

Projekt „Wirus na mapie PL” – geolokalizowanie przypadków zakażenia w okresie pandemii na przykładzie zachorowań na COVID-19 w Polsce

Lokalizacja kluczem do zdrowia

Wybuch epidemii koronawirusa spowodował ogromne zapotrzebowanie na szczegółowe dane o rozmieszczeniu zachorowań w kraju. Skoro nie zaspokoili go państwowe służby, sprawy w swoje ręce wzięli studenci geodezji i kartografii z Uniwersytetu Warmińsko-Mazurskiego w Olsztynie.

Efektem tych działań jest projekt „Wirus na mapie PL” zainicjowany w połowie marca przez MKN GeoSiN z Olsztyna. Później dołączyły do niego studenckie koła z: AGH w Krakowie (KNG Dahlta), Uniwersytetu Przyrodniczego we Wrocławiu (SKN Geodetów) oraz Uniwersytetu Warszawskiego (KN Geografii Regionalnej i Politycznej). Przedsięwzięcie jest odpowiedzią na potrzebę zgromadzenia w jednym miejscu szczegółowych danych o zakażeniach i zgonach oraz o osobach zdrowych. W wyniku podjętych działań w krótkim czasie powstała ogólnodostępna, aktualna i łatwa do implementacji baza danych, która następnie otworzyła drogę do wykonania i publikacji różnorodnych opracowań statystycznych i kartograficznych dostępnych na stronie www.wirusnamapie.pl. Produkty te zwiększają świadomość społeczną na temat pandemii i potencjalnego ryzyka zakażenia, a tym samym pomagają prze-

ciwdziać rozprzestrzenianiu się wirusa. Niezależnie bowiem od tego, jakimi narzędziami GIS dysponujemy, to serce systemu zawsze stanowią dane. Proces ich zbierania nie jest jednak prosty i wymaga od nas bycia na bieżąco z doniesieniami na temat nowych przypadków zachorowań, zgonów lub ozdowień. Monitorowanie komunikatów rządu i stacji sanitarno-epidemiologicznych z całego kraju, a także doniesień medialnych oraz profili społecznościowych urzędów i niektórych decydentów pozwoliło nam na dokonanie oceny rzetelności i spójności informacji docierających do opinii publicznej.

• Gdzie jest chory?

Na początku marca, w pierwszych dniach raportowania o pandemii, dochodziło do dość dużych rozbieżności pomiędzy danymi podawanymi przez komórki rządowe a doniesieniami medialnymi (przy czym to te drugie z reguły były bar-

ziej precyzyjne). Ministerstwo Zdrowia wskazywało na miasta, w których potwierdzano nowe przypadki. Nie były to jednak miejsca zamieszkania pacjentów, ale miejscowości, w których znajdują się szpitale zakaźne, gdzie doszło do potwierdzenia infekcji. Między innymi z tego powodu początkowo mówiono o 5 pacjentach z Ostródy, podczas gdy pierwszy przypadek w powiecie ostródzkim został *de facto* zaraportowany dopiero 23 marca (był to jednocześnie 25. zakażony w województwie warmińsko-mazurskim).

Działania stacji sanitarno-epidemiologicznych początkowo były z kolei niespójne w skali kraju. Osoby z objawami zakażenia nie zawsze wybierały szpitale zakaźne w swoim województwie – często jechały tam, gdzie miały najbliżej. W konsekwencji stacje sanepidu zaliczały na poczet wykrytych przypadków zakażenia osoby zamieszkujące inne województwo. Jednocześnie musiały jednak informować o tym stację właściwą dla miejsca zamieszkania pacjenta. Czasami zdarzało się więc tak, że jeden przypadek raportowany był przez dwie różne stacje, a w komunikatach urzędów wojewódzkich i Ministerstwa Zdrowia pojawiały się dwie różne liczby zakażonych. Przypadki takie występowały sporadycznie i niekiedy były prostowane, są jednak dowodem na to, że proces raportowania i wymiany danych powinien zostać dopracowany.

