informacje statystyczne zebrano ze wszystkich państw członkowskich, dotyczą różnych obszarów. UNdata łączy duże, międzynarodowe bazy danych w jednym środowisku internetowym, a innowacyjna konstrukcja portalu pozwala na dostęp do nich poprzez przeglądanie serii danych pogrupowanych tematycznie lub poprzez wyszukiwanie słów kluczowych. Jest to istotne ułatwienie ze względu na różne sposoby formatowania i publikacji baz danych przez poszczególne państwa członkowskie. Wśród przydanych funkcji znajduje się np. *Country Profiles*, czyli zaawansowane wyszukiwanie leksykonów dla poszczególnych państw. Bazy danych, tabele i słowniki zawierają ponad 60 milionów zagadnień w bardzo szerokim zakresie tematycznym, obejmującym m.in. przestępczość, edukację, rolnictwo, energetykę, środowisko, zdrowie, rozwój człowieka, przemysł, technologie informacyjne i komunikacyjne, ludność, turystykę czy handel. Portal jest nieustannie rozwijany i rozbudowywany; jak informują twórcy, „w docelowym kształcie UNdata będzie dysponować kompleksowym zestawem międzynarodowych i krajowych baz danych, zapewniających światu natychmiastowy dostęp do bogactwa informacji statystycznych”¹¹⁶. Projekt powstał w celu zapewnienia swobodnego dostępu do światowych statystyk i danych, a także by edukować o ich znaczeniu.

Strona główna zawiera wyszukiwarkę, element centralny portalu, umożliwiający przeszukiwanie bazy danych według wybranych zagadnień lub słów kluczowych. Obok zamieszczono także słowniczek (ang. *glossary*), dział „meta-dane” oraz „API”, czyli dział dla programistów korzystających z interfejsu API (gotowy format danych do dalszej obróbki). Poniżej znajdują się trzy bloki tematyczne: bazy danych (przy których zastosowano podział tematyczny oraz na źródła danych, np. UNESCO), aktualizacje (ang. *updates*), w których strumieniowane są wpisy z oficjalnego konta na Twitterze¹¹⁷, oraz serwisy danych dla poszczególnych krajów (w Polsce jest to serwis Głównego Urzędu Statystycznego¹¹⁸). UNdata umożliwia ponadto sprawdzenie najczęściej wyszuki-

¹¹⁵ UNdata.un.org, *About UNdata*, <https://data.un.org/Host.aspx?Content=About> (dostęp: 29.09.2019), tłumaczenie własne.

¹¹⁶ Tamże, tłumaczenie własne.

¹¹⁷ <http://twitter.com/undata> (dostęp: 29.09.2019).

¹¹⁸ <http://stat.gov.pl> (dostęp: 29.09.2019).

wanych haseł, a także opisów i recenzji portalu. Strona połączona jest także z Twitterem (oficjalny profil) i Facebookiem (w postaci grupy otwartej, zrzeszającej osoby zainteresowane tą tematyką). Autorzy zachęcają również do „sprzężenia zwrotnego” (ang. *feedback*) – użytkownicy mogą odpowiedzieć na dotyczące serwisu pytania, jak: „czy znalazłeś to, co szukałeś”, „czy portal obejmuje główne obszary statystyczne”, „czy uważasz, że strona jest przyjazna użytkownikowi i poprawnie zaprojektowana”¹¹⁹.

2.9. Data blog i data blogging

Data blog w najprostszym rozumieniu to blog powstający przy wykorzystaniu programów do wizualizacji danych¹²⁰, na którym każdy wpis jest wizualną historią stworzoną na podstawie danych zebranych przez autora. Aby pełniej zdefiniować *data blog* i *data blogging*, za punkt wyjścia należy obrać definicję samego blogu, a także wymienić istniejące jego rodzaje i odmiany¹²¹.

2.9.1. Definicja blogu jako tło teoretyczne

Blog w potocznym rozumieniu definiuje się jako internetowy pamiętnik lub dziennik – stąd blogi nazywane są też „weblogami”¹²². Cechy charakterystyczne takiego dziennika, według Wolnego-Zmorzyńskiego, to:

klarowność wywodu połączona z interpretacją zdarzeń politycznych, kulturalnych bądź społecznych – w zależności od tego co jest polem zainteresowania autora. Różni się jednak od gatunków publicystycznych tym, że w dzienniku kładzie się nacisk na poufalość, jaka zachodzi między nadawcą a odbiorcą. Autor wyraźnie nie krępuje

¹¹⁹ Zob. <https://data.un.org/Feedback.aspx> (dostęp: 29.09.2019).

¹²⁰ Por. P. Kozłowski, *Kurs databloga: data blogging, czyli blogowanie z danymi – jak założyć data blog?*, <http://datablog.pl/data-blogging-blogowanie-danymi-zalozyc-data-blog/> (dostęp: 29.09.2019).

¹²¹ W tym podrozdziale powołuję się na fragmenty z artykułu P. Szews, *Mikroblog – odmiana blogu czy oddzielny gatunek*, „Acta Universitatis Lodzensis. Folia Litteraria Polonica” 2013, nr 2(20), s. 271–291.

¹²² Definicję i genezę tego pojęcia szerzej wyjaśnia L. Olszański, *Dziennikarstwo ...*, s. 34. W istocie jest to zbitok dwóch angielskich wyrazów „web” (sieć) i „log” (kojarzony z dziennikiem okrętowym, dosł. zapis, rejestr). Moim zdaniem należy jednak stosować uproszczoną i powszechniejszą formę „blog”, uznając „weblog” za wyrażenie już archaiczne i nieaktualne.

się mówić o sobie i swoich problemach tak, jakby miał nadzieję, że czytelnik jest jego powiernikiem¹²³.

Na potrzeby tego podrozdziału zastąpiono węższe pojęcie weblogu „blogiem”, uznając ten termin za szerszy i bardziej pojemny. Za twórcę nazwy „weblog” uznaje się Johna Barger, założyciela dziennika „Robot Wisdom” (1997). Skrócona wersja „blog”, utrwalona w powszechnym użyciu, została użyta dwa lata później przez Petera Merholza, który podzielił słowo „weblog” na zaimek *we* (my) oraz *blog* (blogujemy). O pojemności i złożoności pojęcia blogu świadczy chociażby liczba dostępnych definicji. W *Słowniku terminologii medialnej* pod redakcją Walerego Pisarka blog został zdefiniowany jako „dziennik (pamiętnik) prowadzony w Internecie, rodzaj osobistej strony WWW, na której autor umieszcza zapiski swoich bieżących przeżyć i przemyśleń”¹²⁴. Opisową definicję przedstawia Katarzyna Kadaj-Kuca: „to często aktualizowana strona internetowa, zwykle prywatna, zawierająca komentarze i linki. Składa się z wielu krótkich wiadomości (postów), zwykle oznaczonych datą i przedstawionych w odwrotnej chronologii”¹²⁵. Określa się w ten sposób również

stronę WWW zawierającą krótkie, rozłączne porcje informacji, zwane artykułami. Artykuły są ułożone w kolejności chronologicznej – od najnowszego do najstarszego. Każdy artykuł jest jednoznacznie identyfikowany przez znacznik zakodowania, a także oznaczony przez trwale łącze, za pomocą którego można się do danego artykułu odwoływać z zewnętrznych stron WWW¹²⁶.

Blog ewoluował z mało znaczącej formy internetowej, w której dominowały treści o niskiej wartości poznawczej i merytorycznej. Z internetowego dyskursu autobiograficznego, w którym „wszystko dzieje się na bieżąco, dzienniki takie mogą być uaktualniane w każdej chwili”¹²⁷, blogi stopniowo przekształciły się w rozbudowane platformy, służące głównie, ale nie tylko, do dzielenia się własną wiedzą i spostrzeżeniami z różnych dziedzin na forum Internetu.

¹²³ K. Wolny-Zmorzyński, A. Kaliszewski, W. Furman, *Gatunki dziennikarskie. Teoria, praktyka, język*, Warszawa 2006, s. 107.

