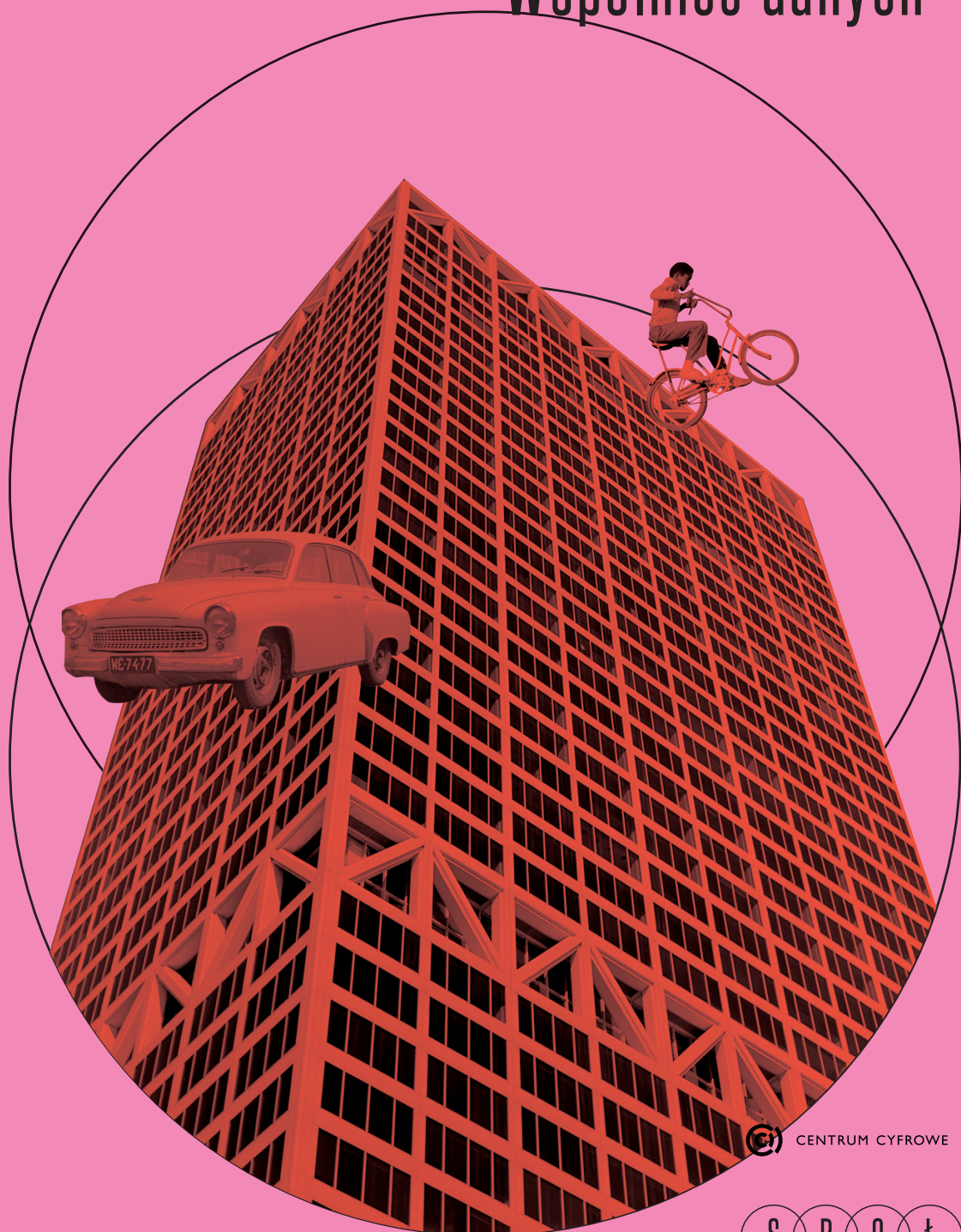


Jan J. Zygmuntowski

Wspólnice danych



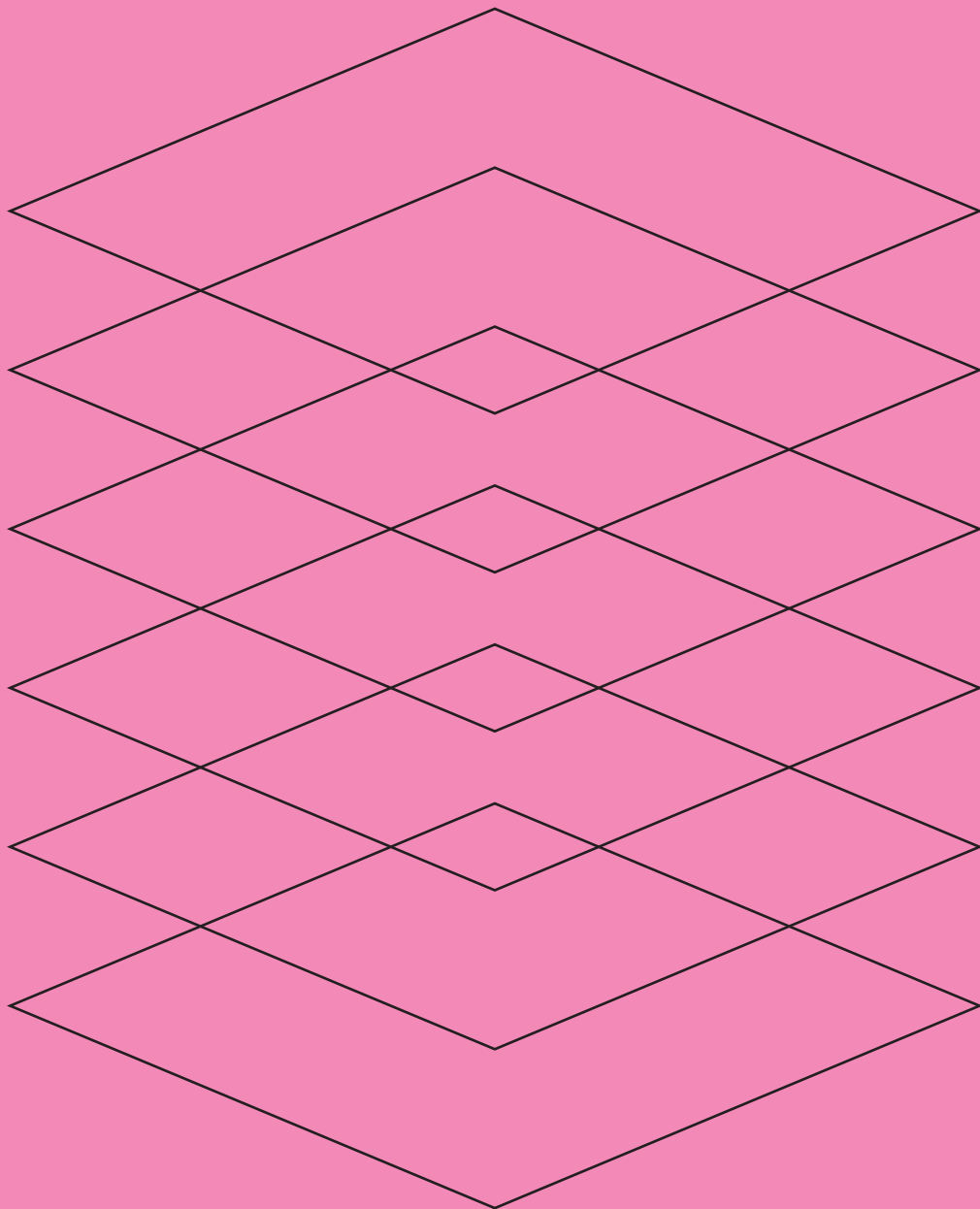
 CENTRUM CYFROWE

Alternatywny model
zarządzania danymi

RAPORT PROJEKTU:



Wstęp



N

Najcenniejszym zasobem świata nie jest już ropa, ale dane.

„The Economist” (2017)

Przydatniej jest myśleć o danych mniej jako o towarze do kupna lub sprzedaży, a bardziej jako o współdzielonym zasobie lub dobru wspólnym.

Theo Bass, Nesta (2020)

Nie da się dziś rozmawiać o gospodarce cyfrowej bez odwołania do cyfrowych danych. Przez infrastrukturę internetu – strony, platformy, aplikacje i portale – przepływają ogromne wolumeny danych, które co więcej z każdym rokiem rosną coraz dynamiczniej. Jeszcze w 2018 r. na świecie powstawały w ten sposób 33 zettabajty danych, ale już w 2025 r. będziemy wytwarzać ich 175 zettabajtów, czyli ilość informacji możliwą do zapisania na blisko 2 miliardach dysków twardej przeciętnego laptopa... dziennie (Komisja Europejska, 2020).

To dane tworzą bogactwo gospodarki cyfrowej. Dane to fragmenty kodu budującego interfejsy, to autoryzacja użycia karty kredytowej przy zakupach online, to wymiana informacji między Tobą a reklamodawcą, to wreszcie strumień rejestrowanych drgań maszyny w hali fabrycznej czy zapis wyników morfologii krwi.

Jeśli to dane tworzą bogactwo gospodarki cyfrowej, to również w nich musi tkwić klucz do zrozumienia sposobu, który pozwoli rozwiązać nękające współczesność problemy. Martwi nas erozja prywatności w warunkach kapitalizmu nadzoru, monopolizacja wielu sektorów gospodarki, w tym prywatyzacja sfery publicznej, a także będące w stagnacji usługi publiczne mimo widniejących na horyzoncie

innowacji technologicznych. Zmianami w dostępie czy jakości aplikacji bazujących np. na sieciach społecznościowych albo GPS są pozorne, a nasza rzeczywistość nie uległa fundamentalnej zmianie. Cyfrowa gospodarka nie jest bardziej demokratyczna, a tym bardziej nie pomaga zniwelować nierówności i stworzyć godnych warunków życia dla wszystkich. Wciąż jesteśmy daleko od wizji świata bez przymusu pracy czy sztucznej inteligencji (AI) pozwalającej nam na cywilizacyjny skok naprzód.

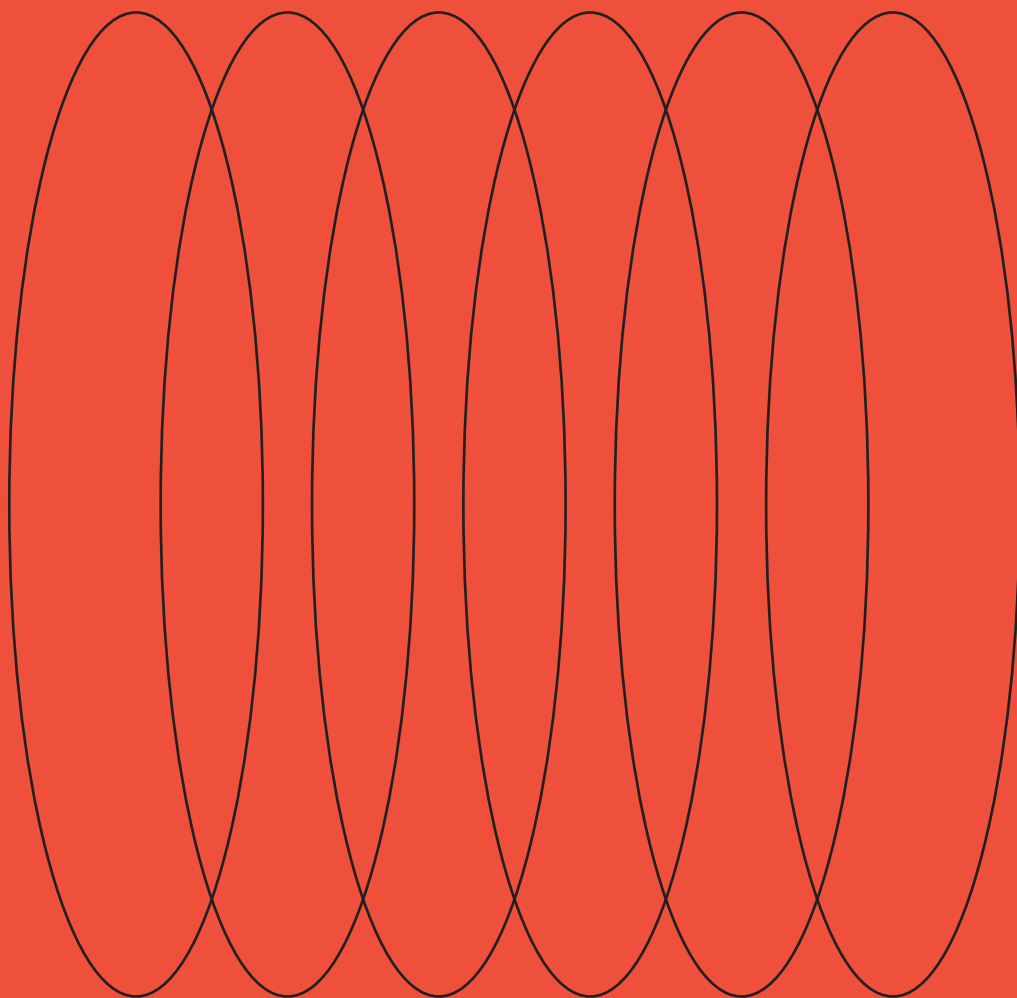
Dane są w sercu tej zagadki. Wartość ekonomiczna danych jest motorem rozwoju gospodarki cyfrowej, bo kolejne informacje tworzą przewagi konkurencyjne, coraz lepsze produkty i usługi. Ale tutaj pojawiają się fundamentalne pytania. **Kto tworzy tę wartość, a kto ją przechwytuje czy ekstrahuje? Do czego służą dane? Czy możemy wykorzystywać je w interesie publicznym, napędzając społecznie korzystne innowacje?**