Fragment bazy danych o zachorowaniach

	A	B	C	D	E	F	G	H	I
1	Timestamp	Province	City	Postal_code	Infection/Death/Recovery	Sex	Where_infected	Who_infected	Age
2	08-03-2020	Dolnośląskie	Wrocław		I	M	GBR		26
3	08-03-2020	Dolnośląskie	Wrocław		I	M	ITA		73
4	09-03-2020	Dolnośląskie	Wrocław		I	F		Family	74
5	11-03-2020	Dolnośląskie	Wrocław		I	F	POL	Family	
6	12-03-2020	Dolnośląskie	Wrocław		I	M	GBR		
7	12-03-2020	Dolnośląskie	Wrocław		I	M	FRA		
8	12-03-2020	Dolnośląskie	Oleśnica		I	M	POL		
9	13-03-2020	Dolnośląskie	Wrocław		I	M			
10	14-03-2020	Dolnośląskie	Wrocław		I				
11	14-03-2020	Dolnośląskie	Wrocław		I				
12	14-03-2020	Dolnośląskie	Bolesławiec		I				
13	14-03-2020	Dolnośląskie	Wrocław		I	M			
14	14-03-2020	Dolnośląskie	Bolesławiec		I	M			
15	15-03-2020	Dolnośląskie	Wrocław		I	F		Other patient	
16	15-03-2020	Dolnośląskie	Legnica		I	F			
17	16-03-2020	Dolnośląskie	Wrocław		I				
18	16-03-2020	Dolnośląskie	Wrocław		I				
19	16-03-2020	Dolnośląskie	Wrocław		I				
20	16-03-2020	Dolnośląskie	Wrocław		I	F	POL	Family	1
21	16-03-2020	Dolnośląskie	Wrocław		I				
22	16-03-2020	Dolnośląskie	Oleśnica		I				
23	16-03-2020	Dolnośląskie	Legnica		I				
24	16-03-2020	Dolnośląskie	Wałbrzych		I	M		Friends	57

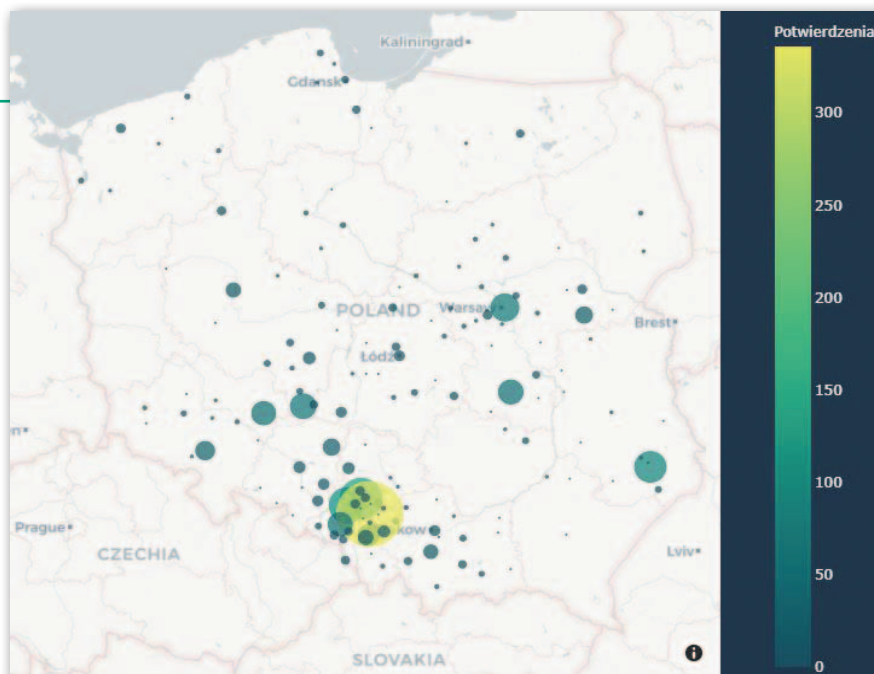
Zachorowania zaprezentowane przy użyciu kartodiagramu. Dzięki „suwakowi czasu” możemy zawęzić zakres danych do dowolnego przedziału (w tym przypadku 10-17 maja)

• Prawo do prywatności czy informacji?

W pierwszych dniach pandemii wszystkie potwierdzone przypadki spotykały się z dużym zainteresowaniem mediów i znane były szczegółowe informacje o każdym zakażonym – płeć, pochodzenie, wiek, potencjalne źródło infekcji i ewentualnie kraj, który w ostatnim czasie odwiedził. Wraz ze wzrostem liczby zakażonych krajowe media podawały coraz mniej danych o pacjentach. Bardziej szczegółowe informacje można było znaleźć w mediach lokalnych, ale nie było to regułą.

