

27. Big Data

Big Data jest terminem niejednoznacznym i przynależnym do różnych środowisk i sektorów. Na inne aspekty zwrócą uwagę informatycy, dla których istotne staną się kwestie agregacji i przetwarzania danych; z kolei dla ekonomistów ważna będzie efektywność ich wykorzystania²⁷. Podobnie zresztą jak dla polityków, dla których możliwość skorzystania z odpowiednich kwerend i zbiorów danych może mieć kluczowe znaczenie nie tylko dla bieżącego sprawowania kontroli nad społeczeństwem, ale także utrzymania się lub zdobycia władzy w przyszłości. *Big Data* oznacza zatem nieograniczoną ilość danych, które można zbierać, porządkować i nimi zarządzać, by uzyskać z góry określony efekt, np. pozyskać wiedzę o klientach określonych produktów lub też wyznaczyć przypuszczalne trendy (zob. rozdz. *Prognozowanie i analiza trendów a rozumienie globalnego świata*). Co istotne, jako kluczowy zasób w ujęciu globalnym zaczęto je traktować dopiero w XXI wieku. Wpływ na tak późne rozpoczęcie publicznych dyskusji miały przede wszystkim ograniczenia techniczne, związane z rozwojem Internetu, m.in. nierównomierny przyrost nowych użytkowników, wsparcie publicznych środków w rozwój infrastruktury oraz, co wydaje się najistotniejsze, pojawienie się mediów społecznościowych, co spowodowało na przełomie pierwszej i drugiej dekady drastyczny przyrost danych publicznych wytwarzanych przez osoby prywatne.

²⁷ M. Tabakow, J. Korczak, B. Franczyk, *Big Data – definicje, wyzwania i technologie informatyczne*, „Informatyka Ekonomiczna” 2014, nr 1 (31), s. 138–139.

Termin *Big Data* przybliżył szerokiemu gronu odbiorców John Mashey w 1998 roku, który utożsamiał go z falą, która pojawiła się dzięki intensyfikacji działań w przestrzeni informatycznej przez sektor prywatny²⁸. Rok później, czyli 31 marca 1999 roku, na ekrany amerykańskich kin wchodzi film braci Wachowskich, pod znamienym tytułem *Matrix*, w którym maszyny stworzyły świat oparty na potężnych strumieniach danych. Rewolucja cyfrowa rozpoczęła się na dobre.

Cyfrowy świat

Wraz z usunięciem ograniczeń technicznych i upowszechnieniem usług, z których przeciętny użytkownik może skorzystać w łatwy sposób za pośrednictwem telefonu, zegarka czy lodówki²⁹, sposób funkcjonowania zwykłych ludzi w XXI wieku uległ znaczącej zmianie, która przyjmuje dwa oblicza: rewolucji technologicznej opartej na samoodtwarzającym się konsumpcjonizmie (zob. rozdz. *ICT w kontekście podziałów Północ–Południe* oraz rozdz. *Konsumpcjonizm. Dyktat globalnych marek*) i wzrastających możliwościach kontroli obywateli przez instytucje państwa³⁰.

Problemem globalnym jest współzależność owych trendów, które w czasach upowszechniania Internetu (lata 80. XX wieku) nie występowały. Przede wszystkim przestrzeń cyfrowa była miejscem, w którym twórcy nie byli ograniczani; ich wolność kreowania rzeczywistości wynikała z niedopasowania reguł prawa do nowych wariantów cyfrowego świata, które pojawiały się bardzo szybko. Wzrastający w siłę ruch hackerski uosabiał nadzieję na rozwiązywanie w Internecie wielu problemów realnego świata, zgodnie z zasadą: *co jest niemożliwe tutaj, tam*

²⁸ J. Mashey, *Big Data... and the Next Wave of InfraStress*, http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf, [dostęp: 12.07.2019].

²⁹ P. Grabiec, *W temacie lodówek nie da się wymyślić nic nowego? To patrz, co zrobił Samsung*, <https://www.spidersweb.pl/2017/01/smart-lodowki-samsung-agd.html>, [dostęp: 12.07.2019].