Od samego początku internetu towarzyszyło mu napięcie funkcjonowania w realiach kapitalizmu, co możemy zobrazować rywalizacją otwartej współpracy Wikipedii i gospodarki „współdzielenia” Ubera (Jemielniak & Przegalinska, 2020). Analizując platformy zbierające dane o nas, korporacyjne zamknięte silosy i strzeżone prawa własności intelektualnej, musimy pamiętać, że alternatywa w postaci większej współpracy, demokracji i równości również może być mediowana technologicznie. Prezentowany w tym raporcie model wspólnic danych jest częścią szerokiego procesu współzarządzania, współpracy i uspołeczniania, który ma swój wymiar społeczny, ekonomiczny, ale też polityczny i kulturalny.

Alternatywny model zarządzania danymi nie jest hipotetycznym pomysłem, ale precyzyjnym planem tworzenia instytucji na miarę XXI wieku. Wspólnice danych ułatwią współdzielenie i wykorzystanie danych i jednocześnie ochronią ich wartość w interesie publicznym. Dzięki temu możliwe będzie rozwiązanie problemów dotyczących danych, które trapią gospodarkę cyfrową, i stworzenie cyfrowej przyszłości dobrobytu.

Raport rozpoczynam od dyskusji na temat potrzeby stworzenia ekosystemu zaufania dla danych, wychodzącego poza koncepcje prywatności i własności danych. Następnie z różnych perspektyw omawiam wspólnice danych jako model instytucji, pokazując schematy działania i elementy ich architektury. W trzecim rozdziale opisuję krótko trzy przykłady możliwych wspólnic danych, które realnie mogą powstać w polskiej przestrzeni publicznej. Na końcu podsumowuję krótko ryzyka oraz regulacyjne możliwości związane z publicznym gospodarowaniem danymi.

1. Ekosystem zaufania dla cyfrowych danych



P

Prowadzony przez kilka lat projekt Sidewalk Labs, spółki-córki Alphabet (konglomeratu posiadającego m.in. Google'a), miał odmienić oblicze Toronto. Pomysł na stworzenie wzorowego „inteligentnego miasta” w dzielnicy Quayside w Toronto był unikalną w skali świata ofertą złożoną Kanadyjczykom zamieszkującym miasto. Jednak projekt szybko napotkał opór. Okazało się, że mnogość sensorów, kamer, elektronicznych kart i aplikacji ma z jednej strony zapewniać komfort korzystania z przestrzeni i usług miejskich, ale z drugiej zbierać dane i wysyłać je do korporacyjnych silosów. Ryzyko naruszenia prywatności, ale także potencjalnej segregacji mieszkańców oraz uzależnienia całej dzielnicy – a być może Toronto – od jednego dostawcy było bardziej niż realne.

Kanadyjscy badacze proponowali rozwiązanie określone żartobliwie jako *BiblioTech*. Mianowicie, potraktujmy dane jako zasób wspólny i przekazmy go zaufanej instytucji miejskiej, w tym przypadku Publicznej Bibliotece Toronto (Ruttan et al., 2019). Podobne głosy od dawna płynęły z Barcelony, gdzie dane miejskie są traktowane jako zasób należący do mieszkańców. Również Komisja Europejska w opublikowanej niedawno strategii danych postuluje stworzenie „europejskich przestrzeni wspólnych danych” w obszarach takich jak m.in. zielony ład, finanse, energetyka, zdrowie, transport, rolnictwo czy umiejętności (Komisja Europejska, 2020).

il. 1



Potrzeba stworzenia instytucji, regulacji i praktyk społeczno-gospodarczych dla danych jest zatem duża. Przyszłość cyfrowych danych opierać się będzie na zaufaniu, że ich wykorzystanie i zarządzanie jest zgodne z naszymi wspólnymi potrzebami.

1.1 Czym są dane?

Dane niejednokrotnie uchodzą za „ropę”, czyli wydobywany za darmo ze środowiska naturalnego zasób. W ujęciu dominującej praktyki biznesowej uchodzą też za kapitał, aktywo strzeżone pilnie na korporacyjnych serwerach – chociaż prawo międzynarodowe nie uznało wciąż, że dane mogą być przywłaszczane przez tego, kto ma techniczny aparat do ich rejestracji. Z pewnością musimy zauważyć, że mamy do czynienia z dwoma rodzajami danych: danymi osobowymi i nieosobowymi.

W kontekście danych osobowych konieczne jest podkreślenie udziału człowieka w ich wytwarzaniu. Ekonomiści i badacze społeczni w ostatnich latach dużo uwagi poświęcili pracy wykonywanej np. w formie pracozabawy (*playbour*) oraz korzystaniu z platform cyfrowych. Jak uzasadnia zespół Instratu:

„Przyjmowanie, analiza treści i reagowanie na nie z uwagą to aktywne czynności człowieka. Jeżeli generują wartość ekonomiczną, to są pracą informacyjną, która może przybierać formę: analizy symboli – rozpoznawania i przetwarzania treści przy użyciu kognitywnych zasobów mózgu (określanych często jako »uwaga«), komunikacji – przekazywania informacji i danych, manipulacji afektami – intencjonalnego używania ekspresji emocjonalnej”. (Zygmuntowski et al., 2020)

Dane osobowe są też nośnikiem wiedzy o konkretnym obywatelu czy użytkowniczce. Z tej perspektywy muszą być też szczególnie chronione, np. przez regulacje takie jak europejskie RODO, aby zapobiec nadużywaniu. Identyfikacja konkretnej osoby z poważną chorobą

przewlekłą mogłaby prowadzić do szantażu, podwyższania kosztów ubezpieczenia zdrowotnego czy dyskryminacji w dostępie do usług.

Dane nieosobowe z kolei często powstają w toku działania maszyn przemysłowych, produktów, ale też mogą pochodzić ze środowiska naturalnego poza człowiekiem, jak np. dane meteorologiczne czy dotyczące fauny i flory. Część z nich można rejestrować niemal swobodnie, jak poziom zanieczyszczenia powietrza czy liczbę aut na drodze, jednak niektóre są dostępne tylko producentowi oprogramowania czy właścicielowi hali fabrycznej. Jeśli zawierają informacje będące tajemnicą handlową przedsiębiorstwa, mogą mieć też poufny charakter, choć taka argumentacja jest powszechnie nadużywana.

Jednak gdy mowa o wykorzystaniu danych, często wskazuje się na ich cechy, które przypominają dobra publiczne lub dobra wspólne (wspólne zasoby). Replikowalność danych i ich użyteczność w dużych wolumenach sprawia, że naturalnym kierunkiem rozwoju gospodarki cyfrowej musi być wspólne wykorzystanie danych, trenowanie dzięki nim AI i ponowne wykorzystanie przez innowatorów, działaczy społecznych czy naukowców. Dziś wciąż jesteśmy daleko od takiej rzeczywistości, a dominujące do niedawna koncepcje były niewystarczające dla pozytywnego rozwoju ekosystemu zaufania dla danych.

1.2 Poza prywatność i własność danych

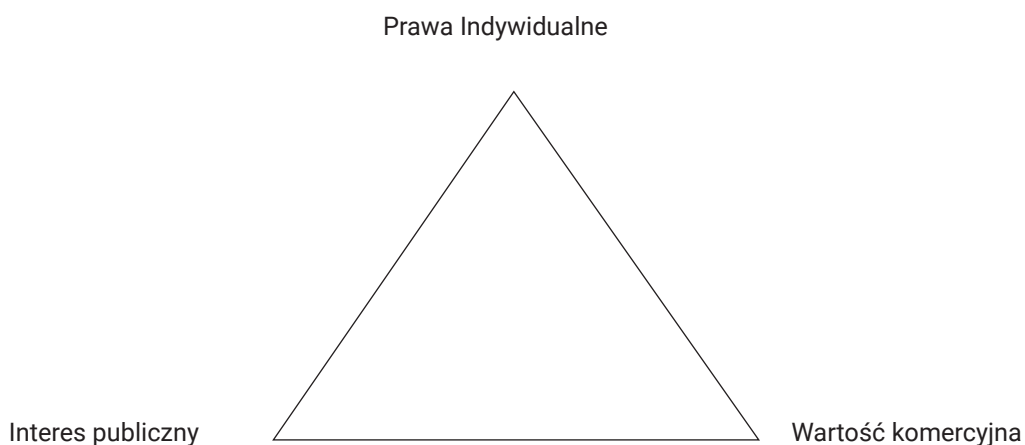
W ostatnich latach duże zainteresowanie w kontekście współdzielenia danych zyskały koncepcje takie jak prywatność czy własność danych. Według pierwszej z nich dane są przedmiotem ochrony, ponieważ mogą zdradzać identyfikujące daną osobę, osobiste i przez to wrażliwe informacje, które mogłyby zostać wykorzystane przeciwko interesom tejże osoby (np. przez profilowanie i wyświetlanie niechcianych reklam lub zdradzanie problemów zdrowotnych ubezpieczycielowi). Zgodnie z podejściem własności danych, w którym

dominuje kontekst komercyjny, obywatel/użytkownik może całkowicie dowolnie rozporządzać danymi, w tym modyfikować je, sprzedawać i dopuszczać lub zastrzegać dostęp.

Dominującą dziś praktyką rynkową jest wykorzystywanie platform cyfrowych, aplikacji i innych technologicznych aparatów do zbierania danych i trzymania ich przez firmy jako prywatnej własności. W większości krajów świata relacja między użytkownikiem a firmą jest skrajnie asymetryczna – tworząc czy podając dane w obrębie cyfrowej infrastruktury, nie ma się wyboru co do przeznaczenia danych. To dlatego największe firmy technologiczne mają tak wielką przewagę w konkurencji z mniejszymi podmiotami, jednocześnie często nie dbając o bezpieczeństwo czy godziwe wynagrodzenie społeczeństwu pracy informacyjnej (czasu, uwagi) włożonej w ich wytworzenie (Zygmuntowski et al., 2020).

Chociaż logika prywatności i własności danych są wartościową przeciwwagą dla dominujących obecnie korporacyjnych silosów danych, są w fundamentalny sposób niewystarczające. Mamy do czynienia raczej z trylematem zarządzania danymi, zobrazowanym poniżej.

Diagram 1. Trylemat zarządzania danymi



W przypadku radykalnego ograniczenia możliwości zbierania i przetwarzania danych z powodu obaw o prywatność musielibyśmy zarazem zrezygnować z korzyści technologii opartych o dane, zarówno w postaci produktów i usług firm, jak i usprawniania administracji i usług publicznych. Badacze renomowanych instytucji wskazują, że tradycyjnie filozofia głosząca, że „mój dom to moja twierdza” ma przecież drugą stronę – w imię interesu publicznego możliwe jest, na ściśle regulowanych zasadach, korzystanie również z prywatnej własności (Evans, 2011; Kerry & Morris Jr, 2019).