¹²⁴ W. Pisarek (red.), *Słownik terminologii medialnej*, Kraków 2006, s. 18.

¹²⁵ K. Kadaj-Kuca, *Moda na blogowanie. Blog jako komentarz, element polityki redakcyjnej i narzędzie kreacji politycznej rzeczywistości*, [w:] K. Wolny-Zmorzyński, W. Furman (red.), *Internetowe gatunki...*, s. 70.

¹²⁶ C. Doctorow, R. Dornfest, J.S. Johnson, S. Powers, B. Trott, M.G. Trott, *Blogging – przewodnik*, tłum. P. Kresak, Warszawa 2003, s. 28.

¹²⁷ M. Kawka, *Blog jako gatunek dziennikarski – ewolucja i transgresja*, [w:] K. Wolny-Zmorzyński, W. Furman (red.), *Internetowe gatunki...*, s. 65.

Andy Wang z „New York Times” określił blogi jako „internetowe przeglądy informacji”¹²⁸, a samo ich prowadzenie „jednym z najszybciej się rozwijających i najbardziej kreatywnym obszarem internetu”¹²⁹. Stewart Allan zauważa, że „popularne blogi były z reguły poświęcone jednemu konkretnemu tematowi, wypełniając określoną niszę, co umożliwiało tworzenie grupy wiernych odbiorców”¹³⁰. Forma blogu od początku umożliwiała dwukierunkową komunikację z natychmiastowym sprzężeniem zwrotnym. Przekaz charakteryzował się brakiem dystansu między nadawcą a odbiorcą, który miał wrażenie, że znajduje się bliżej wiarygodnych informacji w porównaniu do treści prezentowanych w mediach tradycyjnych.

W miarę rozwoju sieci i popularyzacji blogów rosła ich wartość informacyjna, a samo zjawisko stawało się coraz bardziej złożone. Wolny-Zmorzyński pośród gatunków informacyjnych wyróżnia blogi reportażowe, zaznaczając, że choć pisane są najczęściej przez amatorów, mogą służyć jako ważne źródło informacji dla „profesjonalnych” reporterów¹³¹. Z czasem, w efekcie konwergencji między prasą a Internetem, na popularności zyskało tworzenie blogów dziennikarskich, prowadzonych przez profesjonalistów z danej redakcji. Można wyróżnić co najmniej dwa rodzaje dziennikarstwa blogowego – dziennikarzy piszących zarówno w sieci, jak i w tradycyjnych mediach drukowanych, oraz dziennikarzy sieciowych prowadzących swoje blogi tylko w Internecie¹³².

Potencjał komunikacyjny blogów szybko został zauważony przez firmy marketingowe i zajmujące się praktykami PR. Tego typu witryny stały się także ważnym narzędziem przy kreowaniu wizerunku i wpływaniu na opinię publiczną, co dało początek blogom politycznym, powstałym również w efekcie zmian sposobu uprawiania polityki, w tym zacieśniającej się relacji polityków z mediami czy „mediatyzacji dyskursu publicznego”¹³³. Takie blogi mogą być postrzegane jako „instrument inicjowania debaty, tworzenia wspólnoty komunikacyjnej, pobudzania internautów do podjęcia debaty społecznej, aktywności politycznej. Należy w nich widzieć narzędzie kontaktu pozwalające

¹²⁸ C. Doctorow, R. Dornfest, J.S. Johnson, S. Powers, B. Trott, M.G. Trott, *Blogging – przewodnik*, s. 59.

¹²⁹ Tamże.

¹³⁰ S. Allan, *Newsy w sieci. Internet i dziennikarstwo*, tłum. A. Sadza, Kraków 2008, s. 55.

¹³¹ K. Wolny-Zmorzyński, A. Kaliszewski, W. Furman, *Gatunki dziennikarskie. Teoria, praktyka...*, s. 82–83.

¹³² M. Kawka, *Blog jako gatunek dziennikarski...*, s. 65.

¹³³ J. Biniewicz, *Blog polityczny, czyli o kreowaniu wizerunku politycznego słów kilka*, [w:] M. Jeziński, A. Seklecka, Ł. Wojtkowski (red.), *Nowe media we współczesnym społeczeństwie*, Toruń 2011, s. 217.

kreować wizerunek polityczny, skutecznie nim zarządzać”¹³⁴. O istotności tego narzędzia można się przekonać, śledząc przebieg kampanii prezydenckich w USA i wykorzystanie Internetu, a nawet przyglądając się działalności polskich polityków w sieci (wpisy polityków często cytowane są w mediach, stając się głównym tematem informacji czy głosem w debatach dotyczących ważnych spraw publicznych)¹³⁵.

Blogi można dzielić nie tylko ze względu na tematykę, ale także z uwagi na dwa inne kryteria: „tworzywo” i treść. Upowszechnienie się komunikacji wizualnej i audiowizualnej w Internecie dało początek fotoblogom i videoblogom (zwanymi także vlogami). Rozwój nowych mediów i towarzyszące temu zjawisku procesy konwergencji, zarówno lokowane w przepływie treści między różnymi kanałami, jak również dotyczące procesów technologicznych i sposobów prezentacji treści, zacierają międzygatunkowe różnice oraz tworzą ściślej niezdefiniowane, wielotworzywowe hybrydy gatunkowe. Podobnie jest z blogami – „ogromna popularność i różnorodność blogów nie sprzyja ich wyrazistości genologicznej, jest to – jak się wydaje – zależność asymetryczna: im więcej powstaje ich wariantów, tym mniejsza ich wyrazistość gatunkowa”, pisał Maciej Kawka¹³⁶. Podobny problem dotyczy samej definicji blogu – trudno o wspólny dla całości opis, choć zasadne byłoby podjęcie takiej próby: blogiem można nazwać cykliczne publikacje wypowiedzi w Internecie tego samego autora (lub autorów), w tym samym miejscu w sieci (pod jednym, niezmiennym adresem), o wyraźnie nakreślonym rysie tematycznym i profilu. Charakter wypowiedzi zazwyczaj jest mocno subiektywny, oddający osobowość autora, a każda ma zwykle podobne konstrukcje i wspólne elementy. Wypowiedzi towarzyszą multimedia lub grafika. Oczywiście definicja ta może być rozwijana i uzupełniana o nowe elementy, takie jak szata graficzna, układ strony, zdjęcie autora, nazwę i podtytuł w nagłówku etc.

W artykule *Mikroblog – odmiana blogu czy oddzielny gatunek*¹³⁷ zaproponowałem następujący podział blogów (rys. 19), ze względu na osobę autora, tematykę i przeważające treści.

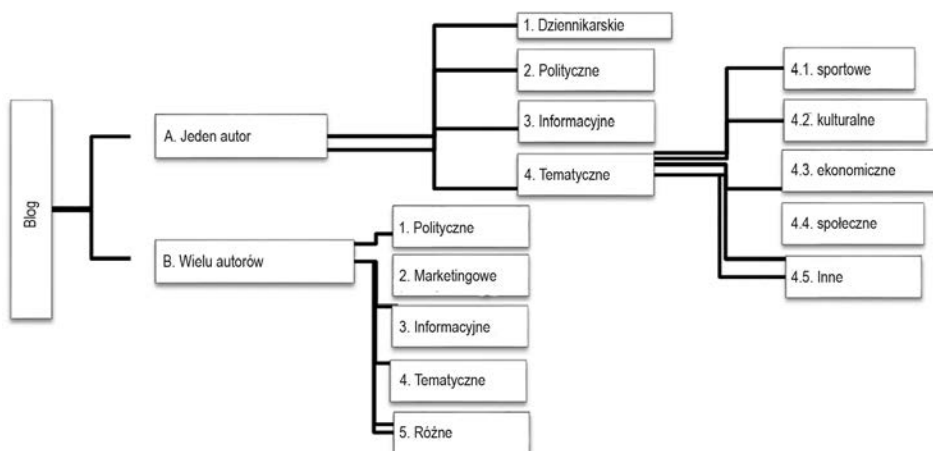
Pierwszym zastosowanym kryterium tego podziału jest rozdzielenie blogów prowadzonych przez jednego i wielu autorów. Choć pierwotnie przyjmowało się, że blog to elektroniczny dziennik/pamiętnik jednej osoby,

¹³⁴ Tamże.