³⁰ Z. Bauman, *Straty uboczne. Nierówności społeczne w epoce globalizacji*, tłum. J. Hunia, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2012, s. 15.

nie obowiązuje³¹. Konsumpcjonizm w takich warunkach nie miał racji bytu, dopóki dostęp do cyfrowego świata ograniczał się do ludzi posiadających wysokie kompetencje teleinformatyczne. Wraz ze wzrostem liczby użytkowników zaczęto wymyślać coraz to prostsze sposoby zarządzania aktywnością, m.in. wprowadzenie myszki komputerowej, które znacznie ograniczyło potrzebę korzystania z klawiatury przy czynnościach funkcyjnych, a w XXI wieku pojawienie się ekranu dotykowego, który znacznie ograniczył liczbę ruchów, jakie wykonuje przeciętny użytkownik.

Liczba użytkowników Facebooka przekroczyła dwa miliardy

Gdyby Facebook był państwem, należałoby go umieścić na pierwszym miejscu rankingu pod względem liczby ludności. Liczba użytkowników Facebooka stale rośnie. Pierwszy milion pojawił się już w pierwszym roku działalności (2004), 500 milionów użytkowników korzystało z Facebooka w 2010 roku, a pierwszy miliard osiągnięty został w 2012 roku. Przekroczenie w 2017 roku liczby dwóch miliardów stanowi dowód na siłę przyciągania wciąż nowych użytkowników, choć o mniejszym przyroście niż w ubiegłych latach. Jest to również serwis, który dzięki posiadanym informacjom na temat aktywności swoich użytkowników z pewnością wiedziałby o swoich obywatelach więcej aniżeli jakiegokolwiek inne państwo (z wyjątkiem Chin – zob. przykład *System Kredytu Społecznego*).

Zarządzanie danymi w perspektywie instytucjonalnej

Należy pamiętać, iż instytucje państwa, ze szczególnym uwzględnieniem służb bezpieczeństwa, nader często korzystają z analizy danych w oparciu o duże bazy, katalogi i zbiory. Aktualnie, dzięki informacjom przekazywanym przez sygnalistów, czyli osoby świadomie upubliczniające tajne dane (m.in. Edwarda Snowdena w 2013 roku), wiemy, iż instytucje nie tylko je przetwarzają, ale także dążą do tworzenia systemów, w których aktywnie pozyskuje się i łączy ze sobą odpowiednie partie materiałów. Problemem z całą pewnością jest zakres ingerencji służb w prywatność użytkowników: czy i w jaki sposób dane mogą być

³¹ J. Janczyk, *W głębi Internetu – inne zastosowania informatyki*, „Dydaktyka Informatyki” 2014, nr 9, s. 115–117.

Rok 1984, George Orwell, 1949

Kiedy w latach 40. XX wieku George Orwell tworzył antyutopię współczesności, opierał się na własnej intuicji i wiedzy z lat minionych. Dobrze przewidział nie tyle konkretne przedmioty, będące w użyciu w Oceanii, co procesy społeczne, które przez wdrożenie mechanizmów kontroli obywateli zostaną umasowione. Owa wszechobecność i oczywistość nadzoru ze strony służb, które dbają o pokój, prawdę, miłość i obfitość nas wszystkich, w realnym świecie XXI wieku nie podlegają dyskusji. Dzięki zaawansowanym możliwości przetwarzania danych sami obywatele godzą się z ograniczeniem własnej prywatności na rzecz zapewnienia pozornego bezpieczeństwa. Po przeczytaniu *Roku 1984* w roku 2019 możemy mieć wątpliwości, czy Ocena, tak naprawdę, sprawowała aż tak dużą kontrolę nad własnymi obywatelami?

zbierane? Czy katalog przetwarzanych danych powinien mieć charakter otwarty czy powinien być zamknięty, regulowany mocą ustawy? Jakie kwalifikacje powinny mieć osoby upoważnione do zarządzania bazami danych oraz przetwarzania konkretnych partii materiałów? Są to kluczowe pytania, na które odpowiedzi coraz częściej poszukiwane są w ramach kluczowych dyskusji politycznych z udziałem podmiotów trzeciego sektora. Na szczególną uwagę zasługuje Fundacja Panoptykon, która w Polsce dba o przestrzeganie standardów bezpieczeństwa prywatnego i publicznego, ze

szczególnym uwzględnieniem przestrzeni sieciowej³².