Propozycja własności danych natomiast łączy dominującą dziś komercyjną logikę w odniesieniu do danych, uzupełniając ją jedynie o prawa indywidualne. To propozycje, by każdy właściciel danych trzymał je we własnej kapsule, która zabezpieczy prywatność i pełnię praw danej osoby do jego lub jej danych osobowych czy wytworzonych samodzielnie. To podejście wciąż nie jest jednak optymalne z punktu widzenia interesu publicznego. Radykalne zastosowanie zasad prywatnej własności danych mogłoby doprowadzić ponownie do monopolizacji danych w rękach podmiotów gospodarczych zdolnych je wykupić lub zdobyć przy pomocy manipulacji od indywidualnych użytkowników. Zespół Darona Acemoglu z MIT zaprezentował model, który pokazuje wady hipotetycznego rynku danych, na którym każdy indywidualnie sprzedaje swoje informacje. Nisko ceniący swoją prywatność użytkownicy sprzedają swoje dane jako pierwsi, umożliwiając predykcję również nieznanej wiedzy przez analizę *big data*. W konsekwencji dochodzi do ogólnego spadku cen, a prywatność traci na znaczeniu (Acemoglu et al., 2019)

Siła negocjacyjna pojedynczego użytkownika jest znikoma, a przeciążenie liczbą koniecznych do podjęcia decyzji w rozbudowanej gospodarce cyfrowej jest raczej wyzwaniem niż wyzwoleniem. Jak wskazują badacze z MIT, w modelu z udziałem aktora pośredniczącego – mediatora stron – zarządzanie danymi odbywa się z większą korzyścią społeczną.

Również w przypadku danych nieosobowych, przykładowo przemysłowych lub rolniczych, ich prosta sprzedaż na rynku nie jest najbardziej efektywnym rozwiązaniem. Po pierwsze, własność tego typu danych nie jest jednoznacznie uregulowana, więc dominuje brak zaufania między podmiotami. Jednak przypisanie praw własności do danych, zdaniem większości pytanym o to europejskich firm, jest znacznie mniej ważne niż zaproponowanie jasnego trybu dostępu do danych przy użyciu ustandaryzowanych kontraktów i protokołów (Komisja Europejska, 2018). Prawne oraz techniczne warunki współdzielenia danych mogą być oferowane właśnie przez instytucję pośredniczącą w wymianie.

Wskazane powyżej problemy wynikają z niezrównoważonego rozwoju obecnego modelu zarządzania danymi. **Interes publiczny – tak dotyczący ochrony przed monopolizacją, jak i budowania zaufania i tworzenia standardów współdzielenia danych – musi zostać uwzględniony w budowie fundamentów gospodarki cyfrowej.** Interes publiczny oznacza również, że dane mogą być wykorzystywane ponownie do rozwiązywania społecznych i cywilizacyjnych problemów – zdrowotnych, środowiskowych czy związanych z jakością życia. Ani perspektywa praw, ani wartości w ujęciu komercyjnym nie zapewnia właściwego zabezpieczenia takiego rozwoju.

Podsumowując, uwzględnienie szerszego interesu publicznego rozwiązuje trylemat, ale pozwala na zbalansowanie potrzeb wszystkich stron. W takim scenariuszu współdzielenie danych i płynąca z niej innowacyjność gospodarki cyfrowej kwitnie, a produkty i usługi tworzą nie tylko cyfrowi giganci, lecz także małe i średnie firmy, naukowcy oraz administracja publiczna i spółki świadczące usługi publiczne. Jednocześnie poszanowane są prawa indywidualne do prywatności i otrzymania korzyści z udostępnianych danych.

Jak zatem wskazuje Evans, to nie pytanie o własność danych ma realne znaczenie, ale raczej kwestia tego, jak konkretnie zabezpieczyć zarówno indywidualną prywatność, jak i interes publiczny dostępu do danych (Evans, 2011). Kim jest mediator maksymalizujący korzyści społeczne?

1.3 Wspólnice danych jako fundament ekosystemu

Czy mediatorem mogą być podmioty komercyjne zbierające dane w celach biznesowych? Większość danych jest dziś zbierana przez platformy, które na ich podstawie udoskonalają własne produkty lub usługi cyfrowe, w tym przez np. ulepszanie algorytmów wyszukiwania, zmiany w ofercie i jej kierowanie do określonych segmentów rynku, lub sprzedaż reklam profilowanych. Takie podmioty nie są zatem neutralnym mediatorem – następuje konflikt interesów między maksymalizacją wartości komercyjnej (np. zysku firmy) a zobowiązaniami wobec właścicieli danych (np. prywatności lub określonego wykorzystania danych) (Delacroix & Lawrence, 2018).

Rzadko w obszarze tak zmiennym jak przecięcie gospodarki i nowych technologii mamy do czynienia z podobnym konsensusem środowiska naukowego, biznesu oraz organizacji obywatelskich, że cyfrowe dane muszą być zarządzane w ramach alternatywnego modelu. Instytucje nowego typu mają wobec tego pilnować poprawnego użycia danych i łączyć strony chcące oferować lub przetwarzać dane.

W pewnym sensie mamy do czynienia z potrzebą stworzenia zupełnie nowego modelu organizacji życia gospodarczego w odniesieniu do cyfrowych informacji. Nowym czynnikiem produkcji i sposobom zarządzania pracą w społeczeństwie współpracy muszą towarzyszyć adekwatne. Korzyści płynące z ich funkcjonowania podsumowują na podstawie London Economics (Godel & Natraj, 2019):

- **rozwiązanie konfliktu interesów w postaci problemu agencji** (*principal-agent problem*) – interesy właścicieli danych – użytkowników, obywateli – oraz firm posiadających dane i tworzących komercyjne produkty lub usługi są rozbieżne, jak w problemie pryncypała i agenta. Wprowadzenie trzeciej strony rozwiązuje konflikt;
- **zwiększenie produktywnego wykorzystania danych dzięki ich współdzieleniu** – dane są wykorzystywane przez podmioty, które nie miały do nich wcześniej dostępu, co podnosi produktywność w gospodarce;

- **pozytywne efekty zewnętrzne z agregacji danych** – łączenie danych w większe zbiory wytwarza pozytywne efekty zewnętrzne, czyli podnosi wartość danych, umożliwiając lepsze predykcje i testowanie modeli AI;
- **pozytywne efekty sieciowe łączenia różnych podmiotów gospodarczych** – zapośredniczona przez nowe instytucje współpraca np. dużych firm, start-upów, uczelni i organizacji pozarządowych łączy je w sieć dialogu oraz wymiany wiedzy i handlu;
- **umożliwienie szybszego rozwoju branży sztucznej inteligencji (AI)** – szczególnie branża AI mierzy się z koniecznością poszukiwania wysokiej jakości obszernych zbiorów danych, więc nowa instytucja bezpośrednio odpowiada na te potrzeby;
- **zwiększenie konkurencyjności na rynku przez dopuszczenie nowych podmiotów do danych** – korporacyjne silosy danych skutkują rosnącą monopolizacją w gospodarce cyfrowej. Aby przywrócić konkurencyjność, niezbędne jest stworzenie równych szans na starcie przez obniżanie kosztów wejścia, takich jak pozyskanie dostępu do danych;
- **podnoszenie zaufania do współdzielenia danych** – ponieważ wciąż współdzielenie danych odbywa się zbyt rzadko, potrzebne jest budowanie odpowiedniej kultury organizacyjnej, która promuje te praktyki i wyznacza dobre standardy dla innych podmiotów.

Najważniejszym wyzwaniem jest wobec tego zaproponowanie schematu budowy instytucji gwarantujących powyższe korzyści. Inne zasady są potrzebne menedżerom danych przemysłowych, wykorzystywanych przykładowo przez firmy z sektora motoryzacyjnego do oceny sprawności silników, a inne opiekunom wrażliwych danych medycznych opisujących historię przebiegu nowotworów. Niektóre podmioty miałyby charakter wyłącznie publiczny (państwowy), ale niektóre operowałyby jako partnerstwa publiczno-prywatne, spółki celowe lub spółdzielnie. Brytyjski ośrodek innowacyjności Nesta proponuje, by zamiast rynku danych rozwijać ekosystem zaufania dla cyfrowych danych – przestrzeń pluralistycznych instytucji mających u podstaw zasadę wspólnego dysponowania danymi (Mulgan & Straub, 2019)

Wykres 1. Ekosystem zarządzania danymi



Niewiele – główna wartość jest osobista

Ile wartości dla interesu publicznego stworzy współdzielenie danych?

Bardzo dużo wartości dla sfery publicznej

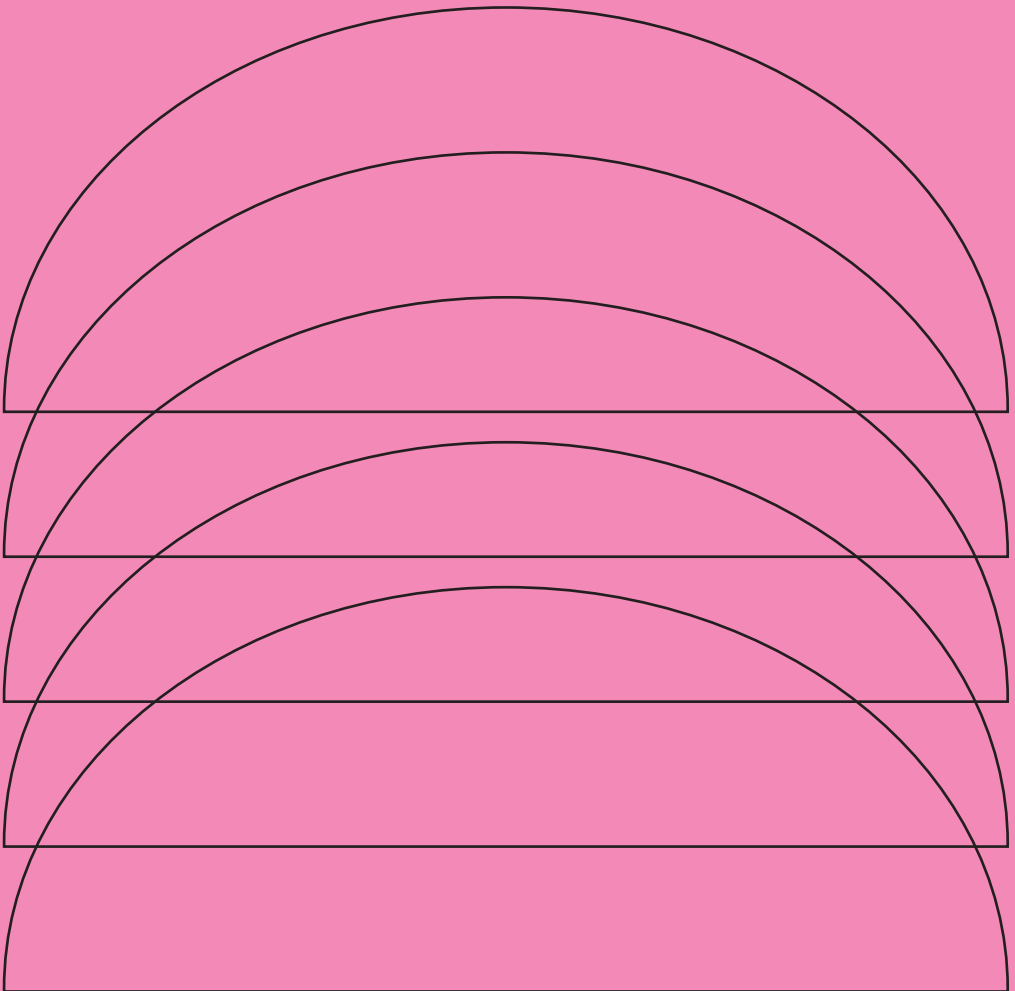
Nowy model zarządzania danymi może czerpać z doświadczeń istniejących modeli zarządzania zasobami. Badane dotychczas modele oraz ich zalety i wady przedstawia poniższa tabela.

