

Biuletyn Informacji Publicznej Polskiej Akademii Nauk

Adres artykułu: <https://bip.pan.pl/artukul/56-758-aby-walka-z-pandemia-byla-efektywna-potrzebny-jest-szerszy-dostep-do-danych-stanowisko-nr-21-z-16-wrzesnia-2021-r>

Aby walka z pandemią była efektywna, potrzebny jest szerszy dostęp do danych (stanowisko nr 21 z 16 września 2021 r.)

Podejmowanie strategicznych decyzji bez danych i analiz może nie tylko okazać się nietrafne, ale i prowadzić do tragicznych skutków. W szczególności dotyczy to sytuacji kryzysowych – na przykład obecnej pandemii COVID-19. Z danych powinni korzystać nie tylko decydenci, ale i naukowcy – im więcej danych dobrej jakości, tym większa szansa na zrozumienie badanych zjawisk. W sytuacji kryzysu epidemicznego, który pokonać można jedynie dzięki racjonalnym zachowaniom podejmowanym w skali całego społeczeństwa, kluczowego znaczenia nabiera również zapewnienie szerszego dostępu do danych dla dziennikarzy i obywateli. W tym stanowisku analizujemy, jak w kontekście pandemii COVID-19 umożliwić pełniejsze wykorzystanie dostępnych danych.

Dane w pandemii

Dane o dużym stopniu zagregowania dotyczące zakażeń, hospitalizacji i zgonów, a także interwencji takich jak testowanie czy szczepienia, są potrzebne do śledzenia rozwoju epidemii w skali ponadnarodowej oraz skuteczności różnych strategii jej zapobiegania. Służą temu **globalne repozytoria danych**, takie jak [Worldometer](#), [Our World in Data](#) czy [COVID-19 Data Repository by the Center for Systems Science and Engineering \(CSSE\) at Johns Hopkins University](#) wykorzystujące między innymi nowoczesne technologie zautomatyzowanego pozyskiwania publikowanych w sieci danych. Repozytoria te są ogólnodostępne, ale mało szczegółowe.

Lokalne decyzje podejmowane w kraju czy w regionie wymagają danych o większej szczegółowości, na przykład zawierających informacje:

- w jakich grupach wiekowych czy zawodowych występują zakażenia,

- czy występują lokalne ogniska,
- jakie grupy są narażone na ciężki przebieg choroby,
- jaki jest status zaszczepienia wśród osób, które chorują.

Zbierane są one w ramach **systemów nadzoru epidemicznego**. Dodatkowe dane generują również **systemy wspomagające procesy administracyjne**, takie jak system wspomagający objęcie izolacją czy kwarantanną lub zlecenie testu. Łącznie z innymi danymi administracyjnymi (np. o statusie zatrudnienia, stanie cywilnym, posiadaniu dzieci itd.) stanowią one wyjątkowo bogaty zasób. W Polsce publicznie udostępniane są jednak jedynie podstawowe statystyki dotyczące zachorowań na COVID-19. W dodatku przez dłuższy czas były one udostępniane w formatach utrudniających ich pobieranie i dalsze wykorzystanie.

Osobną kategorią danych są te **powstające w wyniku użycia nowych technologii cyfrowych**. Są to pomiary mobilności oparte o:

- dane z telefonów komórkowych,
- aplikacje na telefon umożliwiające śledzenie kontaktów lub przestrzegania kwarantanny,
- czy też aplikacje, w których można dokumentować swoje objawy i np. zamówić test w kierunku SARS-CoV-2.

Część tych danych znajduje się w sektorze prywatnym, ale w pewnym zakresie zostały one publicznie udostępnione, jak np. COVID-19 Mobility Reports. Natomiast dane zbierane przez aplikacje publiczne nie są praktycznie w ogóle udostępniane społeczeństwu.

W trakcie pandemii przeznaczono dodatkowe finansowanie na badania naukowe służące poznaniu zarówno samego wirusa, patofizjologii choroby, dróg szerzenia się zakażeń, jak i procesów społecznych i szeroko rozumianych konsekwencji epidemii. Niektóre z danych płynących z tych badań są udostępniane innym naukowcom. Powstały także inicjatywy **repozytoriów danych pozyskanych w ramach projektów badawczych** finansowanych z pieniędzy publicznych, ale są one na razie dość nieliczne i ograniczone do wąskich tematów lub dyscyplin.

Dobrym przykładem są także globalne przedsięwzięcia badawcze takie jak [Rapid-Response COVID-19 Project \(PSACR\)](#). Działania wchodzące w skład tego projektu mają na celu przeprowadzenie rygorystycznych międzynarodowych badań w celu zrozumienia psychologicznych i behawioralnych aspektów kryzysu COVID-19. Zaletą takich działań jest duża skala zbieranych danych, co nie tylko zwiększa rzetelność uzyskanych wyników, ale daje znakomite możliwości porównań międzykulturowych.