¹³⁵ Więcej na ten temat w artykułach w: M. Jeziński, A. Seklecka, Ł. Wojtowski (red.), *Nowe media...*

¹³⁶ M. Kawka, *Blog jako gatunek dziennikarski...*, s. 64.

¹³⁷ P. Szewski, *Mikroblog...*, s. 275.



Rysunek 19. Podział blogów

Źródło: opr. własne.

obecnie odnajdujemy wiele przykładów blogów prowadzonych przez kilka osób, które tworzą swoisty zespół redakcyjny, zespół ekspertów, specjalistów skoncentrowanych wokół tej samej tematyki (punkt B4, dla którego możemy przyjąć takie samo rozwinięcie jak dla punktu A4; rys. 19). Charakter publikowanych tam treści i sposób ich edycji nie pozwala na nazwanie takich stron np. portalem tematycznym czy serwisem internetowym. Prowadzone przez wielu autorów bywają także blogi polityczne (np. prowadzone przez partie polityczne, instytucje), marketingowe i wizerunkowe, przy czym autorzy często pozostają anonimowi, występując pod jednym pseudonimem (np. nazwa firmy, marki). Takie witryny wykorzystuje się jako narzędzia w kampaniach wyborczych (punkt B1 na rys. 19), pozycjonowaniu i zwiększaniu świadomości marki, nowych produktów, kształtowaniu wizerunku firmy wśród internautów (punkt B2). Blogi marketingowe, obok których można by wpisać także PR-owe, cieszą się umiarkowaną popularnością. Użytkownicy sieci podchodzą do nich sceptycznie, gdyż blogi z założenia nie powinny być komercyjne; od początku uważano je bowiem za niezależne i alternatywne źródło informacji, często bardziej wiarygodne od tradycyjnych mediów.

Popularniejsze są blogi prowadzone przez jednego autora, który często przyciąga czytelników swoją osobowością bardziej niż treścią publikowanych materiałów. Blog w tym przypadku może stać się narzędziem kreacji „bohatera” Internetu, specjalisty w danej dziedzinie czy po prostu osoby, która w sposób bezinteresowny dzieli się swoją wiedzą, przemyśleniami etc. W tej kategorii umieściłem blogi dziennikarskie (A1 na rys. 19), prowadzone przez profesjonalnych dziennikarzy. Tutaj autorzy mogą pozwolić sobie na publikowanie

wypowiedzi niemających możliwości ukazać się w mediach, w których pracują na co dzień. Często blogi stanowią dla dziennikarzy użyteczne narzędzie w skracaniu dystansu między nimi a czytelnikami, budowaniu relacji z odbiorcami czy pomocy w zbieraniu i uzupełnianiu materiałów. Drugą kategorię stanowią blogi polityczne, które można podzielić jeszcze na te prowadzone przez samych polityków (nie przez sztaby wyborcze czy specjalistów od *public relations* – te zostały przyporządkowane do innej kategorii) i przez osoby, które zajmują się komentowaniem politycznej rzeczywistości. Ostatnia kategoria wyróżniona w powyższym podziale to blogi tematyczne, stanowiące większość spośród tych, które możemy odnaleźć w Internecie. Spektrum tematyczne w tej kategorii jest ogromne i w zasadzie trudno znaleźć dziedzinę, której nie został poświęcony ani jeden blog w sieci.

W powyższym schemacie nie umieszczono fotoblogów i wideoblogów, które tworzą odmianę gatunkową blogu ze względu na typ publikowanych treści i sposób ich publikowania. Można jednak przyjąć, że zarówno blogi oparte na obrazie, gdzie systematycznie pojawiają się pokazy slajdów, fotokasty i galerie zdjęć, jak i blogi oparte na materiałach wideo¹³⁸ można systematyzować i dzielić według powyższych kryteriów, ustalonych z myślą o blogach tradycyjnych¹³⁹.

2.9.2. *Data blog* – definicja pojęcia

Jak wobec powyższych kryteriów sytuje się *data blog*? Nawiązując do przedstawionego wykresu, do blogów zawierających zwizualizowane dane można zastosować podział uwzględniający osobę autora – mogą one być prowadzone zarówno przez jedną, jak i przez wiele osób. Charakter zamieszczanych na takim blogu wypowiedzi trudno jednak uznać za czysto subiektywny – ideą dziennikarstwa danych jest wszakże obiektywizacja przekazu. Przy wpisach, w których dane i statystyki są źródłem informacji, a zadaniem autora ich agregacja, obróbka i wizualizacja, miejsca na własne komentarze pozostaje

¹³⁸ Najczęściej jest to autorski film, w którym autor blogu komunikuje się za pomocą kamery internetowej ze swoimi odbiorcami – taki blog ma podobny charakter do blogów tradycyjnych, różni się sposobem prezentacji treści, zachowany jest jednak osobisty i subiektywny ton, a także wyraźny podział tematyczny.

¹³⁹ Choć oczywiście dla tego typu blogów istnieją inne podziały – wewnętrzne, np. ze względu na typ publikowanych zdjęć, ich tematykę, rodzaje fotokastów, sposób realizacji, ujęcia. To samo dotyczy wideoblogów, które publikowane są także w serwisach społecznościowych, np. YouTube, gdzie autorzy zakładają własne „kanały”, które można subskrybować.

bardzo mało. Prowadzeniu tego typu blogów przyświeca też inny cel – nie ma tutaj miejsca na prywatne przemyślenia nadawcy, pisanie o sobie i swoich problemach czy traktowania odbiorcy jako powiernika informacji i komentarzy na temat aktualności i rzeczywistości otaczającej autora.

Wartość informacyjna tego typu blogów jest niepodważalna – wśród wielowarstwowych infografik, uzupełnień tekstowych i przy możliwości swobodnego poruszania się po materiale (w przypadku infografik interaktywnych) odbiorca odnajduje wiele informacji, które może dostosowywać i dopasowywać do swoich potrzeb. Ponadto bez wątpienia zapełniają internetową niszę, wobec czego ich autorzy natychmiast mogą odnaleźć grupę odbiorców zainteresowanych dziennikarstwem danych i wizualizacją informacji. Choć ogólnie da się wskazać zależność, że im więcej powstaje blogów, tym trudniej wskazać ich rodzaj, gdyż różnice się coraz bardziej zacierają, w przypadku *data blogs* wyrazistość gatunkowa jest mocno zaakcentowana. W przypisaniu bloga do kategorii *data* pomaga jasno nakreślony rys tematyczny i profil, a także informacyjny charakter wypowiedzi.

Na innych odmianach blogów występowanie multimediiów lub grafiki we wpisach jest opcjonalne. Dla *data blogs* to wyznacznik gatunkowy. Za najważniejszy wyróżnik, czasem jedyny prócz tytułu, należy uznać obecność wizualizacji danych w formie wykresu, grafu czy infografiki. Wizualizacje te mogą być statyczne lub interaktywne, uzupełniane tekstem lub nie.

Specyfika blogów opartych o dane sprawia, że tworzenie treści jest dużo bardziej czasochłonne od prowadzenia blogów tradycyjnych, w przypadku których teksty często powstają pod wpływem chwili, są zapisem własnych przemyśleń, komentarzem do aktualnych wydarzeń, podczas gdy przy tych pierwszych „prócz zabawy w dziennikarza, musimy się wcielić również w analityka i początkującego grafika”¹⁴⁰. Każdy materiał pojawiający się na takim blogu powstaje wieloetapowo. Najpierw trzeba odszukać odpowiednie dane, przefiltrować je, dokładnie sprawdzić. Następnie, opierając się na nich, zaplanować wizualizację, która jest realizowana za pomocą specjalistycznych programów – ten proces również przebiega w kilku etapach (planowanie, tworzenie, wypełnianie treścią, publikacja).