Tabela 1. Wady i zalety dotychczasowych modeli zarządzania danymi

Model	Opis	Zalety	Wady
Prywatny silos danych	Zbierający dane traktuje je jako własność prywatną.	<ul style="list-style-type: none"> • proste w zarządzaniu 	<ul style="list-style-type: none"> • prywatyzacja wartości ekonomicznej danych • negatywny wpływ na konkurencję • naruszanie prywatności • niepełne wykorzystanie danych
Otwarte dane	Pula danych udostępnianych w otwartym dostępie.	<ul style="list-style-type: none"> • maksymalne upowszechnianie danych 	<ul style="list-style-type: none"> • dominujące podmioty mogą użyć danych lepiej • ryzyko dla prywatności i poufności danych
Kooperatywa danych	Spółdzielnia zrzeszająca osoby współdzielące dane.	<ul style="list-style-type: none"> • właściciele danych mają udział w zarządzaniu • wartość ekonomiczna wraca do właścicieli 	<ul style="list-style-type: none"> • spółdzielcami mogą być tylko osoby fizyczne
Trust danych	Zarząd powierniczy oparty o prawo anglosaskie.	<ul style="list-style-type: none"> • dobrze zdefiniowane obowiązki pośrednika • nastawienie na współpracę różnych interesariuszy 	<ul style="list-style-type: none"> • brak odniesienia do prawa kontynentalnego • ryzyko erozji interesu publicznego

Doświadczenia związane z powyższymi modelami wskazują, że każdy z nich ma swoje silne strony, ale żaden nie jest optymalny dla szerokiego spektrum sektorów, w których współdzielenie danych jest potrzebne. Dlatego proponuję opracowanie nowego modelu – wspólnic danych, które będą czerpały z prowadzonych na świecie pilotażowych inicjatyw. Wspólnice danych są uniwersalnym schematem czerpiącym z podmiotowości właścicieli danych w kooperatywach danych, powtarzalności powierzenia i obowiązków w trustach danych oraz traktowania danych jako wspólnej własności, podobnie jak robią to otwarte dane.

2. Zagadnienia organizacyjne, prawne i techniczne wspólnic danych



P

Proponowany w tym raporcie model **wspólnic danych** jest konceptualizacją takiej instytucji nowego typu. Wspólnica jest zebraniem doświadczeń i najlepszych praktyk z istniejących pilotaży, prac naukowych i obserwacji rynkowych. Dzięki temu możemy zaproponować konkretne, szczegółowe rozwiązania dotyczące np. architektury technicznej czy modelu biznesowego.

Definicja: Wspólnica danych to zaufana instytucja zarządzająca danymi zgodnie z interesem publicznym.

Tworzenie wspólnic danych rozpoczyna się od zidentyfikowania ekosystemu danych w danej branży, sektorze lub obszarze życia społeczno-gospodarczego. To odpowiedź na pytania: „Które podmioty dysponują danymi?”, „Jakie są ich cele i do czego służą dane w osiągnięciu tych celów?”, „Kto zyska najwięcej na współdzieleniu danych, a kto straci?”.

Następnie inicjatorzy stworzenia wspólnic danych – najczęściej władze lokalne lub krajowe, czasem również organizacje pozarządowe lub firmy – muszą rozważyć kilka kluczowych elementów, z których składa się wspólnota. Są to:

- 1) **architektura** techniczna, czyli warstwa oprogramowania,
- 2) **właściwości architektury** i związane z nią działania operatora wspólnoty,
- 3) **pozyskiwanie danych i model biznesowy**,
- 4) **forma prawna operatora** wspólnoty,
- 5) **licencja** dla danych, czyli zagadnienia związane z prywatnością lub poufnością danych.

2.1 Architektura

Architektura wspólnicy danych odpowiada na najważniejsze pytania, jest swego rodzaju planem budowy. Dzięki niej możemy odpowiedzieć na pytanie, gdzie znajdują się dane, jak są przetwarzane i co z takiego układu wynika dla możliwości np. dawania dostępu do danych czy monitorowania ich użycia.

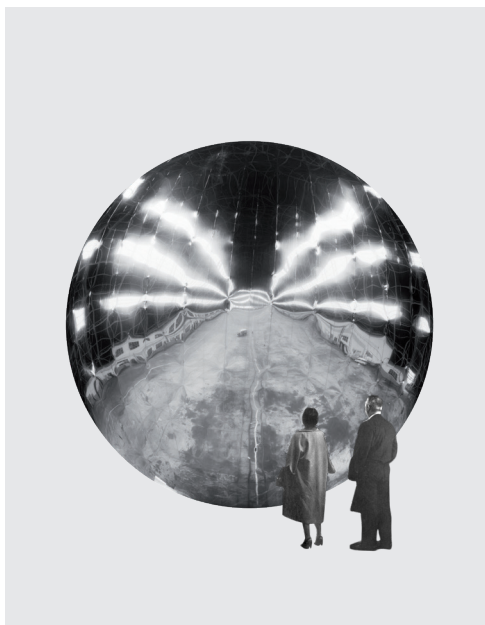
Architektura „określa podział oprogramowania na komponenty oraz definiuje funkcje tych komponentów i występujące między nimi relacje” (Sacha, 2010). Wbrew pozorom nie jest to zadanie wyłącznie dla specjalistów branży IT. Jak podkreślają twórcy zunifikowanego języka modelowania (UML), „architektura oprogramowania dotyczy nie tylko jego struktury i zachowania, ale także stosowania go, jego funkcjonalności, efektywności, odporności, możliwości ponownego użycia, ograniczeń ekonomicznych i technologicznych oraz estetyki” (Booch et al., 2005).

Tworząc architekturę wspólnicy, wybieramy zatem cele, które ma ona spełniać. Proponowana tutaj architektura jest zoptymalizowana jednocześnie na zapewnienie prywatności i selektywne przyznawanie dostępu do danych, ponieważ te cele można rozwiązać już w fazie projektowania (zgodnie z zasadą *privacy by design*), w przeciwieństwie do innych celów, które zabezpiecza odpowiedni model biznesowy czy forma prawna.

Rozważmy podstawowy model, w którym wspólnica danych jest pośrednikiem pomiędzy właścicielem danych – osobą, firmą lub inną organizacją mającą pewne prawa do danych – oraz klientem, który wysyła zapytanie o możliwość uruchomienia swojego algorytmu (np. uczenia maszynowego) na tych danych. Dla uproszczenia przyjmijmy, że serwer należy bezpośrednio do odpowiednich stron¹.

¹ Zrezygnowanie z uproszczenia nie prowadzi do fundamentalnej zmiany w tym modelu. Włączenie większej liczby stron do współdzielenia danych, przykładowo serwerów, może odbywać się z użyciem protokołu UAM (*User-Managed Access*), który bazuje na szeroko akceptowanym otwartym standardzie autoryzującym OAuth, tzn. korzysta z formatowania i tokenów tego standardu (Maler et al., 2017).

Architektura wspólnicy może zatem różnić się przede wszystkim po stronie właściciela danych oraz po stronie klienta. W pierwszym przypadku chodzi o sposób przechowywania danych (zcentralizowany lub rozproszony/federacyjny), w drugim natomiast o miejsce przetwarzania danych (serwer klienta lub serwer wspólnicy, lub serwery właścicieli danych). Te rozwiązania łączą się w konieczne możliwości, przykładowo zcentralizowane przechowywanie danych umożliwia jedynie przetwarzanie na serwerze klienta lub samej wspólnicy.



il. 2

Przy rozważaniu, jaki model najlepiej gwarantuje optimum społeczne, warto wziąć pod uwagę zastosowanie rozwiązań technicznych gwarantujących ochronę danych:

- **federacyjne uczenie** (*federated learning*) – rodzaj uczenia maszynowego, w którym algorytm jest trenowany na urządzeniach/serwerach krańcowych bez konieczności wysyłania danych (przesyłowi podlega algorytm i wyniki obliczeń);
- **prywatność różnicowa** (*differential privacy*) – sposób publikowania informacji o grupach w zbiorze danych tak, aby nie ujawniać danych pojedynczej osoby;
- **szyfrowanie homomorficzne** (*homomorphic encryption*) – szyfrowanie pozwalające na przeprowadzenie obliczeń na zaszyfrowanej treści bez konieczności jej odszyfrowania
- **bezpieczne obliczenia wielopodmiotowe** (*secure multi-party computation*) – protokół prowadzenia obliczeń między wieloma podmiotami bez ujawniania ich danych (*WIPO Technology Trends 2019: Artificial Intelligence, 2019*).

Te rozwiązania znajdują już zastosowanie w sektorze nowych technologii, co więcej, mamy z nimi do czynienia właściwie codziennie. Przykładowo, federacyjne uczenie jest stosowane przez Google do doskonalenia Gboard – podpowiedzi słów w klawiaturze systemu Android (Hard et al., 2018).

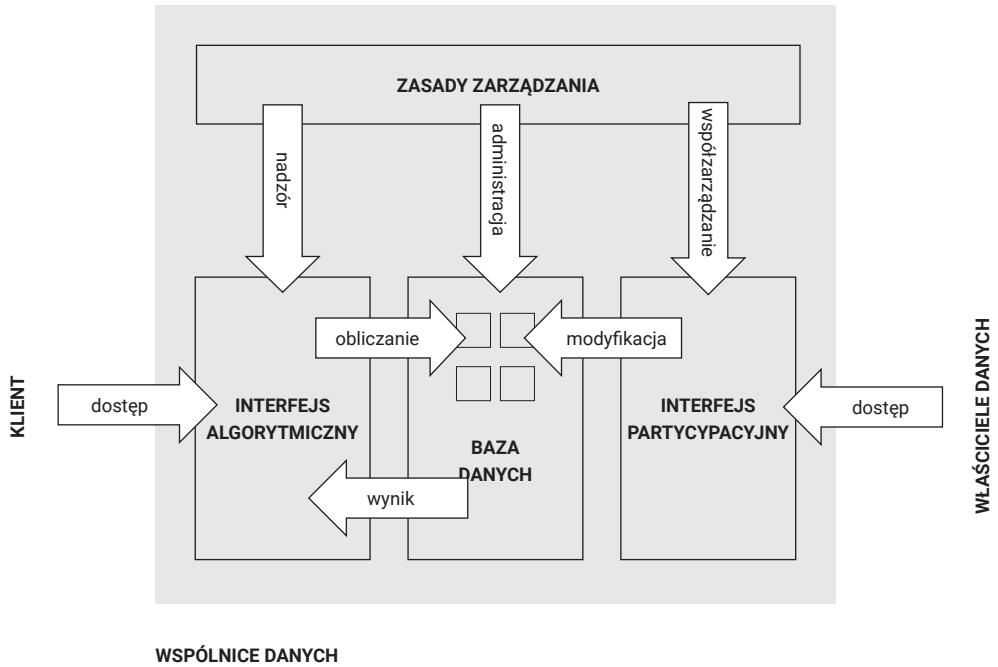
Biorąc pod uwagę możliwości techniczne, całkowicie zasadne jest przyjęcie koncepcji „otwartych algorytmów”, jaką proponuje projekt MIT OPAL – Open Algorithms (Hardjono, Pentland, 2019). Składa się ona z poniższych zasad:

- **przenoszenie algorytmu do danych** – zamiast kopiować dane do klienta, pobierz algorytm i przetwarzaj dane na urządzeniu krańcowym,
- **dane nigdy nie mogą opuścić repozytorium** – dla maksymalnego bezpieczeństwa,
- **weryfikacja algorytmów** – przykładowo przez przeprowadzenie oceny skutków algorytmu (OSA)
- **zwracanie tylko bezpiecznych wyników** – tylko zagregowane wyniki uniemożliwiają ponowną identyfikację danych jednostkowych.

Na podstawie zasad MIT OPAL należy odrzucić możliwość kopiowania danych przez serwer klienta. Proponuję zatem dwie wersje architektury – zcentralizowaną oraz rozproszoną/sfederowaną. Wersje nie są rozłączne, czyli wspólnica danych może jednocześnie przechowywać część pełnego zbioru danych na jednym lub wielu własnych serwerach oraz korzystać z bazy metadanych do lokalizowania pozostałych danych przechowywanych przez właścicieli danych.

Taka architektura znajduje zastosowanie, gdy przechowywanie danych na urządzeniach krańcowych jest utrudnione, lub samo stanowi ryzyko dla prywatności (np. dane medyczne pacjentów publicznego systemu ochrony zdrowia). Dane są trzymane na jednym lub wielu serwerach należących do wspólnicy, która przetwarza je (zarządza) zgodnie z ustalonymi uprzednio warunkami. Algorytm klienta (z zaszyfrowanymi wagami) po przesłaniu przez interfejs prowadzi obliczenia bezpośrednio na danych, a wyniki są zwracane klientowi.