Warto również zwrócić uwagę na współpracę instytucji publicznych z sektorem prywatnym, dzięki czemu możliwe jest efektywniejsze wykorzystywanie zasobów, m.in. poszukiwanie informacji na portalach społecznościowych o osobach ukrywających się przed wymiarem sprawiedliwości, czy też korzystanie z infrastruktury publicznej przez gigantów internetowych³³. Jeszcze inaczej kwestia jest rozpatrywana w państwach autorytarnych, gdzie zasoby *Big Data* są wykorzystywane do sprawowania pełniejszej kontroli nad działaniami obywateli³⁴. Co więcej, narzędzia stworzone w świecie państw demokratycznych, mogące służyć społeczeństwu (m.in. przy wdrażaniu koncepcji *smart city*), są sprzedawane rządowi państw, które wdrażają model większej inwigilacji obywa-

³² Fundacja Panoptykon, <https://panoptykon.org/> [dostęp: 12.07.2019].

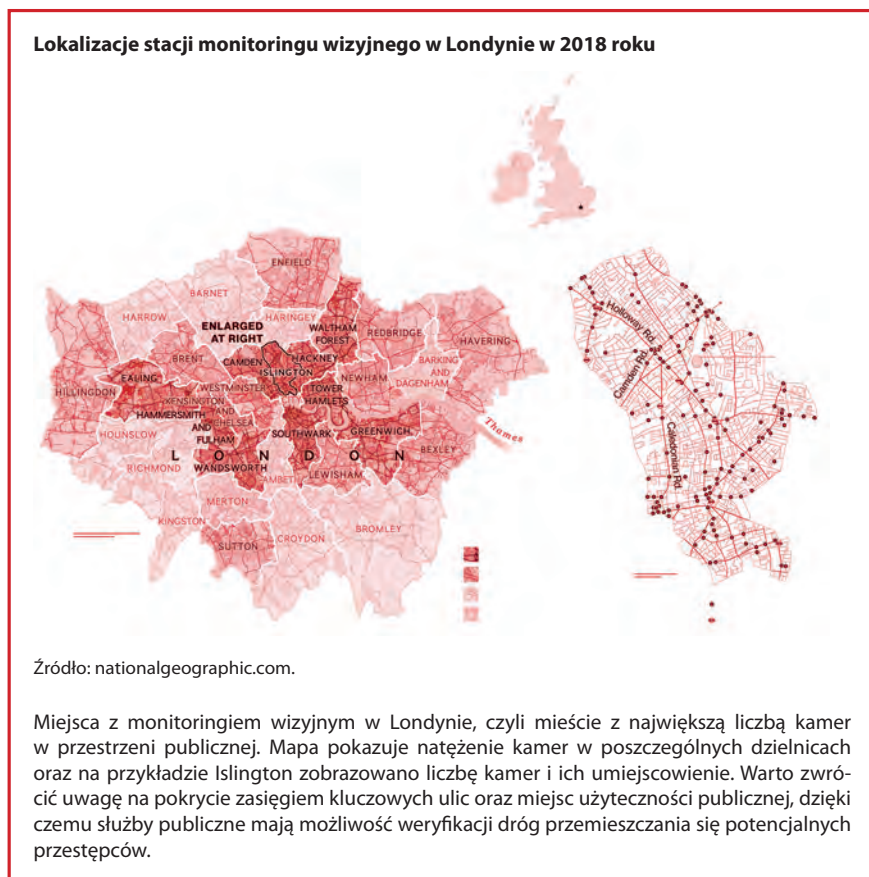
³³ K. Polańska, A. Wassilew, *Analizy big data w serwisach społecznościowych*, „Nierówności Społeczne a Wzrost Gospodarczy” 2015, nr 4 (44), s. 119.