W przedstawionym na ostatnim wykresie podziale (rys. 19) *data blog* stanowiłby oddzielną kategorię. Nie można go bowiem przyporządkować tematycznie, z uwagi na to, że za pomocą danych przedstawia się różne informacje – z pogranicza polityki, ekonomii, sportu czy mediów. Jest to swego rodzaju podzbiór z wykształconymi cechami gatunkowymi, formalnymi oraz ściśle określonym celem – informowania za pomocą wizualizacji danych.

¹⁴⁰ P. Kozłowski, *Kurs databloga...*

Podsumowując i zbierając powyższe informacje w krótką definicję, można stwierdzić, że *data blog* to ściśle określona odmiana gatunkowa bloga, o określonym procesie powstawania treści, w którym zebrane przez autora (profesjonalistę, dziennikarza) dane są wizualizowane przy użyciu specjalistycznych narzędzi. Blogi te cechuje wysoka wartość informacyjna i poznawcza oraz brak subiektywnego charakteru i elementów zdradzających osobiste przekonania twórcy. Stopień aktualności publikowanych materiałów jest uzależniony od zebranych danych.

2.10. Dane jako źródło rozrywki

2.10.1. Infotainment jako punkt wyjścia

Jeżeli przyjąć, że dane same w sobie stanowią informację oraz są jej źródłem dla złożonych gatunków medialnych, to pojawienie się kontekstu łączącego dane z rozrywką było tylko kwestią czasu. Od lat w mediach obecny jest bowiem termin „infotainment”, w analogii do którego przywołano pojęcie „datatainment”¹⁴¹. Wyjściowe „infotainment”, tłumaczone w Polsce jako „inforozrywka”¹⁴² poprzez kontaminację i bezpośredni przekład dwóch słów: *information* (informacja) i *entertainment* (rozrywka), definiuje się jako „komunikaty medialne łączące format informacyjny i rozrywkowy”¹⁴³. Ich istotą jest „nie tylko występowanie obok siebie przekazów informacyjnych i dostarczających rozrywki, ale także ich wzajemne przenikanie się”¹⁴⁴. Gatunek ten przypisywano początkowo jedynie telewizji, jednak odnosi się również do prasy, radia i Internetu.

Ostatnio tego typu krótkie informacje, najczęściej w mediach komercyjnych, nie tylko informują o otaczającej rzeczywistości, ale bawią i szokują. Krótka informacja ma także zaskoczyć odbiorców. Jest to świadome przekształcanie informacji w infotainment, który jest doskonałym towarem, nie liczy się bowiem co jest przekazane, ale w jaki sposób¹⁴⁵.

¹⁴¹ Zob. P. Szews, *Liczby, dane i statystyki...*, s. 257–260.

¹⁴² Warto wspomnieć, że pojęcie to jest przekładane dwojako: obok siebie występują „inforozrywka” i „infozrywka”.

¹⁴³ A. Grabarczuk, *Infotainment w amerykańskiej telewizji*, [w:] K. Burska, P. Czarnek (red.), *Rozrywka w mediach*, Łódź 2012, s. 146.

¹⁴⁴ Tamże.

¹⁴⁵ K. Wolny-Zmorzyński, A. Kaliszewski, W. Furman, *Gatunki dziennikarskie. Teoria, praktyka...*, s. 38–39.

O nowej formie gatunkowej Wolny-Zmorzyński pisał jako kontrowersyjnej, traktującej informację niczym towar, podawany w atrakcyjnym, często krzykliwym opakowaniu masowemu odbiorcy, nastawionemu na rozrywkę. Z jednej strony budzi to obawy ryzyka manipulacji, deformacji i selekcji faktów, a także ośmieszania i przekraczania granic prywatności przedstawianych postaci, z drugiej jednak taki sposób informowania spełnia zasadę „bawiąc, uczyć”. Ponadto wpisuje się w stwierdzenie McLuhana, że już sam środek przekazu jest przekazem.

W szerszym znaczeniu infotainment to także sposób podawania informacji w mediach; informowanie o faktach i wydarzeniach w formie jak najbardziej atrakcyjnej, przypominającej przekaz rozrywkowy¹⁴⁶.

2.10.2. Datatainment

„Datatainment” jest terminem lokującym się bardzo blisko inforozrywki, który można definiować w podobny sposób, zakładając, że w jego obrębie przenikają się wzajemnie dane liczbowe i statystyki z treściami o charakterze rozrywkowym. W odróżnieniu od infotainmentu, obecnego w Internecie, ale i w mediach tradycyjnych (szczególnie w szerszym znaczeniu tego pojęcia), datatainment jako odmiana gatunkowa występuje głównie w Internecie, choć, w okrojonej formie, pojawia się także w telewizji.

„Danorozrywka” (jeśli zastosować dosłowny przekład jak w przypadku inforozrywki) upodabnia się do rozrywki zarówno w węższym, jak i szerszym kontekście. Z jednej strony jest to łączenie informacji w postaci danych, statystyk z rozrywką, a z drugiej jest to także sposób prezentowania danych, angażowanie odbiorcy za ich pomocą, pobudzanie do interakcji. Można w tym miejscu przywołać przykład Hansa Roslinga – „w jego rękach dane ożywają, zaczynają mówić i zabierają nas w historyczną podróż, drogą wypełnioną ekscytacją i chęcią powiedzenia nam, jak jest naprawdę”¹⁴⁷. Rosling zajmuje się problemami zdrowia na świecie. Podczas swojej prezentacji na jednej z konferencji zamiast prezentowania dziesiątek surowych liczb i statystyk, posłużył się interaktywną animacją, wyświetlaną za pomocą projektora, dzięki której przedstawił angażującą historię dotyczącą sytuacji w Stanach Zjednoczonych i Wietnamie po wojnie, Chin po wprowadzeniu reformy zdrowia oraz stosunków między Jemenem a Emirataми Arabskimi¹⁴⁸. Był to jeden z pierwszych przykładów kreatywnego

¹⁴⁶ Zob. tamże.

¹⁴⁷ Digiredo.nl, *Data-tainment*, <http://www.digiredo.nl/data-tainment/> (dostęp: 16.09.2019), tłumaczenie własne.

¹⁴⁸ Zob. tamże.

połączenia liczb i przekazu rozrywkowego (w tym przypadku audiowizualnego) dla zaprezentowania określonych historii.

Dziedziną, która generuje bardzo dużo danych, jest sport (który również może być traktowany jako forma rozrywki). Za przykład można podać tutaj piłkę nożną – w trakcie jednego meczu dyscyplina ta pozwala na zebranie takich statystyk, jak czas utrzymywania się przy piłce, liczba wymienionych podań (celnych i niecelnych), skuteczność poszczególnych zawodników, statystyka fauli czy nawet liczba przebiegniętych kilometrów (przez zawodnika lub całą drużynę). Podobnie jak w przypadku innych danych, także i te są łatwo dostępne dla wszystkich zainteresowanych, dzięki istnieniu dedykowanych aplikacji – *live streaming stats*. Same w sobie mają wartość tylko informacyjną, ale wykorzystuje się je np. w tzw. *fantasy leagues*, czyli internetowych grach, w których kibice wybierają swój skład rzeczywistych piłkarzy. Każdemu z takich zawodników wirtualnej drużyny przyznaje się punkty, których liczba zależy od statystyk dotyczących wybranego piłkarza. „Większość zbieranych informacji jest jednak nastawiona na >>zaawansowanych<< fanów”¹⁴⁹.