Diagram 2. Zcentralizowana architektura wspólnicy

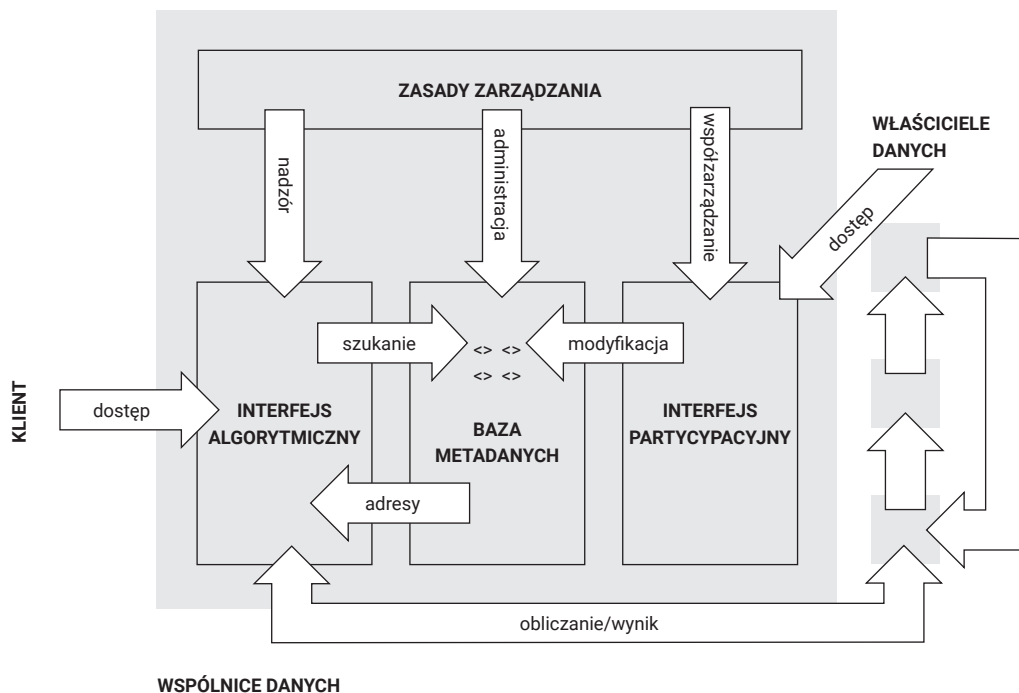


Rozproszona architektura znajduje zastosowanie, gdy dane są przechowywane przez strony chcące zachować pełną kontrolę nad danymi, np. dlatego, że same intensywnie je wykorzystują (dane z procesów przemysłowych). Federacja wymaga ustandaryzowania danych i sposobu ich przechowywania przez właścicieli. Metadane są kopiowane do wspólnej bazy danych. Algorytm klienta po przesłaniu przez interfejs wyszukuje w bazie metadanych odpowiednie zestawy, po czym za pośrednictwem protokołu współdzielenia przeprowadza federacyjne uczenie na serwerach właścicieli danych.

2.2 Właściwości architektury

Planując wdrożenie wybranej architektury, trzeba wziąć pod uwagę konieczne do jej poprawnego działania właściwości. O'Hara wska-

Diagram 3. Federacyjna (rozproszona) architektura wspólnicy



zuje pożądane właściwości w ośmiu głównych obszarach działania, od momentu odkrycia istnienia danych do badania skutków korzystania ze wspólnicy danych (O'Hara, 2019).

1. **Odkrywanie** – klienci mają możliwość odkrycia istnienia zbioru danych, jego właściwości i jakości.
2. **Pochodzenie** – klienci mają dostęp do metadanych umożliwiających ocenę jakości i potencjalnie również pochodzenie danych (szczególnie ważne w przypadku federacyjnej architektury).
3. **Kontrola dostępu** – wspólnica zachowuje możliwość kontrolowania, którzy klienci mogą używać danych. Robi to w imieniu właścicieli danych, którzy muszą mieć zapewnioną możliwość określenia zasad, na jakich wspólnica udziela tego dostępu (lub w niektórych przypadkach sprawować ostateczną kontrolę osobiście przez udzielanie zgody).

4. **Dostęp** – dostęp klienta i właściciela danych odbywa się przy pomocy standardowych protokołów i interfejsów umożliwiających przesył danych zgodnie z wybraną architekturą i sposobem ochrony prywatności (np. anonimizacja, pseudonimizacja).
5. **Zarządzanie tożsamością** – spółnica rozpoznaje tożsamość klientów i właścicieli danych korzystających z dostępu.
6. **Audytywanie użycia** – spółnica prowadzi transparentny rejestr dostępu i życia danych, aby umożliwić audyt i skontrolowanie, czy dane zostały użyte zgodnie z prawem i z zasadami przyjętymi przez spółnicę i właścicieli danych.
7. **Odpowiedzialność** – spółnica ponosi odpowiedzialność za nieprawne użycie danych².
8. **Wpływ** – właściciele danych i spółnica analizują wpływ używania danych, np. przez ocenę skutków trenowanych algorytmów (OSA), wartości tworzonej dla wszystkich stron i dla społeczeństwa czy konsekwencji przyjętego modelu biznesowego.

Właściwości po stronie klienta mogą mieć miejsce na portalu internetowym, który, jak proponuje O'Hara, może zawierać katalog metadanych, narzędzia do przeglądania, wizualizacji danych czy elementy tablicy ogłoszeniowej i budowania społeczności. Spółnica danych oczywiście musi prowadzić własny rejestr klientów i ich uprawnień oraz gromadzić logi korzystania z zasobów. Właściciele danych natomiast muszą dysponować interfejsem na wzór klientów banku, z możliwością przeglądania wniesionych do spółnicy zasobów, historii ich użycia, oraz narzędziami do ich modyfikacji, wycofania czy zmiany zasad udostępniania.

Właściciele danych powinni mieć wpływ na strategię zarządzania danymi – weryfikację niektórych klientów, ustalanie zasad oraz szeroko rozumiany model biznesowy i rozwój spółnicy.

**Warszawska
Wspólnica Danych**

Warszawska Wspólnica Danych zarządzana przez Bibliotekę Publiczną m.st. Warszawy agreguje dane zbierane przez spółki komunalne i jednostki samorządu (m.in. Miejskie Przedsiębiorstwo Wodociągów i Kanalizacji, Zarząd Transportu Miejskiego), czujniki prywatnych osób (m.in. smogu) oraz obowiązkowo przekazywane przez duże platformy cyfrowe operujące w mieście (m.in. Uber, Wolt, Airbnb, Booking). Tworzy to atrakcyjny, różnorodny zbiór danych. Średniej wielkości firma tworzy oprogramowanie dla inteligentnych latarni miejskich. Przez portal internetowy na stronie biblioteki sprawdza zakres danych, ich szczegółowość i kompletność. Wypełnia prosty formularz dostępu, akceptuje warunki licencyjne (firma może chronić oprogramowanie patentem maksymalnie 5 lat, ale miasto ma prawo wglądu na cele OSA) i opłaca skorzystanie z bazy. Wspólnica potwierdza autoryzowane użycie i je rejestruje. Firma korzysta z danych, aby stworzyć zaawansowany system sterujący oświetleniem, który może sprzedać wielu miastom. Na koniec roku rada złożona z prezydenta Warszawy, dyrektorów jednostek samorządowych dostarczających dane oraz organizacji obywatelskich ocenia, czy dostęp był prawidłowo udzielany. Decyzją rady mieszkańcom z zysków zostanie wypłacona dywidenda od danych, odliczona od rachunków za elektryczność.

2.3 Pozyskiwanie danych i model biznesowy

Wspólnice danych będą zarządzały cennym w gospodarce cyfrowym aktywem, niezbędnym do tworzenia konkurencyjnych produktów i świadczenia usług klientom. Wobec tego polityka „otwartych danych” będzie tutaj zgubna, ponieważ z otwartości danych najbardziej skorzystają największe podmioty, zaś właściciele danych nie dostaną żadnej rekompensaty za dane, które udostępniłi.

Jak już zauważyliśmy wcześniej, wiele niezwykle cennych dla społeczeństwa czy biznesu zestawów danych znajduje się w korporacyjnych silosach. Z punktu widzenia dużych firm technologicznych mo-

nopolizacja danych jest rozsądna biznesowo, przynosi korzyści – i to w skali, która przekracza uczciwy zysk, a stanowi nieproduktywną rentę. Nic dziwnego, że znalezienie idealnego modelu biznesowego, zaspokajającego wszystkie strony, wydaje się niezwykle trudne. Przecież prawo antymonopolowe nie szuka korzystnej umowy dla monopolisty i społeczeństwa płacącego rentę, czy dla monopolisty i nieuczciwie pokonywanych małych firm, a raczej sięga po środki przymusu, od kar wobec spółki po zwiększenie konkurencji przez rozbitcie podmiotu.

Dlatego konieczne jest, żeby wspólnice danych nie były jedynie oddolnym wysiłkiem np. organizacji pozarządowych czy mikroprzedsiębiorców. **Wykorzystując progresywne regulacje, samorządy i rządy mogą nakazać przeniesienie prywatnych danych korporacji z domeny prywatnej do publicznej.** Takie rozwiązanie jest dziś powszechnie postulowane w kręgach zarówno ekspertów prawa (co Viktor Mayer-Schönberger nazywa *progressive data sharing mandate*; Mayer-Schönberger & Ramge, 2018), jak i ekonomistów (Mazzucato, 2018). Pierwsze miasta i kraje już wdrażają reformy nakazujące dominującym podmiotom dzielenie się danymi, np. o transporcie w mieście, nawykach zakupowych czy zdrowiu.

Aby być samowystarczalna lub chociaż minimalizować straty, wspólnota danych powinna znaleźć adekwatny model biznesowy, czyli schemat pozyskiwania środków na swoje funkcjonowanie w zamian za dostarczaną wartość (Hardinges et al., 2019). Przykładowe modele finansowania dla wspólnoty to:

- pobieranie opłat za dostęp do danych przez API (różnicowane wg celu zastosowania lub liczby przetwarzanych danych),
- sprzedaż abonamentów obejmujących obsługę klienta, pomoc techniczną (*troubleshooting*),
- sprzedaż dodatkowych usług przez centrum kompetencji wspólnoty danych,
- organizowanie wydarzeń i konkursów wokół danych z udziałem partnerów komercyjnych,
- pozyskiwanie grantów od instytucji publicznych i organizacji pozarządowych.

Chcąc wybrać model biznesowy, warto czerpać lekcje z teorii zarządzania wspólnymi zasobami. Noblistka Elinor Ostrom wskazuje, że aby uniknąć tzw. tragedii wspólnego pastwiska, konieczne jest m.in. dopasowanie zasad przywłaszczania i dostarczania zasobów do lokalnych warunków, stosowanie progresywnych sankcji i prostych mechanizmów rozwiązywania konfliktów (Ostrom, 1990) Wspólnica danych w swoim modelu biznesowym powinna różnicować podmioty niekomercyjne, inicjatywy naukowe czy społeczne, od klientów mających znaczące dochody z działalności na globalnych rynkach i pozycję dominującą.

W przypadkach, gdy wspólnica będzie otrzymywała dane nieodpłatnie, np. na podstawie regulacji nakazujących podmiotom przekazywanie danych z racji na interes publiczny, można przewidywać inne formy rekompensaty. Taką rekompensatą może być wystawianie listów intencyjnych na udział w projektach badawczych, darmowe konsultacje techniczne czy preferencyjne opłaty za dostęp.

**Wspólnica
E-Handlowa**

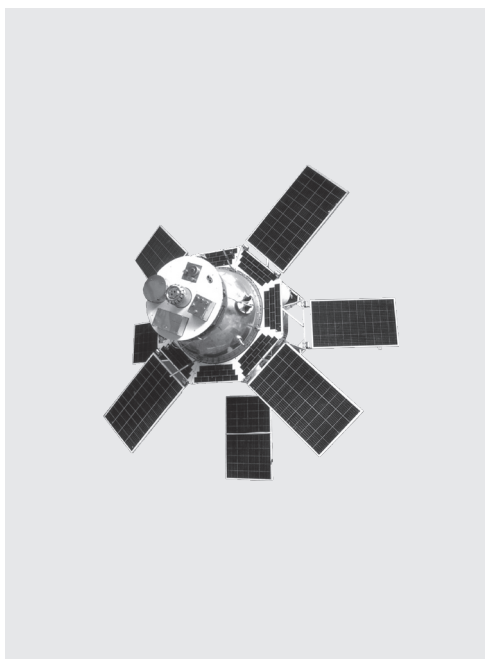
Wspólnica Handlowa zbiera dane zakupowe z kilkuset największych sklepów i sieci handlowych w Polsce, łącząc je z danymi platform e-handlu. Dzięki wdrożeniu Polskiej Karty Suwerenności Cyfrowej również firmy technologiczne takie jak Amazon, Alibaba i eBay przekazują swoje dane. Takie informacje to niezwykle ciekawy temat badawczy, ale też wiedza o trendach rynkowych. Zarząd wspólnicy na wniosek organizacji konsumenckiej decyduje, że uczelnie i organizacje pozarządowe będą mieć dostęp za darmo na licencji badawczej, natomiast firmy zapłacą za korzystanie z API. Oddający dane do wspólnicy zapłacą symboliczne kwoty, ale zewnętrzne podmioty dostaną stawki rynkowe. Dodatkowo, Wspólnica Handlowa oferuje odpłatne szkolenia z analizy danych handlowych i wizualizacji danych.