Obecnie w wielu powiatach ograniczono się do informowania jedynie o nowych przypadkach zakażeń i wyzdrowień, bez podawania informacji o osobach, których dotyczyły. Wyjątkiem są województwa małopolskie i podlaskie, gdzie znana jest także płeć. Sporadycznie podawane są ponadto informacje o wieku.

Ministerstwo Zdrowia od samego początku unikało publikacji precyzyjnych danych dotyczących pochodzenia osób zakażonych koronawirusem. Najpierw informacje o potwierdzonych przypadkach podawane były z dokładnością do miast, w których zdiagnozowano infekcje. Po około 14 dniach zdecydowano jednak, że dalsze informacje będą upubliczniane z dokładnością jedynie do województw, co uzasadniono zbyt dużą liczbą nowych przypadków. Mniej więcej w podobnym czasie stacje epidemiologiczne zaczęły publikować raporty na temat kolejnych zakażeń na terenie poszczególnych powiatów, więc na swój sposób szczegółowość danych podawanych przez instytucje państwowe została utrzymana. Dodajmy, że dotyczy to tylko informacji dla opinii publicznej, bo wójtowie otrzymywali komunikaty z podziałem na gminy, choć bez dodatkowych szczegółów. Przykładowo, 10 kwietnia wójt gminy Teresin poinformował mieszkańców, że nie otrzymuje oficjalnych danych o miejscu pobytu osób zakażonych, ponieważ „informacja ta traktowana jest jako informacja tajna, zawierająca dane wrażliwe”. Co ciekawe, w tym samym komunikacie podał nazwy miejscowości, z których pochodzą osoby z potwierdzonym zakażeniem. Nie był to odosobniony przypadek. Również starosta wysokomazowiecki publikował na swoim prywatnym profilu facebookowym bardziej szczegóło-



we informacje o miejscu zamieszkania osób chorych na COVID-19. Wiele wskazuje jednak na to, że uzyskał je nieoficjalnymi kanałami.

W tym miejscu warto zadać pytanie, czy w sytuacji pandemii ustawa o ochronie danych osobowych powinna wiązać ręce stacjom sanitarno-epidemiologicznym w zakresie udostępniania szczegółowych danych o miejscu przebywania osób zakażonych? A jeśli nie, to czy na pewno sanepid powinien podawać tak szczegółowe dane? Jak trafnie zauważa wspomniany wójt Teresina, „mieszkańcy muszą wiedzieć, w których miejscach powinni zachować szczególną ostrożność i absolutnie stosować wszelkie zalecane rygory bezpieczeństwa”. Świadomość, że w naszym otoczeniu przebywają osoby zakażone, znacznie bardziej działa przeciw niż informacja o liczbie chorych w powiecie czy województwie. Uwolnienie takich danych otworzyłoby również furtkę do bardziej złożonych analiz przestrzennych obrazujących kierunek i tempo rozwoju pandemii. Przykład dobrych praktyk płynie ze Stanów Zjednoczonych, gdzie studenci w ramach konkursu zorganizowanego przez Center for Global Data Visualization opracowali mapy obrazujące związki pomiędzy zachorowaniami na COVID-19 a nierównościami społecznymi, jakie dotyczą mniejszości rasowe (projekt „Mapping Racial and Ethnic Differences with COVID-19”). Jeśli chcemy, by równie szczegółowe opracowania pojawiały się w Polsce, musimy zapewnić dostęp do lepszych danych.

Na przeciwległej szali leży prawo obywateli do poszanowania informacji o stanie ich zdrowia i ochrona ich dóbr osobistych. Szczególnie w małych miejscowościach osoby zakażone mogą być przecież narażone na ostracyzm ze strony lokalnej społeczności.

• Płynność lokalizacji i pojęć

Epidemie rządzą się swoimi prawami i nie zawsze można przewidzieć wszystkie scenariusze. Czasami dochodziło do sytuacji niestandardowych, gdy trudno było nam właściwie przypisać pacjentów do konkretnej lokalizacji. Na przykład 14 marca na przejściu granicznym w Bezedach zatrzymano autokar rejsowy z Kaliningradu. Po zmierzeniu temperatury pasażerów dwoje z nich (obywatele UE) zabrano do szpitala w Ostródzie, a reszcie odmówiono wjazdu do Polski. Trudno zatem tej dwójce przypisać konkretne województwo, bo ani nie pochodzili z Polski, ani prawdopodobnie nie zostali tu zarażeni. W komunikacie rządowym pojawiła się jednak Ostróda i województwo warmińsko-mazurskie.