Pojęcie *datatainmentu* ma swoje korzenie właśnie w sporcie. Po raz pierwszy w takim kontekście użył go Richard Ayers, podczas konferencji South By Southwest. Pracownik pionu marketingu klubu Manchester City stwierdził, że *data journalism* i wizualizacja danych są świetnym sposobem prezentowania informacji, ale nadal niewystarczającym¹⁵⁰. Według Ayersa dane powinny angażować i, jako dodatkowa wartość, służyć rozrywce, co jest zresztą zgodne z definicją wg *Collins Dictionary*: „*Datatainment* – informacja w postaci danych jako forma rozrywki”¹⁵¹. Etymologia słowa oddaje istotę *datatainmentu*. To po prostu proces, w którym dane przedstawiane są w taki sposób, by stanowiły dla odbiorcy źródło rozrywki. „Surowe” na początku liczby poddaje się obróbce – graficznej, wizualnej – i w nowej formie prezentuje się odbiorcy. Statystyki to bardzo ważny element dla miłośników sportu, stanowią często argument w dyskusji – zresztą coraz częściej i w większej ilości udostępniane są w mediach (omawiane w studiu w przerwach spotkań sportowych, przytaczane przez komentatorów). W odpowiedzi na rosnącą liczbę infografik (wpisanych na stałe w prasowe rubryki sportowe, m.in. przy podsumowaniach i relacjach

¹⁴⁹ J. Beckham, *Soccer Sexes Up Stats to Create ‘Datatainment’*, <http://www.wired.com/playbook/2012/03/soccer-datatainment/> (dostęp: 17.09.2019).

¹⁵⁰ Tamże.

¹⁵¹ Collins Dictionary, *Datatainment*, <http://www.collinsdictionary.com/submission/6055/Datatainment> (dostęp: 17.09.2019), tłumaczenie własne (oryginał brzmi: „*Datatainment* – information data as a form of entertainment”).

z wydarzeń sportowych) wizualizacja danych jest rozwijana tak, by nie tylko informować, ale także bawić i pobudzać interakcję¹⁵².

W mediach, szczególnie w Internecie, nie brakuje narzędzi i stron zbierających dane i przedstawiających je w postaci interaktywnej. Na Twitterze i serwisie internetowym swoje statystyki w atrakcyjnej i przystępnej formie prezentuje firma Opta, która współpracuje z największymi europejskimi klubami piłkarskimi i redakcjami sportowymi. W piłkarskiej „netosferze” nie brak również blogów zajmujących się szczegółową analizą danych i statystyk, jak np. Zonalmarking¹⁵³ czy jego polski odpowiednik – Taktycznie.net¹⁵⁴.

Datatainment jest terminem blisko związanym z zyskującym coraz większą popularność zjawiskiem *second screen watching*, które narodziło się w wyniku konwergencji telewizji i mediów społecznościowych i zmienia postrzeganie dzisiejszej telewizji. Warto w tym kontekście powiedzieć, że „z badań Interactive Research Center wynika, że 73% polskich internautów w wieku 19–30 lat podczas oglądania telewizji korzysta z laptopa, smartfona lub tabletu. Z co najmniej dwóch urządzeń naraz korzysta co czwarty polski internauta w wieku 19–30 lat”¹⁵⁵. To właśnie dla takiego odbiorcy dane mogą być dodatkowym źródłem rozrywki. Korzystają z tego m.in. twórcy aplikacji (oficjalnych i nieoficjalnych) przygotowanych dla fanów Formuły 1¹⁵⁶. W trakcie wyścigu w czasie rzeczywistym dostarczane są na ekran komputera lub smartfona szczegółowe dane dotyczące czasów pojedynczych sektorów, zmian kół czy najlepsze okrążenia i różnice czasowe między poszczególnymi kierowcami. Nie jest to przedstawienie kilkunastu tysięcy liczb, ale przejrzysta ich wizualizacja – wykresy, interaktywna miniatura toru z nazwiskami kierowców etc.

Na podstawie danych generowanych przez sport tworzy się w Internecie specjalne aplikacje i gry służące rozrywce osobom, które oczekują od sportu czegoś więcej niż tylko widowiska. Udostępnianie i odpowiednia prezentacja danych liczbowych pozwalają angażować odbiorcę jeszcze długo po zakończeniu meczu, igrzysk czy innych popularnych wydarzeń sportowych, stanowiąc

¹⁵² S. Walsh, *Manchester City Datatainment – player data goes public*, <http://digital-football.com/featured/manchester-city-datatainment-player-data-stats-go-on-the-website/> (dostęp: 17.09.2019).

¹⁵³ <http://www.zonalmarking.net/> (dostęp: 17.09.2019).

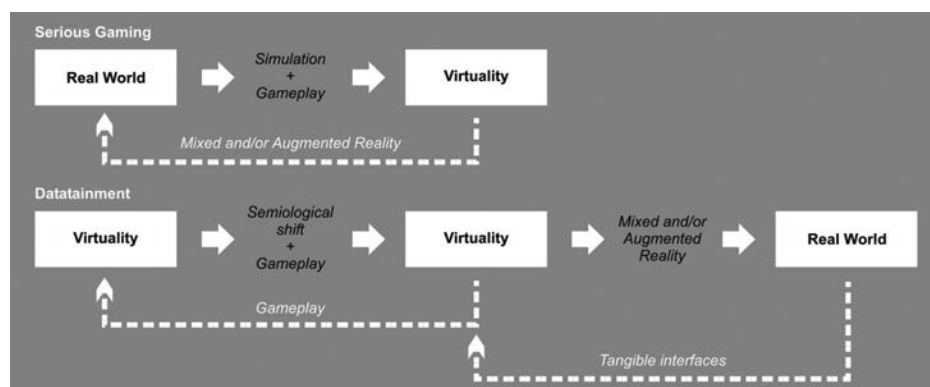
¹⁵⁴ <http://taktycznie.net/> (dostęp: 17.09.2019).

¹⁵⁵ A. Robotycka, *Social TV daje więcej niż TV i social razem wzięte*, <http://nowymarketing.pl/a/1991,social-tv-daje-wiecej-niz-tv-i-social-razem-wziete> (dostęp: 18.09.2019).

¹⁵⁶ T. Hirst, *Sports Data Journalism and „Datatainment”*, <http://blog.ouseful.info/2011/11/04/sports-data-journalism-and-datatainment/> (dostęp: 17.09.2019).

dla organizatorów i osób odpowiedzialnych za marketing i PR wyzwanie, ale i szansę na dodatkowy dochód i zwiększanie zasięgu marki.

Inne rozumienie datatainmentu zaproponował Gregoire Cliquet¹⁵⁷. Definiuje ten termin jako nowy sposób przedstawiania danych cyfrowych związanych z działalnością osób w świecie informacji. Nowa koncepcja miała za zadanie wdrożyć dynamiczne transpozycje działań związanych z informacją dla odbiorcy zorientowanego na rozrywkę. W tym przypadku dane rozumiane są jako plastyczny materiał bazowy, któremu można nadawać dowolne formy i kształtować w różnych kierunkach. Autor zwraca uwagę, że mogą zostać przedstawione w sposób elementarny – w formacie cyfrowym binarnym lub szesnastkowym. Datatainment, według Cliqueta, wywodzi się z eksperymentów prowadzonych w dziedzinie odwzorowania, przede wszystkim tego, co stanowi obszar cyfrowy. W związku z tym konieczne stają się zapożyczenia z obszaru graficznej semiologii – zmienne (dane cyfrowe) są zmaterializowane i przedstawione w sposób metaforyczny, za pomocą graficznych symboli (kontekstualizacja danych lub wiadomości). Datatainment stanowi więc bardziej sposób reprezentacji, który można wykorzystać do przekazywania informacji, a także w edukacji i nauce. Polega na zasobach niematerialnych (pochodzących z sieci web) i dąży do ponownego skontekstualizowania ich w tym samym środowisku, ale w inny sposób. Ideą tego procesu jest zatem zmaterializowanie wirtualnych elementów w rzeczywistych aplikacjach za pośrednictwem specjalnych interfejsów¹⁵⁸.



Rysunek 20. Definicja terminu „Datatainment” autorstwa Gregoire’a Cliqueta

Źródło: G. Cliquet, *Datatainment*, <http://turing.lecolededesign.com/gcliquet/2007/10/06/datatainment/> (dostęp: 26.09.2019).

¹⁵⁷ G. Cliquet, *Datatainment*, <http://turing.lecolededesign.com/gcliquet/2007/10/06/datatainment/> (dostęp: 26.09.2019).

¹⁵⁸ Tamże.