2.4 Forma prawna operatora

Ostatnim zagadnieniem przy tworzeniu wspólnicy danych jest wybranie adekwatnej formy prawnej. Jak ustaliliśmy na początku, podmioty stricte komercyjne znajdują się daleko od optimum społecznego z powodu inherentnego konfliktu interesów. Dwa ze wspomnianych w rozdziale 1. modeli – trustu danych i kooperatywy danych – nie są adekwatne dla różnych zastosowań wspólnic danych na gruncie polskiego czy szerzej europejskiego prawa. Instytucja trustu nie istnieje w prawie kontynentalnym (stanowionym; *civil law*).

Kooperatywy danych są natomiast dobrym i działającym już w praktyce modelem dla danych osobowych. Przykładem może tu być Midata, szwajcarska kooperatywa danych medycznych do celów badawczych (Burnicka, Zygmuntowski, 2019). Nie znajdują jednak zastosowania w przypadku danych posiadanych nie tylko przez osoby fizyczne (które mogą być spółdzielcami), ale również przez osoby prawne – firmy, organizacje rządowe i pozarządowe.

Jednym z możliwych wariantów jest potraktowanie wspólnicy danych podobnie jak klastra energii, który nie istnieje jako osobny podmiot prawny – jest po prostu wielostronną umową. Dotychczasowe doświadczenia z pilotaży współdzielenia danych wskazują, że stosowanie umowy oznacza konieczność wypracowania i ustalenia wszystkich zasad z góry, co utrudnia kalibrację np. modelu biznesowego (Hardinges et al., 2019). Co więcej, istnieją zasadne obawy o ustalanie odpowiedzialności za błędy w procesach administrowania, przetwarzania i ochrony danych, szczególnie w świetle RODO.



il.3

Warto podkreślić, że być może w gospodarce cyfrowej trzeba stworzyć zupełnie nowy rodzaj wehikułu prawnego. Obecnie istniejące osoby prawne to przecież nie tylko dość uniwersalne spółki i stowarzyszenia, lecz także samodzielne publiczne zakłady opieki zdrowotnej, koła gospodyń wiejskich i instytuty badawcze. „Ustawa o wspólnicach danych” byłaby ambitnym ujednoczeniem prawnych zasad, na których ma funkcjonować współdzielenie danych.

Korzystając z istniejących form prawnych, wskazuję poniższych dobrych kandydatów na operatora wspólnicy danych.

Repozytorium naukowe

Forma prawna:	uczelnia lub instytut badawczy
Charakterystyka:	repozytorium dorobku intelektualnego instytucji badawczej to neutralna przestrzeń opierająca się o rygor akademicki (Alsaad et al., 2019)
Przykłady:	dane zdrowotne i medyczne, dane edukacyjne, dane przemysłowe i branżowe, poufne dane handlowe

Repozytorium biblioteki

Forma prawna:	państwowa lub samorządowa instytucja kultury (biblioteka)
Charakterystyka:	cyfrowe archiwum biblioteki to dobry punkt wyjścia do agregacji danych związanych z samorządem lub obszarem społecznego życia rutt
Przykłady:	dane miejskie, dane internetu rzeczy (IoT), dane użyteczności publicznej (energetyczne, komunikacyjne)

Jednostka publiczna

Forma prawna:	instytucja gospodarki budżetowej
Charakterystyka:	IGB ma osobowość prawną i samodzielnie gospodaruje mieniem, ale jest elementem sektora finansów publicznych, tak jak np. Główny Urząd Statystyczny czy Centralny Ośrodek Informatyki
Przykłady:	dane demograficzne, dane telekomunikacyjne, dane użyteczności publicznej (energetyczne, sanitarne, komunikacyjne), dane skarbowe

2.5 Licencja dla danych a prywatność/poufność

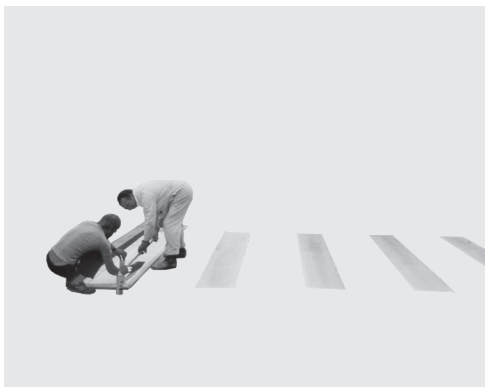
Ta sekcja odnosi się w największej mierze do danych osobowych i prywatności, która przysługuje ludziom. Jednak wysokie standardy bezpieczeństwa danych gwarantowane przez RODO mogą być też wskazówką, jak projektować wspólnice danych przetwarzające dane nieosobowe, które strony uważają za poufne (np. dotyczące tajemnic handlowych firmy).

Podstawową regulacją stojącą na straży prawa do prywatności w Unii Europejskiej jest Rozporządzenie o Ochronie Danych Osobowych (RODO). Artykuły RODO nie przypisują własności danych do żadnej ze stron, jednak nadają obywatelom szereg niezwykłych praw związanych z przetwarzaniem danych osobowych, takich jak m.in. prawo do sprostowania, ograniczenia przetwarzania, przenoszenia czy usunięcia danych (znane też jako „prawo do bycia zapomnianym”).

Istnieją pewne obawy, że RODO może ograniczać łatwość lub zakres współdzielenia danych, jednak dotychczasowe prognozy dotyczące np. znaczącego spadku PKB i utraty milionów miejsc pracy (Deloitte, 2013) nie zostały potwierdzone empirycznie. RODO nie stoi na przeszkodzie powtarzalnego przetwarzania danych ani udostępniania go stronom trzecim, konieczne jest jednak zadbanie, by wspólnice danych stosowały się do wymogów rozporządzenia.

Artykuł 25 RODO ma tutaj kluczowe znaczenie, ponieważ nakazuje uwzględnianie ochrony danych w fazie projektowania (*data protection by design*) oraz domyślną ochronę danych (*data protection by default*). Oznacza to konieczność zadbania o organizacyjne i techniczne mechanizmy, które spełniają wskazane w Artykule 5 RODO zasady dotyczące przetwarzania danych osobowych. Dla ułatwienia implementacji tych zasad w projektach współdzielenia danych zespół z University of Southampton opracował kwestionariusz ochrony danych (Stalla-Bourdillon et al., 2020), którego polska wersja wraz z omówieniem znajduje się w Aneksie 1. do tego raportu.

Te zasady muszą znaleźć odbicie zarówno w infrastrukturze technicznej (np. API), jak i w podstawie prawnej przetwarzania. Dobrą praktyką jest opracowanie standardowej licencji. Taka licencja stanowi dokumentację prawną, towarzyszącą dokumentacji technicznej interfejsów. Projektując licencję, warto wzorować się na dobrych przykładach zidentyfikowanych przez Support Center for Data Sharing dla Komisji Europejskiej:

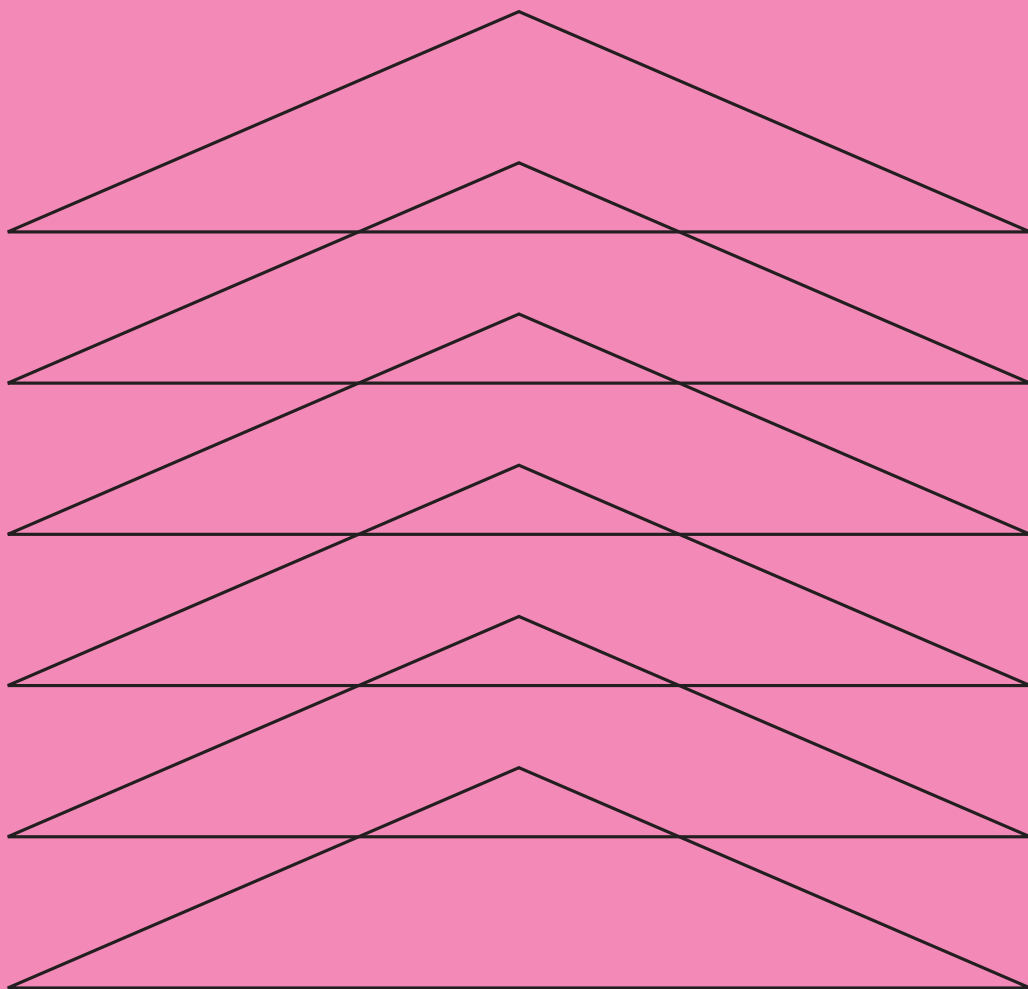


il. 4

- Swedish API License – elastyczna licencja prawna i techniczna opracowana przez szwedzką agencję innowacyjności VINNOVA dla twórców aplikacji przetwarzających dane osobowe;
- Data Use Agreement for Open AI Model Development (DUAOAI) – licencja przygotowana przez Microsoft do wykorzystywania danych osobowych oraz poufnych do trenowania modeli AI;
- Montreal Data Licence – elastyczna, ogólna licencja przygotowana przez zespół prawników i badaczy AI do trenowania modeli AI;
- Norwegian Licence for Open Government Data (NLOD) 2.0 – licencja służąca udostępnianiu nieosobowych danych publicznych.

Proces powstawania licencji musi włączać branżę, ekspertów, ale też społeczeństwo, tak, by te zasady były nie tylko zgodne z RODO, ale zabezpieczały wspomniany interes publiczny (Ctrl-Shift, 2018). Przedmiotem dyskusji może być wynagrodzenie za przetwarzanie danych dla ich właścicieli; licencja może też precyzować czy powstałe produkty technologiczne będą strzeżone patentami, czy umieszczone w otwartym dostępie.

3. Implementacja wspólnic danych w Polsce. Trzy propozycje



O

Opisana w raporcie wspólnota danych nie jest abstrakcyjnym, utopijnym niemal pomysłem, ale całkowicie możliwym do wdrożenia modelem zarządzania danymi. Polska może już teraz korzystać w pełni z gospodarki cyfrowej i stać się liderem społecznie korzystnej AI. W tym rozdziale opiszę krótko trzy propozycje wdrożeniowe w sektorach szczególnie potrzebujących modernizacji i cyfryzacji: opieki zdrowotnej, gospodarki komunalnej oraz energetyki.