³⁴ Egipt i wyłączenie Internetu w czasie *Arabskiej Wiosny*.

teli, ze szczególnym uwzględnieniem ugrupowań opozycyjnych wobec dyktatury.

Dostęp do publicznych treści a sposób zarządzania danymi

Można spotkać się z niejednoznacznymi przykładami przetwarzania danych, które są związane z funkcjonującym modelem dostępu do treści w Internecie. Jeszcze do niedawna całkiem powszechną



praktyką administratorów było stosowanie abonamentów, których sprzedaż umożliwiała pokrycie kosztów wytworzenia i dostarczenia informacji. Z chwilą umasowienia dostępu do Internetu model ten został znacząco ograniczony i zastąpiony przez optykę dostarczania bezpłatnych treści w zamian za wyświetlanie na ekranach urządzeń użytkowników reklam, a wraz z rozwojem technologicznym podzielenie się przez ludzi śladami swojej internetowej aktywności. Wydaje się, że aktualnie jest to model powszechnie akceptowany, gdyż wciąż zakłada dobrowolność uczestnictwa wszystkich stron. Użytkownik, chcący korzystać z informacji dostarczanych przez daną stronę, musi wyrazić zgodę na przetwarzanie i zarządzanie danymi, które są związane z jego internetowym awatarem, ale z reguły nie mają wpływu na jego aktywność w rzeczywistym świecie. Dopiero od niedawna te elementy można łączyć i coraz więcej osób decyduje się jeszcze mocniej ograniczać własną prywatność na rzecz m.in. potencjalnych zniżek przy zakupie towarów lub unikalnym dostępie do treści. Jednoznacznym zagrożeniem będzie zatem sytuacja, w której użytkownikom nie zostanie pozostawiona alternatywa wyboru. Przeprowadzenie aktu wyborczego wyłącznie za pośrednictwem Internetu; komunikowanie się z urzędnikami publicznymi tylko w trybie *online*; przeprowadzanie transakcji finansowych bez możliwości użycia gotówki, to tylko niektóre z opcji, które powinny budzić poważne zastrzeżenia. Warto zwrócić uwagę na usługi oferowane przez Cambridge Analytica, firmę bazującą na profilowaniu i targetowaniu oferty politycznej za pośrednictwem dużych zbiorów danych pozyskanych bezpośrednio z sieci społecznościowych. Dzięki wykonanej pracy możliwe stało się wykorzystanie precyzyjnie zogniskowanych komunikatów skierowanych w stronę zarówno zwolenników, jak i przeciwników kandydatury Donalda Trumpa na urząd prezydenta Stanów Zjednoczonych w 2016 roku. Mieliśmy wtedy do czynienia z przykładem efektywnego wykorzystania danych, które, gdy weźmiemy pod uwagę realną nieświadomość całego procesu użytkowników, budziło znaczne kontrowersje natury moralnej. Dlatego też z jednej strony wdrożenie w życie projektów w obszarze *Big Data* może spowodować wykluczenie dużej grupy obywateli, którzy nie korzystają z dobrodziejstw cyfrowego świata, z drugiej natomiast brak realnej alternatywy może świadczyć

Zarządzanie danymi osobowymi uchodźców w obozie Zaatari



Płatności elektroniczne w obozie dla uchodźców w Zaatari

Zdjęcie dzięki uprzejmości World Food Programme.

Każdy uchodźca, zarejestrowany w obozie, dysponuje swoim profilem sieciowym, dzięki czemu może w przystępny sposób realizować płatności w formie elektronicznej. W systemie zarejestrowanych aktualnie jest ponad 40 tysięcy uchodźców, ze szczególnym uwzględnieniem kobiet, które zazwyczaj za pośrednictwem systemu dokonują zakupów. Nie jest do tego potrzebny papierowy dokument tożsamości czy dysponowanie smartfonem, gdyż osoby są rozpoznawane na podstawie zarejestrowanych danych biometrycznych, a płatności odbywają się w formie bezgotówkowej. Co istotne, dzięki systemowi agregacji i przetwarzania danych administracja obozu oraz władze Jordanii mają również możliwość kontroli przepływu ludności do i z obozu. Istnieje również podejrzenie, iż dane biometryczne są udostępniane służbom bezpieczeństwa innych państw, dzięki czemu rejestracja w obozie Zaatari wiąże się z potencjalnymi trudnościami w kolejnych rejestracjach uchodźczych, m.in. w Turcji czy państwach Europy Zachodniej.

o zbyt dużej władzy instytucji publicznych, co wymagałoby stworzenia odpowiednich zabezpieczeń prawnych.