ZAKOŃCZENIE

Umiejętność opowiadania historii za pomocą danych wymaga wieloetapowych przygotowań, a często także żmudnej selekcji danych, poszukiwania i agregacji. Oddzielić w tym miejscu należy dziennikarstwo ekonomiczne czy finansowe od dziennikarstwa danych, które, jak zostało to zdefiniowane wcześniej, ma inne cele i zadania. To stosunkowo nowa w Polsce specjalizacja, z nieukonstytuowanym schematem redakcji, zorientowana na Internet, który oferuje najwięcej źródeł informacji i narzędzi odpowiadających specyfice tej specjalizacji. Mimo dużego potencjału tkwiącego w informacji cyfrowej, konieczność dostrzeżenia ciekawych historii i procesów ukrytych między danymi sprawia, że przygotowanie materiałów dziennikarstwa danych jest bardziej czasochłonne od tworzenia tradycyjnych form. Banalizacja i szybkość, skrótowność przekazu, nastawienie na jak największą ilość odsłon, walka o kliknięcia tytułami typu *click bait* mogą przeważać nad chęcią publikacji materiałów „jakościowych”, niszowych, skierowanych do wąskiej grupy odbiorców.

Szansą rozwoju tej specjalizacji jest upowszechnienie się narzędzi do pracy z danymi, jak również powstawanie nowych programów i aplikacji pozwalających na łatwą selekcję i wizualizowanie informacji. Skrócenie czasu potrzebnego do opracowania artykułów *data journalism* poszerzyłoby możliwości tej specjalizacji i wydaje się bardzo korzystne z perspektywy odbiorcy. Nie bez znaczenia dla nadawcy pozostaje zwiększająca się liczba źródeł informacji w formie cyfrowej, które ewoluują, umożliwiając łatwą selekcję, agregację i dostęp.

Obserwując popularność treści i portali infotainmentowych, można także pokusić się o stwierdzenie, że dziennikarstwo danych, ale w formie przedstawiającej dane w ramach datatainmentu, ma jeszcze większe szanse na rozwój i zaistnienie w świadomości odbiorcy. Bliskoznaczna w wielu kontekstach i głęboko już zakorzeniona w mediach inforozrywka przetrzała szlaki dla takiego sposobu przekazywania informacji. Obserwując kreatywne łączenie informacji liczbowych i rozrywki w sporcie, można śmiało stwierdzić, że datatainment z powodzeniem sprawdzi się również w innych dziedzinach, jak np. muzyka czy film, mając na względzie coraz popularniejsze zjawisko *second screen watching*.

Zdefiniowanie nowej specjalizacji dziennikarskiej, jaką jest *data journalism*, otwiera horyzont poznawczy do dalszych badań i analiz tej odmiany. Choć

ta forma jest obecna też w prasie, zdecydowaną przyszłość i perspektywę rozwoju dostrzegać należy w Internecie. To właśnie w sieci dziennikarstwo danych ujawnia pełnię możliwości: od szerszych perspektyw wizualizacji po interaktywność i urozmaicone sposoby tworzenia narracji.

Ukazanie dziennikarstwa danych i infografiki na tle zjawiska wizualizacji danych i samej komunikacji wizualnej daje pole do dalszych badań, jako że są wciąż nowe formy, dynamicznie rozwijające się w mediach, ewoluujące, generujące nowe odmiany i realizacje. Prezentując informacje i dane w sposób ciekawy dla odbiorcy, umożliwiając wprowadzenie interaktywnej narracji i efektownych wizualizacji, są przyszłościowe w mediach i redakcjach, gdzie jeszcze częściej będą zatrudniani specjaliści w tych dziedzinach. Pomogą tym samym w uzyskaniu przewagi w walce o czytelnika i liczbę odsłon lub zakupionych subskrypcji, jak chociażby w przypadku portalu BIQdata.

Szansą dla dziennikarstwa danych, jako odmiany jakościowej, jest także wzrost popularności płatnych subskrypcji, z których Polacy korzystają coraz chętniej. Dodatkowe finansowanie czy zamknięcie materiałów *data journalism* za tzw. paywallem mogą zapewnić środki twórcom i dać gwarancję ciekawych treści odbiorcom. Możliwość wglądu do treści w modelu subskrypcyjnym rokrocznie zyskuje na popularności, obecna zarówno na platformach wideo, jak i w prasie cyfrowej czy na portalach z podcastami. Choć często panuje przekonanie, że w sieci wszystko dostępne jest za darmo, należy sądzić, że liczba odbiorców skorych do wykupienia płatnych subskrypcji będzie wciąż rosła. Czas pokaże, czy i w jakim stopniu przyczyni się to do rozwoju dziennikarstwa danych i profesjonalnego wizualizowania informacji.

BIBLIOGRAFIA

- Adamski A., *Media w analogowym i cyfrowym świecie*, Warszawa 2012.
- Allan S., *Newsy w sieci. Internet i dziennikarstwo*, tłum. A. Sadza, Kraków 2008.
- Barańska A., Źródła informacji dziennikarza internetowego, [w:] G. Habrajska, J. Mikosz (red.), *Dziennikarstwo internetowe. Teoria i praktyka*, Łódź 2010.
- Bauer Z., *Dziennikarstwo wobec nowych mediów*, Warszawa 2009.
- BBC.com, *Robot Writes LA Times Earthquake Breaking News Article*, <http://www.bbc.com/news/technology-26614051> (dostęp: 06.09.2019).
- Beckham J., *Soccer Sexes Up Stats to Create 'Datatainment'*, <http://www.wired.com/playbook/2012/03/soccer-datatainment/> (dostęp: 17.09.2019).
- Biecek P., *Odkrywać! Ujawniać! Objaśniać! Zbiór esejów o sztuce prezentowania danych*, Warszawa 2014.
- Biniewicz J., *Blog polityczny, czyli o kreowaniu wizerunku politycznego słów kilka*, [w:] M. Jeziński, A. Seklecka, Ł. Wojtkowski (red.), *Nowe media we współczesnym społeczeństwie*, Toruń 2011.
- Bitner M., Chłopecki A., Gomulka S., Otto W., Stępień J., Wiśniewski M., *Obywatelski kontrraport – krytyczna analiza rządowego przeglądu systemu emerytalnego*, Warszawa 2013, <http://kobe.org.pl/wp-content/uploads/2013/07/Obywatelski-Kontraport.pdf> (dostęp: 15.07.2019).
- Bounegru L., *Computer-Assisted Reporting and Precision Journalism*, [w:] J. Gray, L. Bounegru, L. Chambers (red.), *The Data Journalism Handbook. How Journalists Can Use Data to Improve the News*, Sebastopol (CA) 2012.
- Bradshaw P., Rohumaa L., *The Online Journalism Handbook: Skills to Survive and Thrive in the Digital Age*, New York 2011.
- Cliquet G., *Datatainment*, <http://turing.lecolededesign.com/gcliquet/2007/10/06/datatainment/> (dostęp: 26.09.2019).
- Cognity.pl, *Manipulacja danymi na wykresach*, <https://www.cognity.pl/manipulacja-danymi-na-wykresach,blog,94.html> (dostęp: 11.07.2019).
- Collins Dictionary, *Datatainment*, <http://www.collinsdictionary.com/submission/6055/Datatainment> (dostęp: 17.09.2019).
- Czapiewski B., *Co to jest chartjunk (wykres śmieciowy) i lie-factor (współczynnik kłamstwa)?*, <https://skuteczneraporty.pl/blog/co-to-jest-chartjunk-wykres-smieciowy-i-lie-factor-wspolczynnik-klamstwa/> (dostęp: 17.08.2019).