3.1 Wspólnota Danych Zdrowotnych

Wspólnota Danych Zdrowotnych to największa w Polsce baza danych zdrowotnych i medycznych służących tworzeniu nowoczesnych rozwiązań telemedycznych, podnoszeniu jakości publicznego systemu opieki zdrowotnej i prowadzeniu przełomowych badań naukowych.



il. 5

Źródła danych:

- Centrum Systemów Informacyjnych Ochrony Zdrowia
- Narodowy Fundusz Zdrowia
- Państwowa Inspekcja Sanitarna
- Państwowa Inspekcja Farmaceutyczna
- Podmioty lecznicze (m.in. szpitale, praktyki lekarskie, fundacje)
- Prywatne sieci medyczne i ubezpieczenia
- Apteki
- Inteligentne urządzenia (m.in. Fitbit, Apple Watch)
- Aplikacje zdrowotne i medyczne

Architektura i model biznesowy:

- Federacyjna architektura dla zwiększonego bezpieczeństwa danych
- Pełen wgląd obywateli w osobiste zestawy danych i możliwości darowizny danych na cele badawcze
- Darmowy dostęp dla naukowców i organizacji pożytku publicznego
- Dostęp do ograniczonego wolumenu („piaskownicy”) dla start-upów
- Komercyjny dostęp oznacza 10% udziału wspólnicy w kontroli i zyskach z wytworzonych praw własności intelektualnej

Zastosowania:

- Profilaktyka dopasowana do lokalnych potrzeb i trendów okresowych
- Prognozowanie ognisk chorobowych na podstawie recept i zakupów leków
- Zdalna oraz zautomatyzowana diagnostyka
- Lepsze wykrywanie raka oraz chorób przewlekłych przez AI
- Inteligentny asystent zdrowotny dla obywateli
- Planowanie zapotrzebowania na usługi medyczne i lepsza alokacja funduszy

3.3 Wspólnica Danych Energetycznych

Wspólnica Danych Energetycznych była kluczowym narzędziem sukcesu transformacji energetycznej w Polsce. Dzięki niej nie tylko inwestorzy zaufali w plan dekarbonizacji kraju, ale możliwe stało się konsekwentne zwiększanie efektywności energetycznej przez precyzyjne zarządzanie energią.

Operator: Główny Urząd Statystyczny

Źródła danych:

- Polskie Sieci Elektroenergetyczne (operator sieci przesyłowej)
- Operatorzy sieci dystrybucyjnej (tzw. energetyczna wielka czwórka oraz Innogy)
- Elektrownie i elektrociepłownie
- Wirtualne elektrownie
- Instalacje prosumenckie

Architektura i model biznesowy:

- Architektura mieszana – scentralizowana archiwizacja starych zbiorów danych i federacyjny dostęp do nowych
- Dostęp do danych w modelu SaaS – abonament miesięczny o niskiej wartości
- Symboliczna „dywidenda od danych” dla małych i średnich firm oraz producentów energii

Zastosowania:

- Budowa ogólnopolskiej inteligentnej sieci (smart grid)
- Zoptymalizowane planowanie inwestycji w energetyce
- Przewidywanie zapotrzebowania na prąd i tworzenia bilansujących się źródeł
- Zwiększanie efektywności energetycznej dzięki wykrywaniu strat
- Personalizacja oferty dla gospodarstw domowych i firm
- Tworzenie oprogramowania dla energetyki odnawialnej

3.2 Wspólnica Danych Komunalnych

Wspólnica Danych Komunalnych powstała jako baza dla projektów inteligentnego miasta w Warszawie, jednak stopniowo włączyła wszystkie miasta wojewódzkie. Dziś jest swoistym „mózgiem miasta” dla polskich aglomeracji.

Operator: Biblioteka Publiczna m.st. Warszawy

Źródła danych:

- Miejskie spółki użyteczności publicznej (m.in. wodociągi, transport, mieszkalnictwo, odpady)
- Biura i departamenty miejskie (np. organizacji ruchu, planowania przestrzennego)
- Czujniki prywatnych osób (m.in. smogu)
- Prywatne platformy cyfrowe (m.in. Uber, Wolt, Airbnb, Booking, Nextbike)
- Deweloperzy i inwestorzy nieruchomości
- Aplikacje i platformy miejskie

Architektura i model biznesowy:

- Scentralizowana struktura obniżająca koszty obsługi technicznej
- Nadzór składający się z przedstawicieli miast i organizacji miejskich
- Darmowy dostęp w przypadku tworzenia rozwiązań na zamówienie uczestniczących miast
- Komercyjny dostęp w modelu płatności za użycie, uzależniony od skonsolidowanych przychodów grupy kapitałowej
- Środki zasilają międzymiastowy budżet partycypacyjny

Zastosowania:

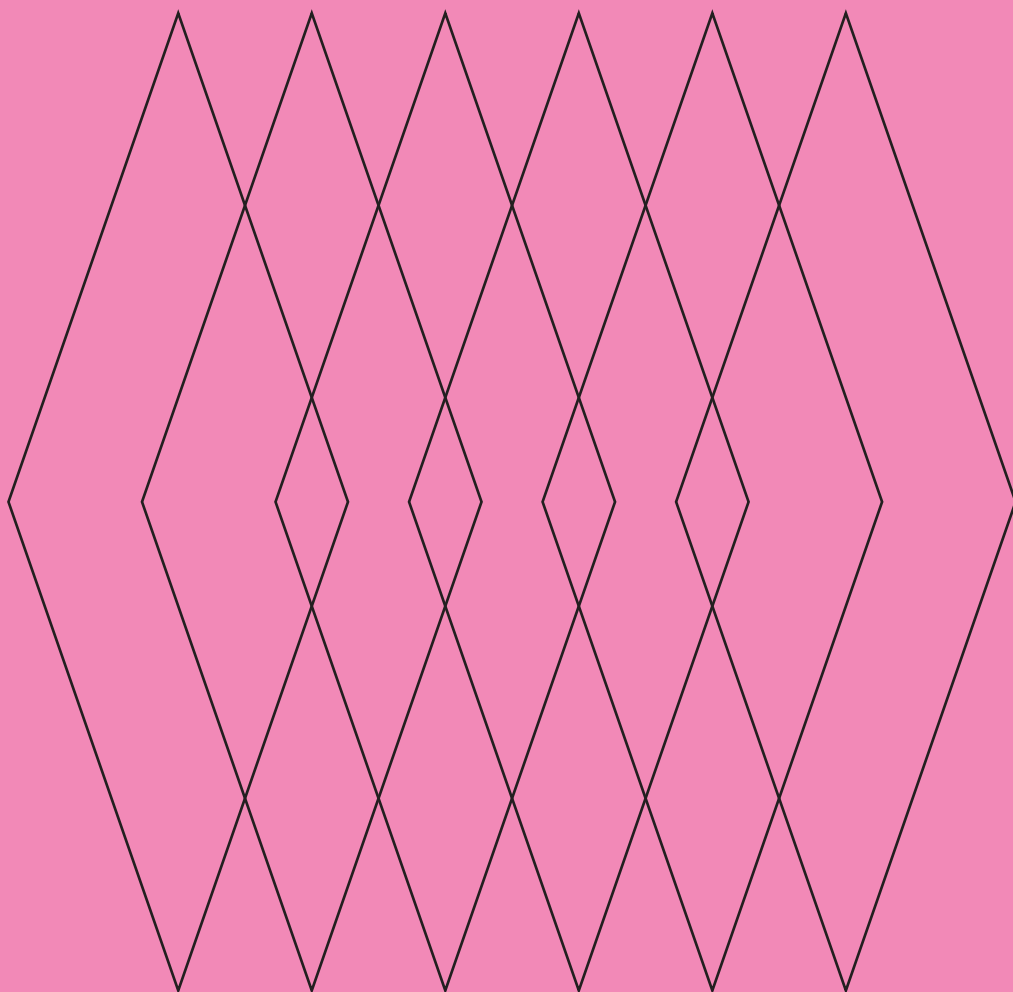
- Inteligentne sterowanie oświetleniem ulicznym
- Zintegrowana wiedza nt. rynku nieruchomości i potrzeb mieszkaniowych

- Optymalizacja multimodalnego transportu miejskiego dzięki AI
- Ochrona środowiska bazująca na wiedzy
- Dopasowanie oferty usług miejskich do spersonalizowanych potrzeb mieszkańców
- Prognozowanie zużycia infrastruktury i koniecznych inwestycji
- Tworzenie nowych firm, produktów i usług w oparciu o znane nawyki mieszkańców



il. 6

4. Publiczne gospodarowanie danymi. Ryzyka i sugestie regulacyjne



W

Wspólnice danych to ambitna propozycja budowy instytucji na miarę gospodarki cyfrowej w XXI wieku. Jednak nie znaczy to, że sukces jest gwarantowany. Implementacja zupełnie nowego modelu działania w obszarze nowych technologii może zwyczajnie okazać się zbyt skomplikowana, szczególnie jeśli wspólnice zostaną uznane za sezonowe pilotaże lub ciekawe projekty dla mikrograntów, a nie fundamentalny sposób reorganizacji życia społeczno-gospodarczego.

Dlatego powodzenie alternatywnego modelu zarządzania danymi zależy od właściwych podstaw i priorytetów. Dotyczy to nie tylko samych decydentów politycznych, lecz także organizacji pozarządowych, szczególnie zajmujących się prywatnością danych czy szerzej gospodar-

ką cyfrową, badaczy oraz innowatorów i przedsiębiorców technologicznych. Publiczne gospodarowanie danymi to logika przenoszenia danych do sfery publicznej, aby umożliwić wspólne wykorzystywanie zasobów w bezpieczny i egalitarny sposób. Wymierne korzyści odniosą wszyscy, choć co jasne koszty zostaną rozłożone sprawiedliwie, by od dominujących rynkowo podmiotów wymagać więcej niż od start-upów czy niedofinansowanych przez lata instytucji publicznych.

Publiczne gospodarowanie danymi to odpowiedź na realne zagrożenia, które stoją przed projektem wdrożenia wspólnic danych:

- **brak woli politycznej i dynamiki wdrożenia zmian** – wobec czego konieczne jest wskazanie konkretnego „gospodarza” danych, który przyjmie rolę koordynatora powstającej współpracy,
- **przejęcie lub wykorzystanie wspólnicy przez państwo do inwigilacji** – sensem „publicznego” gospodarowania jest nie wyłącznie administracyjna, państwowa kontrola, ale prawne wpisanie

interesariuszy w procedury zarządzania, dzięki czemu nie będą zależeć od arbitralnych decyzji o np. ujawnieniu kodu czy pojedynczych dokumentów,

- **wyciek wartości przez zasady nieproporcjonalnie korzystne dla prywatnych firm** – od początku trzeba zaznaczyć nadrzędny cel wspólnicy danych, którym jest dbanie o interes publiczny, a to oznacza ustawienie priorytetów zrównoważonego dostępu i odzyskiwania wartości na pierwszym miejscu.

Aby ułatwić wdrożenie wspólnic danych, można posłużyć się regulacjami prawnymi, by umożliwić sprawniejsze działanie i szybciej doprowadzić do uwspólnienia niektórych zasobów danych. Wśród sugestii regulacyjnych przede wszystkim warto wymienić:

- uchwalenie *Ustawy o wspólnicach danych* definiującej nową osobę prawną wzorowaną na prawie spółdzielczym z systemu kontynentalnego i prawie powiernictwa z tradycji precedensowej,
- znowelizowanie *Ustawy o ochronie danych osobowych* tak, aby wzmocnić prawo do przenoszenia danych, co ułatwi ich transfer do wspólnic danych,
- nałożenie na firmy prywatne o określonej znaczącej obecności cyfrowej ustawowego obowiązku przekazywania danych do wspólnic danych, szczególnie w branżach interesu publicznego, jakimi są zdrowie i medycyna,
- jednoznaczne poparcie na arenie unijnej dla powstawania wspólnych przestrzeni danych.