W Polsce wielu badaczy realizuje prace na temat różnych aspektów COVID-19, ale tematyka tych prac jest rozdrobniona, a same prace są prowadzone na małą skalę i w izolacji od innych badaczy. Brak koordynacji, współpracy i nawyku dzielenia się pomysłami i danymi utrudnia wykorzystanie istniejącego w Polsce potencjału badawczego, a także znacząco obniża wagę i rangę uzyskanych wyników.

Reasumując: w trakcie pandemii wiele danych zbieranych jest na bieżąco. Są to dane epidemiczne, administracyjne, pochodzące z projektów badawczych czy od użytkowników aplikacji i serwisów. Są one wykorzystywane przez decydentów, ale nie są one wykorzystywane w pełni. Łączenie zasobów administracyjnych umożliwiłoby na przykład badanie zachorowalności na COVID-19 w wybranych grupach zawodowych, ciężkości przebiegu choroby wśród pacjentów z chorobami współistniejącymi, czy porównania częstości hospitalizacji wśród osób zaszczepionych i niezaszczepionych przeciwko COVID-19. Połączenie danych epidemicznych z danymi psychologicznymi lub społecznymi umożliwiłoby także lepsze zrozumienie wpływu czynników pozamedycznych na rozwój i przebieg choroby.

Otwarcie baz danych dla badaczek i badaczy stanowiłoby unikalną okazję wykorzystania zainteresowania środowisk naukowych epidemią i w efekcie ich pogłębionych analiz być może lepsze, oparte na dowodach decyzje administracyjne. Dostęp do danych pozwoliłby również na weryfikację – a co a tym idzie – większą wiarygodność racjonalnych decyzji rządowych dotyczących walki z epidemią.

Dane dotyczące zdrowia to **dane wrażliwe, dlatego przy udostępnianiu należy zadbać o ich pełną anonimowość** i brak możliwości identyfikacji poszczególnych osób. Należy też zwrócić uwagę, że osoba może nie być identyfikowalna w pierwotnym zbiorze danych, ale już po złączeniu z dodatkowymi informacjami może tak się stać. Im więcej informacji obejmuje zbiór danych, tym większe ryzyko identyfikacji osoby. Stąd udostępnianie danych indywidualnych musi być zawsze rozważane pod tym kątem i podlegać szczególnym regulacjom.

Kultura ponownego używania danych

Podczas pandemii utworzono wiele repozytoriów danych. Warto wspomnieć o danych zbieranych, agregowanych i publikowanych przez [Europejskie Centrum ds. Prewencji i Kontroli Chorób \(ECDC\)](#). ECDC jest tu dobrym przykładem, gdyż wiele z ich danych jest udostępnianych bezpłatnie i są dostępne do użycia dla dowolnych celów. Jednak dostęp do szczegółowych danych indywidualnych, które mogą być danymi wrażliwymi, jest udzielany na podstawie konkretnego wniosku badaczy dokładnie określającego zakres danych i cele badawcze. [Procedura ta](#) z jednej strony zapewnia

transparentność procesu zbierania i udostępniania danych, z drugiej możliwość podejmowania prac badawczych w oparciu o dane z obszaru całej Europy.

Transparentny i efektywny proces udzielania dostępu do danych, zwłaszcza danych administracyjnych z publicznych wykazów (rejestrów), wymaga uwzględnienia tego aspektu już przy konstrukcji baz danych. Udostępnianie danych wymaga również wyznaczenia instytucji odpowiedzialnej za udzielanie tego dostępu. W przypadku złożonych baz danych i konieczności integrowania danych z różnych źródeł przygotowanie zbioru danych do badań może wymagać niebanalnych operacji na bazach, a co za tym idzie zespołu fachowców znających strukturę zasobów, których zadaniem jest przygotowywanie danych do dalszych analiz. Obecnie rejestry dotyczące COVID-19 znajdują się w kilku instytucjach ([Centrum e-Zdrowia](#), [Główny Inspektorat Sanitarny](#), [Narodowy Instytut Kardiologii – Państwowy Instytut Badawczy](#), [Narodowy Instytut Zdrowia Publicznego – PZH – Państwowy Instytut Badawczy](#)), a dane są wymieniane pomiędzy rejestrami. Zasady ewentualnego udostępniania danych do badań nie zostały jednak określone, w tym nie została określona jedna instytucja odpowiedzialna za ten proces.