Kolejna nieścisłość związana jest z przypadkami wyzdrowień i/lub ozdrowień. Definicje osoby zdrowej i ozdrowiałej (tj. osoby, u której ustąpiły objawy) różnią się od siebie i nie powinny być stosowane zamiennie. Stacje epidemiologiczne informowały jednak o osobach ozdrowiałych, bez informacji o zdrowych. Niektóre używały tych terminów zamiennie i nie do końca było jasne, czy osoby ozdrowiałe faktycznie wyzdrowiały, czy też może należy je zaliczyć tylko w poczet ozdrowieńców. Z kolei w rządowych komunikatach wszystkie te osoby zaliczano do osób zdrowych, co nie było do końca zgodne ze stanem faktycznym. Z powodu braku jednolitej polityki dotyczącej klasyfikacji tych osób monitorowanie wyzdrowień (a tym samym przypadków aktywnych) staje się bardzo trudne.

Istniały też różne podejścia do informowania o innych cechach osób zakażonych. Dla przykładu: niektóre ze źródeł podawały wiek chorych. Częściej jednak zdarzały się komunikaty, które zawierały tylko kategorię wiekową pacjenta (osoba



Kierunki ostatnich podróży pacjentów z potwierdzonym zakażeniem w całym zakresie czasowym

młoda, starsza, w sile wieku itp.). W sytuacji, gdy w danym miejscu zakażonych było kilku, komunikaty ograniczały się do przedziału wieku, w którym znajdowała się cała grupa (prawdopodobnie wiek najstarszego i najmłodszego). Podobna sytuacja miała miejsce też w przypadku płci osób niepełnoletnich, gdy zamiast „kobieta/mężczyzna” pojawiała się określenie „dziecko”.

● W kierunku automatyzacji

Obecnie komunikaty publikowane przez powiatowe stacje sanitarno-epidemiologiczne są na tyle systematyczne i spójne, że proces automatyzacji pozyskiwania geodanych wydaje się kolejnym naturalnym krokiem w naszym projekcie. Wiąże się z nim jednak dwa wyzwania, z którymi musimy się zmierzyć. Pierwsze to uwzględnianie korekt do wcześniejszych komunikatów związanych z ograniczoną skutecznością testów. Co jakiś czas zdarzają się przypadki, że osoby, które uzyskały pozytywny wynik na obecność koronawirusa, po drugim badaniu próbki okazują się zdrowe. Wówczas nie jest poprawiany komunikat z błędem, ale publikuje się nowy, korygujący i jednocześnie aktualizowana jest liczba wszystkich zakażonych. Taki przypadek miał miejsce chociażby 30 kwietnia, gdy Ministerstwo Zdrowia poinformowało o 63 błędnych (fałszywie

pozytywnych) wynikach w województwie śląskim z okresu 16-19 kwietnia. Oprócz tego wcześniej zdarzały się korekty spowodowane dublowaniem pacjentów w komunikatach różnych PSSE lub zmiany przyczyn śmierci (choćby przypadek 27-letniej pacjentki ze szpitala w Łańcucie). Zdarza się, że korekty do komunikatów podawane są w sposób niejednoznaczny – przykładem może być zgon z 12 maja. Ministerstwo Zdrowia poinformowało o zdublowanym zgonie, ale bez podania konkretnych informacji – wieku, płci, miejscowości. W takiej sytuacji nie jest do końca jasne, który zgon należy usunąć z bazy.