- Czapiewski B., *Dziewięć zasad skutecznego użycia koloru*, <https://skuteczneraporty.pl/blog/dziewiec-zasad-skutecznego-uzycia-koloru/> (dostęp: 17.08.2019).
- Czapiewski B., *Wykresy, których należy unikać*, <http://skuteczneraporty.pl/blog/wykresy-ktorych-nalezy-unikac/> (dostęp: 13.08.2019).
- Dewdney A., Ride P., *The New Media Handbook*, London–New York 2010.
- Diep F., *Associated Press Will Use Robots to Write Articles*, <http://www.popsci.com/article/technology/associated-press-will-use-robots-write-articles?dom=PSC&loc=recent&lnk=4&con=associated-press-will-use-robots-to-write-articles> (dostęp: 04.09.2019).
- Digiredo.nl, *Data-tainment*, <http://www.digiredo.nl/data-tainment/> (dostęp: 16.09.2019).
- Dijk van J., *The Network Society: Social Aspects of New Media*, London 2006.
- Doctorow C., Dornfest R., Johnson J.S., Powers S., Trott B., Trott M.G., *Blogging – przewodnik*, tłum. P. Kresak, Warszawa 2003.
- Fisher B., Świerczyńska-Głównia W., *Dostęp do informacji ustawowo chronionych, zarządzanie informacją. Zagadnienia podstawowe dla dziennikarzy*, Kraków 2006.
- Fundacja Naukowa SmarterPoland.pl, *Jeden zły rysunek jest wart 1000 słów... korekty*, <http://smarterpoland.pl/index.php/2012/03/jeden-zly-rysunek-jest-wart-1000-slow-korekty/> (dostęp: 27.06.2019).
- Fundacja Naukowa SmarterPoland.pl, *Wykres ministerstwa finansów z „lie-factor” przekraczającym 200%*, <http://smarterpoland.pl/index.php/2013/04/wykres-ministerstwa-finansow-z-lie-factor-przekraczajacym-200/> (dostęp: 27.06.2019).
- Fundacja Naukowa SmarterPoland.pl, *Zmiany w OFE a wykresy Ministerstwa Pracy i Polityki Społecznej*, <http://smarterpoland.pl/index.php/2013/11/zmiany-w-ofe-lie-factor-a-ministerstwo-pracy-i-polityki-spoecznej/> (dostęp: 15.07.2019).
- Gluzza R., *Newsy z Excela*, „Press” 2013, nr 09.
- Główny Urząd Statystyczny, *Monitoring pracy. Wejście ludzi młodych na rynek pracy*, http://stat.gov.pl/cps/rde/xbcr/gus/PUBL_Wejscie_ludzi_mlodych_na_rynek_pracy.pdf (dostęp: 11.07.2019).
- Goban-Klas T., *Pochwała medioznawstwa czyli renesans McLuhana*, „Studia Medioznawcze” 2000, nr 1.
- Grabarczuk A., *Infotainment w amerykańskiej telewizji*, [w:] K. Burska, P. Czarnek (red.), *Rozrywka w mediach*, Łódź 2012.
- Hirst T., *Sports Data Journalism and „Datatainment”*, <http://blog.ouseful.info/2011/11/04/sports-data-journalism-and-datatainment/> (dostęp: 17.09.2019).
- Hopfinger M., *Rekonfiguracja komunikacji społecznej*, [w:] K. Wolny-Zmorzyński, W. Furman (red.), *Internetowe gatunki dziennikarskie*, Warszawa 2010.
- Jakubowicz K., *Nowa ekologia mediów. Konwergencja a metamorfoza*, Warszawa 2011.
- Jastrzębski J., *Chaos, baza danych i internetowe gatunki dziennikarskie*, [w:] K. Wolny-Zmorzyński, W. Furman (red.), *Internetowe gatunki dziennikarskie*, Warszawa 2010.

- Jeziński M., Seklecka A., Wojtkowski Ł. (red.), *Nowe media we współczesnym społeczeństwie*, Toruń 2011.
- Jędrysik M., *Dziennikarstwo danych – szansa i wyzwanie*, <https://centrumcyfrowe.pl/czytelnia/milada-jedrysik-o-data-journalism-dla-szkoly-infografiki/> (dostęp: 30.08.2019).
- Kadaj-Kuca K., *Moda na blogowanie. Blog jako komentarz, element polityki redakcyjnej i narzędzie kreacji politycznej rzeczywistości*, [w:] K. Wolny-Zmorzyński, W. Furman (red.), *Internetowe gatunki dziennikarskie*, Warszawa 2010.
- Kahneman D., *Pułapki myślenia. O myśleniu szybkim i wolnym*, tłum. P. Szymczak, Poznań 2012.
- Kancelaria Prezesa Rady Ministrów, *Wsparcie dla dzieci z uboższych rodzin w nowym roku szkolnym*, <https://www.premier.gov.pl/mobile/en/news/news/en-wsparcie-dla-dzieci-z-ubozszych-rodzin-w-nowym-roku-szkolnym.html> (dostęp: 11.07.2019).
- Kawka M., *Blog jako gatunek dziennikarski – ewolucja i transgresja*, [w:] K. Wolny-Zmorzyński, W. Furman (red.), *Internetowe gatunki dziennikarskie*, Warszawa 2010.
- Kononiuk T., *Profesjonalizacja w dziennikarstwie. Między modernizmem a ponowoczesnością*, Warszawa 2013.
- Kozłowski P., *Kurs databloga: data blogging, czyli blogowanie z danymi – jak założyć data blog?*, <http://datablog.pl/data-blogging-blogowanie-danymi-zalozyc-data-blog/> (dostęp: 29.09.2019).
- Kozłowski P., *Kurs dziennikarstwa danych #1: Czym są dane?*, <http://datablog.pl/kurs-dziennikarstwa-danych-1-czym-sa-dane/> (dostęp: 29.08.2019).
- Levinson P., *Nowe nowe media*, tłum. M. Zawadzka, Kraków 2010.
- Linksvayer M., *Getting Data*, [w:] J. Gray, L. Bounegru, L. Chambers (red.), *The Data Journalism Handbook. How Journalists Can Use Data to Improve the News*, Sebastopol (CA) 2012.
- Logan R.K., *Understanding New Media*, New York 2010.
- Lorenz M., *Why Journalist Should Use Data*, [w:] J. Gray, L. Bounegru, L. Chambers (red.), *The Data Journalism Handbook. How Journalists Can Use Data to Improve the News*, Sebastopol (CA) 2012.
- Manovich L., *Język nowych mediów*, tłum. P. Cypriański, Warszawa 2006.
- Mayer-Schonberger V., Cukier K., *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Boston 2013.
- Meyer P., *Paper Route: Finding My Way to Precision Journalism*, <http://www.unc.edu/~pmeyer/book/Chapter1.htm> (dostęp: 28.08.2019).
- Ministerstwo Pracy i Polityki Społecznej, *Planowane zmiany w systemie emerytalnym*, <http://www.mpips.gov.pl/aktualnosci-wszystkie/ubezpieczenia-spoeczne/art,6499,planowane-zmiany-w-systemie-emerytalnym.html?serwis=2> (dostęp: 28.06.2019).