Sferą, w której większość użytkowników dostrzega pozytywną wartość wykorzystywania mechanizmów *Big Data*, jest szeroko rozumiane bezpieczeństwo. Zarówno w ujęciu wewnętrznym, jak i międzynarodowym punktem wyjścia będzie stwierdzenie, iż ludzie chcą żyć bez obawy o zagrożenie, które dopiero może nadejść. Dlatego też zgadzają

się zrezygnować z (niemałej) części prywatności na rzecz zapewnienia wyższego poziomu bezpieczeństwa. Uczestnicząc w ruchu drogowym, liczymy się z tym, iż możemy zostać ukarani za wykroczenie, które zostanie zdalnie zarejestrowane, a cały proces wymierzania sprawiedliwości będzie w znacznym stopniu zautomatyzowany. Podobnie, gdy trwa akcja poszukiwawcza zaginionego dziecka, a jedyną możliwością będzie sięgnięcie do nagrań monitoringu i danych zamieszczanych w portalach społecznościowych. Istnienie wielu rodzaju furtek dostępu, z których korzystają pracownicy służb bezpieczeństwa, jest przedmiotem dyskusji, ale niewielu ludzi decyduje się zanegować istnienie całego systemu, w którym taki dostęp jest w ogóle możliwy.

Dane w przyszłości

Wieloznaczność koncepcji efektywnego zarządzania danymi w czasach, kiedy technologia nie ogranicza, a stwarza realne szanse przeobrażenia otaczającej nas rzeczywistości, jest kluczem do zrozumienia zmian, jakie będą zachodzić w kolejnych latach w globalnym świecie Internetu. Przedstawione w rozdziale rozważania miały jedynie określić linie podziału: prywatne *versus* publiczne oraz indywidualne *versus* grupowe.

To, w jaki sposób i przez kogo gromadzone i przetwarzane będą dane, stanie się jednym z wyzwań przyszłości, a to, czy cyfrowy świat bardziej będzie przypominał *Matrix*, *Oceanię z Roku 1984*, a może *Wakandę* z marvelowskiego hitu *Czarna Pantera*, w dużej mierze zależy od świadomości samych użytkowników. W wymiarze zarządzania problemami globalnymi szczególną rolę należy zatem przypisać potrzebie ustawicznej edukacji użytkowników społeczeństwa sieciowego.