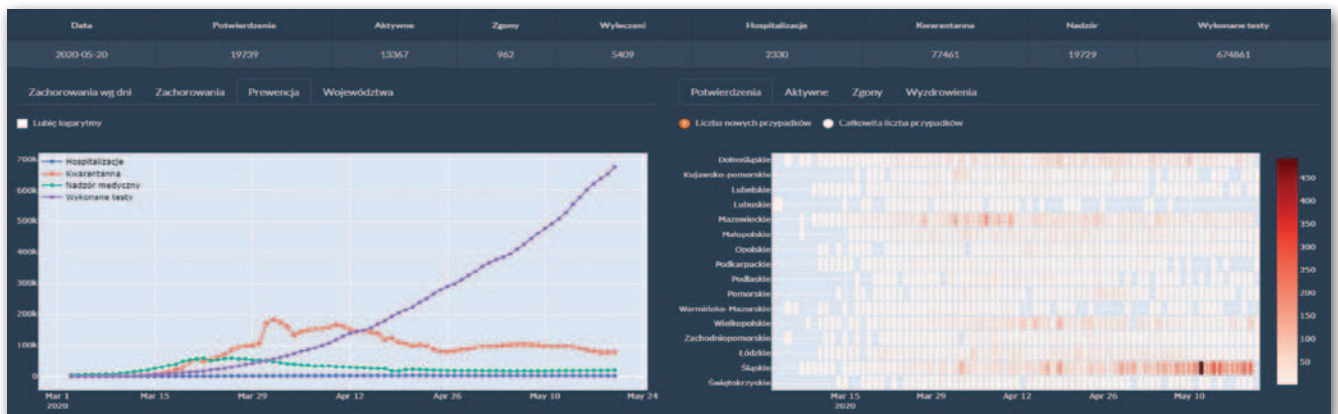
Drugie wyzwanie wynika z tego, że nie istnieje jeden szablon urzędowego komunikatu o osobach zakażonych. Takie informacje, w zależności od źródła, pojawiają się w formie tabelarycznej, tekstowej lub graficznej. Zautomatyzowanie procesu ich pozyskiwania wymaga przygotowania narzędzi, które zadziałają niezależnie od tego, kto i w jakiej formie przygotowuje komunikat (muszą one np. odczytać tekst z obrazka). Wiąże się z tym także rezygnacja z większej dokładności danych o zakażeniach, do których można dotrzeć, monitorując doniesienia medialne. Automatyczne pozyskiwanie i przetwarzanie informacji z takich źródeł jest praktycznie niewykonalne, bo mają one niejednorodną formę, bywają również pub-

likowane niekonsekwentnie lub są powielane przez różne media.

● Nie tylko obrazek

Na koniec warto zastanowić się nad rolą GIS w zarządzaniu pandemią w Polsce. Mogłoby się wydawać, że sytuacja tego rodzaju to idealny scenariusz dla wszelkiej maści analiz i prognoz przestrzennych. Mamy tu bowiem zarówno lokalizację nowych przypadków zakażeń, jak i oznaczanie obszarów, gdzie zakażeni mogli zarażać inne osoby. Tymczasem odnosimy wrażenie, że GIS został zepchnięty na drugi plan, do roli jedynie efektywnej wizualizacji danych. Trudno dostrzec, aby kolejne decyzje władz centralnych podejmowane były na podstawie analizy danych przestrzennych. Zarówno nowe obostrzenia, jak i proces ich luzowania odbywa się bowiem w skali całego kraju, a nie poszczególnych regionów. Dopiero w ostatnich dniach zaczęto mówić np. o tym, że w niektórych województwach obowiązek noszenia maseczek zniknie szybciej niż w innych. Ma to sens, ponieważ dane przestrzenne pokazują, że z jakichś powodów epidemia w jednych miejscach postępuje gwałtowniej niż w innych. Gdyby decyzje podejmowane były na podstawie aktualnego stanu pandemii w poszczególnych regionach, być może udałoby się złagodzić konsekwencje społeczno-ekonomiczne niektórych rozporządzeń wprowadzających obostrzenia związane z epidemią. Bardziej zrozumiałe byłby wpływ podejmowanych działań na faktyczny rozwój pandemii na danym obszarze, co w przyszłości pozwoliłoby opracować najlepszy scenariusz postępowania w podobnych sytuacjach. To oczywiście są jedynie nasze domniemania, bo – niestety – nie wiadomo, czym w podejmowaniu decyzji kieruje się Ministerstwo Zdrowia.

Marta Augustynowicz, Patrycja Borsuk, Mateusz Czyniak, Aleksandra Gleba, Tomasz Kozakiewicz, Michał Lasia, Łukasz Łobko, Adam Parol, Piotr Poskrobko, Dariusz Tanajewski



„Wirus na mapie” to także spora dawka statystyk m.in. w formie wykresów. Po lewej: liczba osób objętych jakąś formą prewencji, po prawej: liczba nowych przypadków wg województw