Dane wtórne pochodzące ze źródeł administracyjnych z reguły mają wiele ograniczeń, które powinny być brane pod uwagę przy ich analizie. Znajomość procesu pozyskiwania danych jest istotna zarówno w przypadku danych rejestrowych, jak i danych pozyskiwanych w trakcie zaplanowanych badań i eksperymentów. Analiza danych wtórnych wymaga więc szczegółowej wiedzy na temat procesu zbierania danych, a w przypadku bardziej złożonych zbiorów – współpracy z instytucjami odpowiedzialnymi za ich zbieranie. **Dobłą praktyką może być publikowanie danych wraz z ich opisem jako osobnej publikacji** (tzw. data paper) skupiającej się na samych danych bardziej niż na wnioskach z nich płynących. Publikacja taka spełniałaby również rolę docenienia samego procesu pozyskiwania danych. Znajomość tego procesu, jak i weryfikacja spójności danych powinna być jednym z zadań instytucji odpowiedzialnej za proces udostępniania danych.

Zalecenia

Stworzenie kultury szerokiego udostępniania danych może budować zaufanie do decyzji administracji, to z kolei jest kluczowe w radzeniu sobie z epidemią, dlatego zalecamy:

- **Publiczne udostępnianie danych na najwyższym możliwym poziomie szczegółowości, nieodpłatnie i bez konieczności rejestracji.** Takie dane

powinny być dostępne dla mediów, przedsiębiorców i ogółu społeczeństwa. Wymaga to utworzenia i utrzymywania trwałej platformy, umożliwiającej wizualizację danych, ale także pobranie aktualnych danych w formie pozwalającej na ich dalszą analizę. Pociąga to za sobą konieczność jasnego ustalenia akceptowalnego i zgodnego z prawem ochrony danych osobowych i wrażliwych poziomu szczegółowości udostępnianych danych.

- **Zdecydowanie większe otwarcie zbiorów danych administracyjnych i badawczych dla celów wtórnych analiz dotyczących COVID-19.** Większe otwarcie zbiorów danych wymaga utworzenia infrastruktury umożliwiającej bezpiecznie udostępnianie zasobów, w tym transparentnych zasad udostępniania danych i, co bardzo ważne, wyznaczenia instytucji odpowiedzialnej za ten proces. Zasady udostępniania danych powinny zostać opracowane we współdziale środowisk naukowych oraz specjalistów ds. ochrony danych osobowych.
- **Utworzenie wyspecjalizowanej i niezależnej jednostki prowadzącej repozytorium danych badawczych,** w szczególności pochodzących z populacyjnych badań społecznych odnoszących się do postaw i zachowań podczas pandemii. Jednostka ta mogłaby również koordynować proces pozyskiwania takich danych, tak by umożliwić niezależną ocenę trendów
- **Udostępnianie wyników badań.** Publikacje naukowe, choć cenne, pojawiają się dopiero po pewnym czasie. W okresie pandemii istotny jest czas i komunikowanie kluczowych wyników jak najszybciej. [W naszym stanowisku nr 18](#) poświęconym komunikacji publicznej w czasie pandemii wskazywaliśmy na decydującą rolę niezależnych instytucji i zespołów eksperckich. Takie zespoły mogłyby również stanowić forum dyskusji nad nieopublikowanymi jeszcze wynikami badań.
- **Udział w międzynarodowych inicjatywach dotyczących zasobów danych dostępnych zarówno dla naukowców, jak i dla przedsiębiorców.** Otwieranie zasobów informacyjnych jest postrzegane jako długofalowy kierunek rozwoju i jest częścią [Europejskiej Strategii dla Danych](#). Strategia ta wprowadza regułę otwartego, darmowego używania i dystrybucji zbiorów danych z systemów i badań prowadzonych ze środków publicznych oraz wskazuje na konieczność określenia uczciwych i jasnych zasad dostępu do danych. Konieczna jest też inwestycja w infrastrukturę (także tę ogólnoeuropejską) oraz zapewnienie

odpowiednich uprawnień, narzędzi i umiejętności instytucjom generującym dane. Zgodnie z tą strategią Komisja Europejska we współpracy ze środowiskami naukowymi podjęła inicjatywę utworzenia [Europejskiej Chmury dla Otwartej Nauki](#), w której ze strony polskiej aktywny udział bierze Narodowe Centrum Nauki. W perspektywie kilku lat Europejska Chmura będzie wirtualnym środowiskiem z łatwo dostępnymi usługami przechowywania, zarządzania, analizy i ponownego wykorzystywania danych badawczych, współdzielonych pomiędzy różnymi dyscyplinami naukowymi i państwami członkowskimi UE. Niezbędna jest jednak intensyfikacja prac w tym zakresie.

[Zespół doradczy ds. COVID-19 przy prezesie Polskiej Akademii Nauk](#)

Tagi: polska akademia nauk, pan, koronawirus, covid19, zespół doradczy, stanowisko

Metryczka

Wytworzył:	Zespół doradczy ds. COVID-19
Data wytworzenia:	16.09.2021
Opublikował w BIP:	Agnieszka Borowa-Malinowska
Data opublikowania:	16.09.2021 14:25
Ostatnio zaktualizował:	Agnieszka Borowa-Malinowska
Data ostatniej aktualizacji:	16.09.2021 14:26
Liczba wyświetleń:	146