System Kredytu Społecznego w Chinach

Występuje również pod nazwą Systemu Wiarygodności Społecznej i jest niczym innym, jak największym projektem społecznej kontroli w historii ludzkości. Ideą przewodnią systemu jest stworzenie wirtualnych profili członków społeczności, które będą oddawały pozytywne lub negatywne aspekty ich zachowań w życiu codziennym, jak też w wirtualnej rzeczywistości. Przyjęto założenie, iż istnieją technologiczne możliwości katalogowania ludzkich dokonań i ocenianie aktów życia codziennego, m.in. łamanie przepisów ruchu drogowego, pozostawianie śmieci na ulicach, czy też używania przemocy. Negatywne zachowania mają wiązać się z przypisaniem do profili punktów ujemnych, a pozytywne – dodatnich. Tym samym członkowie społeczności biorą udział w grze, której wynik rzutuje na ich realne życie. Projekt wdrażany jest w Chinach od 2014 roku, a w 2020 roku system ma być gotowy do wejścia w fazę pełnego użycia. W pierwszych latach bazował na dobrowolnym uczestnictwie (projekt testowano w 40 miastach, a kontrolą objęto prawdopodobnie ponad 10 milionów ludzi); obecnie objętych systemem jest coraz więcej mieszkańców, również z terenów wiejskich, gdzie system wykorzystuje się również do działań związanych z identyfikacją indywidualną obywateli. Wszystkie dane elektroniczne, które są możliwe do zebrania (system telefonii komórkowej, system monitoringu, płatności mobilne, serwisy społecznościowe etc.), ulegają przetworzeniu i dopasowaniu do wzorca zachowań. Osoby są oceniane, a ich wyniki umożliwiają przypisanie do konkretnej kategorii. Potwierdzone zostały informacje, że osoby z kategorii D (najgorszej) mają niską zdolność kredytową i nie mają możliwości kariery w instytucjach publicznych. W Chinach testowane są również algorytmiczne systemy ustalania limitów kredytowych (Alipay) oraz aplikacje bazujące na bezgotówkowej wymianie środków pieniężnych, dokonywanej niejako przy okazji innych działań, m.in. w ramach komunikacji wirtualnej za pomocą programu WeChat, który ma już ponad miliard użytkowników! Dalszy rozwój systemu kredytu społecznego i rozciągnięcie go na całą populację jest już możliwe pod względem technicznym; zastrzeżenia moralne po stronie chińskiej praktycznie nie występują.



Słowniczek

- Chmura – usługa sieciowa, umożliwiająca utrzymywanie danych na zewnętrznych serwerach; oprócz zalet należy zwrócić uwagę na przekazanie dostępu do plików zewnętrznemu usługodawcy.
- Dane – kluczowy zasób świata XXI wieku; są zbierane, przetwarzane i wykorzystywane zarówno przez nas, jak i zewnętrzne podmioty; za naszą zgodą i bez niej.
- Data Mining – eksploracja danych, która umożliwia odkrycie zależności pomiędzy poszczególnymi zasobami, wartościami lub konkretnymi informacjami.

RODO – Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych).



Kalendarium

- 2001 – powstanie Wikipedii, niekomercyjnego projektu opartego na dzieleniu się wiedzą przez użytkowników.
- 2008 – rok, który uznaje się za początek masowego rozwoju mediów społecznościowych (największy w historii względny przyrost użytkowników w serwisie Facebook).
- 2009 – w Polsce powstaje Fundacja Panoptykon, która zajmuje się ochroną prywatności obywateli, ze szczególnym uwzględnieniem Internetu.
- 2016 – przyjęcie przez Parlament Europejski aktu prawnego (RODO) i ustanowienie tym samym unijnego systemu ochrony danych osobowych.



Problemy do dyskusji

1. W jakim kierunku będzie zmierzał rozwój sektora zarządzania danymi? Jak wiele danych możliwych jest jeszcze do zebrania, przetworzenia i wykorzystania?
2. Czy zarządzanie danymi w przestrzeni globalnej stanowi narzędzie rywalizacji wielkich mocarstw o prymat nad światem?
3. W jaki sposób i w jakim zakresie przeciętny użytkownik Internetu jest w stanie świadomie korzystać z jego dobrodziejstw i unikać niebezpieczeństw?



Dodatkowa literatura

- Bilton N., *Twitter. Prawdziwa historia o bogactwie, władzy, przyjaźni i zdradzie*, tłum. A. Romanek, Wydawnictwo Onepress, Warszawa 2018.
- Bauman Z., *Straty uboczne. Nierówności społeczne w epoce globalizacji*, tłum. J. Hunia, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2012.
- Domingos P., *Naczelny Algorytm. Jak jego odkrycie zmieni nasz świat*, tłum. R. Ociepa, Wydawnictwo Helion, Warszawa 2016.
- Kozielski R., Olsztyński A. i in., *Internet ludzi. Organizacja jutra*, Wydawnictwo Nieoczywiste, Warszawa 2018.
- Lem S., *Dialogi*, Wydawnictwo Agora, Warszawa 2010.
- Schönberger V., Cukier K., *Big data. Efektywna analiza danych. Rewolucja, która zmieni nasze myślenie, pracę i życie*, tłum. M. Gładki, Wydawnictwo MT Biznes, Warszawa 2018.