- Ministerstwo Pracy i Polityki Społecznej, *Przegląd funkcjonowania systemu emerytalnego*, <http://www.mpips.gov.pl/aktualnosci-wszystkie/ubezpieczenia-spoleczne/art,6266,przeglad-funkcjonowania-systemu-emerytalnego.html> (dostęp: 28.06.2019).
- Moll J., *Sztuczna inteligencja może w przyszłości zastąpić dziennikarzy*, <http://tylkonauka.pl/wiadomosc/sztuczna-inteligencja-moze-w-przyszlosci-zastapic-dziennikarzy> (dostęp: 06.09.2019).
- Money.pl, *Premier ujawnia, która wersja zmian w OFE wygra*, <http://www.money.pl/gospodarka/wywiady/artykul/premier;ujawnia;ktora;wersja;zmian;w;ofe;wygra,54,0,1421110.html> (dostęp: 28.06.2019).
- Murdock G., P. Golding, *Cyfrowe możliwości a realia rynku: sprzeczności konwergencji*, tłum. M. Łuczak [w:] G. Murdock, *Media, kultura, ekonomia. Krytyczne pytania. Antologia tekstów*, opr. J. Fras, Wrocław 2010.
- Olśzański L., *Dziennikarstwo internetowe*, Warszawa 2006.
- Olśzański L., *Media i dziennikarstwo internetowe*, Warszawa 2012.
- Pacula P., Nils Mulvad: *Dziennikarstwo danych to współczesny punk*, <https://pl.ejo-online.eu/dziennikarstwo-polityczne/nils-mulvad-dziennikarstwo-danych-wspolczesny-punk> (dostęp: 30.08.2019).
- Piekarski K., *Kultura danych. Algorytmy wzmacniające uwagę*, Gdańsk 2017.
- Pisarek W. (red.), *Słownik terminologii medialnej*, Kraków 2006.
- Playfair W., *The Commercial and Political Atlas*, London 1786.
- The Recorded Future Team, *What Is Open Source Intelligence and How Is it Used?*, <https://www.recordedfuture.com/open-source-intelligence-definition/> (dostęp: 19.03.2019).
- Robotycka A., *Social TV daje więcej niż TV i social razem wzięte*, <http://nowymarketing.pl/a/1991,social-tv-daje-wiecej-niz-tv-i-social-razem-wziete> (dostęp: 18.09.2019).
- Rostowski J., *Piramida finansowa PiS*, <http://rostowski.natemat.pl/30143,piramida-finansowa-pis> (dostęp: 13.08.2019).
- Saramak B., *Wykorzystanie otwartych źródeł informacji w działalności wywiadowczej. Historia, praktyka, perspektywy*, Warszawa 2015.
- SAS Institute, *Cztery powody, dla których nie możesz się obyć bez wizualizacji danych*, https://www.sas.com/pl_pl/offers/sas-wizualizacja-danych-ebook/index.html (dostęp: 30.04.2019).
- Schmidt S.J., *Blickwechsel. Umrisse einer Medienepistemologie*, [w:] G. Rusch, S.J. Schmidt (red.), *Konstruktivismus in der Medien- und Kommunikationswissenschaft*, Frankfurt 1999.
- Schneider K., *Visual Display of Information*, <https://technogenii.net/visual-display-of-information/> (dostęp: 11.07.2019).
- Siemes A., *Liczby w mediach: rola i funkcja rzeczywistości liczb w komunikacji medialnej – wyniki badania empirycznego*, [w:] B. Bogólewska, M. Worsowicz (red.), *Styl, dyskurs, media*, Łódź 2010.

- Siwek K., *Drożęca energia podnosi ceny mieszkań*, <http://abcnieruchomosci.pl/drozejaca-energia-podnosi-ceny-mieszkan> (dostęp: 27.06.2019).
- Szews P., *Algorytmizacja przestrzeni internetowej*, „Acta Universitatis Lodziensis. Folia Litteraria Polonica” 2015, nr 2(28).
- Szews P., *Liczbę, dane i statystyki w dziennikarstwie internetowym*, „Acta Universitatis Lodziensis. Folia Litteraria Polonica” 2014, nr 1(23).
- Szews P., *Mikroblog – odmiana blogu czy oddzielny gatunek*, „Acta Universitatis Lodziensis. Folia Litteraria Polonica” 2013, nr 2(20).
- Świerkot M., Kolany K., *Gospodarcze osiągnięcia partii politycznych*, <http://www.bankier.pl/wiadomosc/Gospodarcze-osiagniecia-partii-politycznych-2417944.html> (dostęp: 11.07.2019).
- Thibodeaux T., *5 tips for Getting Started in Data Journalism*, <http://www.poynter.org/how-tos/digital-strategies/147734/5-tips-for-getting-started-in-data-journalism/> (dostęp: 11.09.2019).
- Tufte E.R., *The Visual Display of Quantitative Information*, Cheshire (CT) 1991.
- UNdata.un.org, *About UNdata*, <https://data.un.org/Host.aspx?Content=About> (dostęp: 29.09.2019).
- Walsh S., *Manchester City Datatainment – player data goes public*, <http://digital-football.com/featured/manchester-city-datatainment-player-data-stats-go-on-the-website/> (dostęp: 17.09.2019).
- Warden P., *What is data journalism?*, <http://petewarden.typepad.com/searchbrowser/2010/10/page/3/2010>, (dostęp: 30.08.2019).
- Wikipedia.org, *Cynthia Brewer*, https://en.wikipedia.org/wiki/Cynthia_Brewer (dostęp: 27.07.2019).
- Wikipedia.org, *Data journalism*, http://en.wikipedia.org/wiki/Data_journalism (dostęp: 30.08.2019).
- Wikipedia.org, *Database journalism*, http://en.wikipedia.org/wiki/Database_journalism (dostęp: 29.11.2013).
- Wikipedia.org, *Data-driven journalism*, https://en.wikipedia.org/wiki/Data-driven_journalism (dostęp: 30.08.2019).
- Wikipedia.org, *William Playfair*, http://en.wikipedia.org/wiki/William_Playfair (dostęp: 27.04.2019).
- Wolny-Zmorzyński K., Kaliszewski A., Furman W., *Gatunki dziennikarskie. Teoria, praktyka, język*, Warszawa 2006.
- Wolny-Zmorzyński K., Kaliszewski A., Furman W., Pokorna-Ignatowicz K., *Źródła informacji dla dziennikarza*, Warszawa 2008.
- World Life Expectancy, *Poland total deaths by cause. 2010*, <http://www.worldlifeexpectancy.com/contry-health-profile/poland> (dostęp: 29.08.2019).
- Wyborcza.pl, *Co trzeci Polak zadowolony z wyniku wyborów*, http://wyborcza.pl/1,75248,18131787,Co_trzeci_Polak_zadowolony_z_wyniku_wyborow___Coraz.html (dostęp: 13.08.2019).

Pozostałe źródła internetowe

<http://bombsight.org/> (dostęp: 29.08.2019).
<http://colorbrewer2.org/> (dostęp: 13.08.2019).
<http://datahub.io> (dostęp: 29.09.2019).
<http://data.worldbank.org> (dostęp: 29.09.2019).
<http://data.worldbank.org/polish> (dostęp: 29.09.2019).
<http://data.worldbank.org/topic/health> (dostęp: 29.09.2019).
<https://data.un.org/Feedback.aspx> (dostęp: 29.09.2019).
<http://getthedata.org> (dostęp: 28.09.2019).
<http://infografika.wp.pl> (dostęp: 11.07.2019).
<https://scraperwiki.com> (dostęp: 29.09.2019).
<http://stat.gov.pl> (dostęp: 29.09.2019).
<http://taktycznie.net/> (dostęp: 17.09.2019).
<http://twitter.com/undata> (dostęp: 29.09.2019).
<http://wykresy.blox.pl/html> (dostęp: 11.07.2019).
<http://zonalmarking.net/> (dostęp: 17.09.2019).

We współczesnych mediach oprócz szybkości przekazu liczy się sposób, w jaki treści trafiają do odbiorcy oraz używane platformy czy atrakcyjność i kondensacja informacji. Dominacja urządzeń mobilnych, nie tylko jako środków komunikacji, ale również jako środków konsumpcji materiałów dziennikarskich, powodują, że nad tekstem często przeważają treści wizualne i multimedialne. Pomaga to w efektywnym odwołaniu się do wyobraźni odbiorcy, ale też maksymalizuje skuteczność przekazu. Na procesie wizualizacji danych autor koncentruje się w pierwszej części publikacji, podkreślając, że mimo obecności niemal w każdej dziedzinie piktogramów, map, infografik, operowania obrazem w różnych mediach i na różnych poziomach, odbiorca dla prawidłowego odczytania takich komunikatów musi posiadać odpowiednie kompetencje. W drugiej części z kolei prezentuje stosunkowo nową, rozwijającą się gałąź dziennikarstwa – opartego na danych, przedstawiając możliwości, jakie niosą one ze sobą w połączeniu z umiejętnością dostrzegania w nich historii i opowiedzenia jej w formie cyfrowej informacji.

Książka dostępna również
jako e-book

 **WYDAWNICTWO
UNIwersYTETU
ŁÓDZKIEGO**

 wydawnictwo.uni.lodz.pl
 ksiegarnia@uni.lodz.pl
 (42) 665 58 63

ISBN 978-83-8220-489-6



9 788382 204896