Kwestionariusz ochrony danych

Kwestionariusz ochrony danych (Stalla-Bourdillon et al., 2020)

- I. Zdefiniuj **cel** współdzielenia danych (Art. 5.1.b – „ograniczenie celu”).
- II. Zidentyfikuj **prawną podstawę** współdzielenia danych (Art. 5.1.b – „ograniczenie celu”).
- III. Określ, które dane są **niezbędne** dla tego celu. Upewnij się, że **minimalizujesz**: a) liczbę **zbędnych aktywności** związanych z przetwarzaniem danych; b) niezbędny **zakres danych**. Jeśli możesz zanonimizować dane, zrób to (Art. 5.1.c – „minimalizacja danych”).
- IV. Określ **czas przechowywania danych** związany ze wskazanym celem (Art. 5.1.e – „ograniczenie przechowywania”).
- V. Upewnij się, że współdzielone dane są **prawidłowe** (Art. 5.1.d – „prawidłowość”).
- VI. Potwierdź **uczciwość** procesu przetwarzania danych (Art. 5.1.a – „zgodność z prawem, rzetelność i przejrzystość”).
- VII. Upewnij się, że dane **nie są zmieniane** lub **udostępniane bez zezwolenia** – przykładowo: zdefiniuj, kto może mieć dostęp do danych – oraz że przetwarzanie ma **poufny** charakter (Art. 5.1.f – „integralność i poufność”).
- VIII. Upewnij się, że przetwarzanie danych jest **monitorowane w transparentny** sposób, np. przez rejestrowanie prowadzonej na danych aktywności w celu wykazania przestrzegania zasad. Oceń ryzyka, zanim rozpoczniesz przetwarzanie (Art. 5.1.a – „zgodność z prawem, rzetelność i przejrzystość” oraz Art. 5.2 – „rozliczalność”).

Określenie powyższych zasad i stosowanie się do nich jest wystarczające, by przetwarzać dane osobowe zgodnie z RODO. Szczegółowe odpowiedzi na pytania zawarte w kwestionariuszu są konieczne do ustalenia w przypadku każdorazowego tworzenia spółnicy i jej modyfikacji. Warto jednak podkreślić podstawową różnicę w prze-

chowywaniu i przetwarzaniu danych między wersjami architektury:

- zcentralizowana architektura oznacza, że zarówno administratorem i podmiotem przetwarzającym pełne dane jest wspólnota danych,
- federacyjna architektura oznacza, że administratorem i podmiotem przetwarzającym metadane jest wspólnota; podmiotem przetwarzającym pełne dane jest sam właściciel danych.

Jan J. Zygmuntowski Ekspert projektu SpołTech. Realizuje projekt: banki danych (data trust) jako alternatywne modele zarządzania danymi. Ekonomista zajmujący się innowacją, cyfrową gospodarką i zrównoważonym rozwojem. Prezes zarządu think-tanku Instrat. Doradza i wspiera w rozwoju kilku projektów technologicznych z pozytywnym wpływem społecznym. Stypendysta G20 Global Solutions i student Szkoły Doktorskiej Akademii Leona Koźmińskiego.

Bibliografia

- Acemoglu, D., Makhdoui, A., Malekian, A., & Ozdaglar, A. (2019). *Too Much Data: Prices and Inefficiencies in Data Markets*. <https://doi.org/10.3386/w26296>
- Alsaad, A., O'Hara, K., & Carr, L. (2019). Institutional repositories as a data trust infrastructure. *WebSci 2019 - Companion of the 11th ACM Conference on Web Science*, 1–4. <https://doi.org/10.1145/3328413.3329402>
- Bass, T. (2020). *It's time to think about our data as a common good | British Council*. British Council. <https://www.britishcouncil.org/anyone-anywhere/explore/communities-connections/rethinking-data>
- Booch, G., Rumbaugh, J., & Jacobson, I. (2005). *The unified modeling language user guide*. Addison-Wesley.
- Burnicka, A., & Zygmontowski, J. J. (2019). *#CoopTech: Platformowy kooperatyzm jako silnik solidarnego rozwoju*. www.instrat.pl
- Ctrl-Shift. (2018). *Data Mobility: The personal data portability growth opportunity for the UK economy*.
- Delacroix, S., & Lawrence, N. (2018). Disturbing the 'One Size Fits All', Feudal Approach to Data Governance: Bottom-Up Data Trusts. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3265315>
- European Commission. (2018). *Towards a common European data space*. <https://ec.europa.eu/digital-single-market/en/news/public-consultation-building-european-data-economy>
- European Commission. (2020). *A European strategy for data*.
- Evans, B. J. (2011). Much Ado About Data Ownership. *Harvard Journal of Law & Technology*, 25(1). <http://www.hhs.gov/ohrp/policy/cdebiol.html>
- Godel, M., & Natraj, A. (2019). *Independent assessment of the Open Data Institute's work on data trusts and on the concept of data trusts*. www.londoneconomics.co.uk
- Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Beaufays, F., Augenstein, S., Eichner, H., Kiddon, C., & Ramage, D. (2018). *Federated Learning for Mobile Keyboard Prediction*. <http://arxiv.org/abs/1811.03604>
- Hardinges, J., Wells, P., Blandford, A., Tennison, J., & Scott, A. (2019). *Data trusts: lessons from three pilots*. <https://docs.google.com/document/d/118RqyUAWP3WlyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit>
- Hardjono, T., & Pentland, A. (2019). *Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management*. <http://arxiv.org/abs/1905.08819>
- Jemielniak, D., & Przegalinska, A. (2020). *Collaborative Society*. MIT Press Essential Knowledge series.
- Kerry, C. F., & Morris Jr, J. B. (2019, June 26). *Why data ownership is the wrong approach to protecting privacy*. Brookings Center for Technology Innovation. <https://www.brookings.edu/blog/techtank/2019/06/26/why-data-ownership-is-the-wrong-approach-to-protecting-privacy/>
- Maler, E., Machulak, M., & Richer, J. (2017, January). *User-Managed Access (UMA) 2.0*. Kantara Initiative, Kantara Published Specification. <https://docs.kantarainitiative.org/uma/ed/uma-core-2.0-10.html>
- Mayer-Schönberger, V., & Ramge, T. (2018). *Reinventing capitalism in the age of big data*.
- Mazzucato, M. (2018, July 27). Let's make private data into a public good. *MIT Technology Review*. <https://www.technologyreview.com/2018/06/27/141776/lets-make-private-data-into-a-public-good/>
- Mulgan, G., & Straub, V. (2019, February 21). *The new ecosystem of trust: how data trusts, collaboratives and coops can help govern data for the maximum public benefit*. Nesta. <https://www.nesta.org.uk/blog/new-ecosystem-trust/>
- O'Hara, K. (2019). *Data Trusts Ethics, Architecture and Governance for Trustworthy Data Stewardship*. <https://www.southampton.ac.uk/wsi/index.page>

- Ostrom, E. (1990). *Governing the commons: the evolution of institutions for collective action*. Cambridge University Press.
- Ruttan, C., Chakarova, R., Apollonova, N., Gill, P., & Kelcey, B. (2019). *BiblioTech: Beyond Quayside: A City-Building Proposal for the Toronto Public Library to Establish a Civic Data Hub*.
- Sacha, K. (2010). *Inżynieria oprogramowania*. Wydawnictwo Naukowe PWN.
- Stalla-Bourdillon, S., Thuermer, G., Walker, J., Carmichael, L., & Simperl, E. (2020). Data protection by design: Building the foundations of trustworthy data sharing. *Data & Policy*, 2. <https://doi.org/10.1017/dap.2020.1>
- The Economist. (2017). Regulating the internet giants - The world's most valuable resource is no longer oil, but data | Leaders | The Economist. *The Economist*. <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>
- WIPO Technology Trends 2019: Artificial Intelligence. (2019).
- Zygmuntowski, J. J., Chojecka, K., & Roy, S. A. (2020). *Podatek cyfrowy od gigantów. Ekspertyza w zakresie wprowadzenia w Polsce podatku cyfrowego (DST)*.

Projekt SpołTech - bo przyszłość potrzebuje przyjaciół

Technologie cyfrowe wpędzają nasze społeczeństwo w stan ciągłej zmiany. Dlatego przyszłość potrzebuje przyjaciół, którzy upomną się o społeczny wymiar technologii.

SpołTech to przestrzeń do namysłu i rozmowy o technologii, przyszłości i Polsce. W Fundacji Centrum Cyfrowe stworzyliśmy zespół analityczny badający społeczny i kulturowy wpływ technologii cyfrowych.

Jednocześnie siecujemy ekspertki i ekspertów, dla których ważny jest społeczny wymiar technologii. Osoby wierzące, że technologie powinny służyć społeczeństwu. Kierownikiem merytorycznym projektu SpołTech jest dr Alek Tarkowski, prezes Fundacji Centrum Cyfrowe.

Projekt SpołTech to wizja świadomego i zrównoważonego rozwoju technologii oparta na następujących założeniach:

- Każda zmiana technologiczna jest zmianą społeczną.
- W Polsce większość dyskusji koncentruje się na rozwoju gospodarki lub infrastruktury cyfrowej. My chcemy mówić o rozwoju społeczeństwa cyfrowego.
- Zbyt wiele debat skupia się na określonej technologii. Dla nas najważniejsze są wartości i cele, którym technologie mogą służyć.
- Technologie muszą być rozwijane w sposób zrównoważony i skupiony na potrzebach społeczeństwa.



Strona projektu: www.centrumcyfrowe.pl/spoltech

Biuletyn SpołTech:
<https://centrumcyfrowe.pl/spoltech-biuletyn/>

Naszym celem jest wspieranie otwartości i zaangażowania w świecie nowoczesnych technologii. Działamy w modelu *think-and-do tank*, wspierając otwartość i zaangażowanie w świecie nowoczesnych technologii. Pracujemy na rzecz zmiany społecznej i zwiększenia zaangażowania obywatelskiego, wykorzystując narzędzia cyfrowe i modele współpracy oparte na dzieleniu się zasobami i wiedzą.

Naszym zdaniem zdrowe społeczeństwo cyfrowe musi uwzględniać społeczny wymiar technologii – to założenie, które przyświeca nam od początku działalności. Nie możemy skupiać się jedynie na rozwijaniu infrastruktury, gospodarce cyfrowej lub rozwoju usług i produktów. Zamiast o technologiach chcemy rozmawiać o SpołTechu.



Spis Ilustracji

il. na okładce

Grafika: Alicja Kobza Studio, CC BY 4.0

na podstawie zdjęć:

- First Wisconsin Center, Milwaukee, Wisconsin

źródło: <https://bit.ly/2CDd0NY>

licencja CC BY-NC 2.0

- Wheelie

źródło: <https://bit.ly/2Outz5>

licencja CC BY-NC 2.0

- First Wisconsin Center, Milwaukee, Wisconsin

źródło: <https://bit.ly/32oolHx>

licencja CC BY-NC 2.0

il. 1 → str. 7

Grafika: Alicja Kobza Studio, CC BY 4.0

na podstawie zdjęć:

- Man in dust coat replacing armful of books onto shelves

źródło: <https://bit.ly/3fAV1BI>

licencja CC BY 2.0

il. 2 → str. 20

Grafika: Alicja Kobza Studio, CC BY 4.0

na podstawie zdjęć:

- Okänd rymdutställning

źródło: <https://bit.ly/3h5u1uol>licencja

licencja CC BY 2.0

- Echo Satellite

źródło: <https://bit.ly/391YWoD>

public domain

il. 3 → str. 28

Grafika: Alicja Kobza Studio, CC BY 4.0

na podstawie zdjęć:

- *Space Pavilion*, Moskwa,

źródło: <http://bitly.pl/w6JGP>

licencja CC BY-SA 2.0

il. 4 → str. 31

Grafika: Alicja Kobza Studio, CC BY 4.0

na podstawie zdjęć:

- Yeaman Bridge

źródło: <https://bit.ly/2DGX48Y>,licencja

licencja CC BY-SA 2.0

il. 5 → str. 33

Grafika: Alicja Kobza Studio, CC BY 4.0

na podstawie zdjęć:

- Masks worn during experiments with plague

źródło: <https://bit.ly/30ilgVR>

public domain

il. 6 → str. 37

Grafika: Alicja Kobza Studio, CC BY 4.0

na podstawie zdjęć:

- Man working at analog computer, 1968

źródło: <https://bit.ly/32ogdXy>

licencja CC BY

Redakcja i korekta: Marta Wojas
Skład: Alicja Kobza
Grafika: Alicja Kobza

Fundacja Centrum Cyfrowe
Projekt SpołTech

Warszawa 2020

Ten utwór jest udostępniony na licencji Creative Commons Uznanie Autorstwa 4.0 Międzynarodowe. Zachęcamy do jego przedruku i wykorzystania. Prosimy jednak o zachowanie informacji o finansowaniu raportu oraz podanie linku do strony <https://centrumcyfrowe.pl/spoltech/>



PROO

Sfinansowano przez Narodowy Instytut Wolności – Centrum Rozwoju Społeczeństwa Obywatelskiego ze środków Programu Rozwoju Organizacji Obywatelskich na lata 2018–2030



NIW