

ROZDZIAŁ 20.

Tradycyjne i nowe źródła danych w geografii społeczno-ekonomicznej i gospodarce przestrzennej – refleksje o dostępie i przydatności opisu procesów przestrzennych

Dariusz Ilnicki
Krzysztof Janc

Streszczenie: Celem opracowania jest zwrócenie uwagi na źródła danych, pozwalające w części wypełnić deficyty poznawcze i zaangażowanie badaczy w rozpoznaniu przestrzennych zjawisk społeczno-gospodarczych w skali lokalnej. Praca ma charakter przeglądowy oraz zawiera interpretacje oraz refleksje nad stanem, możliwością wykorzystywania przywoływanych źródeł danych. Zwrócono szczególną uwagę na te dane i bazy danych, które wykorzystywane są przy analizie ruchliwości przestrzennej. Odwołano się zarówno, do statystyki publicznej, baz danych edukacyjnych, transakcji bankomatowych, jak i danych związanych z rozwojem technologii cyfrowych – informacje o przemieszczeniach użytkowników telefonów komórkowych. Poza opisem tych źródeł, wskazano ich użyteczność oraz trudności związane z dostępem oraz korzystaniem z nich.

Wprowadzenie

Geografia społeczno–ekonomiczna i gospodarka przestrzenna zmienia się wraz z upływem czasu. Zmiany te są konsekwencją przekształceń instytucjonalnych, a może przede wszystkim pojawianiem się nowych zjawisk, kształtowania nowych struktur przestrzennych, nowych metod i źródeł wykorzystywanych w opisywaniu przestrzeni, otaczającej nas rzeczywistości. Dyscyplina rozwija się z jednej strony z „bagażem” twierdzeń, teorii, aparatu pojęciowego, metod, które są już uznane, a z drugiej zaś strony podlega presji wynikającej z konieczności sprostania nowym wyzwaniom. T. Strykiewicz [2011] pisząc o roli geografii, a w domyśle całej dyscypliny, w badaniu procesów i struktur społeczno–ekonomicznych w skali globalnej i regionalnej wskazuje na jej silne i słabe strony. Jakkolwiek odnosi je do zamkniętego katalogu zjawisk takich jak: transformacja strukturalna gospodarki, globalizacja, metropolizacja i integracja, a swoją dyskusję prowadzi w kontekście zmieniającego się pojmowania procesu globalizacji, to można uogólnienia sformułowane przez wspomnianego autora odnieść do całej geografii społeczno–ekonomicznej i gospodarki przestrzennej. Wskazuje on na przyczyny renesansu zainteresowania geografiami, przestrzeniami. Jedną z nich jest znajomość i posługiwanie się kartograficzną metodą badań, dla której szczególne możliwości stwarzają systemy informacji geograficznej (GIS) [zob. Churski, Zwoliński, 2011]. Jednak ukazanie przestrzennego zróżnicowania zjawisk, a następnie identyfikacja uwarunkowań oraz mechanizmów do nich prowadzących, bez względu na wykorzystywane narzędzia, wymaga odpowiednich (merytorycznych, rzetelnych, aktualnych, dostępnych) danych.

Dane są tym, na czym badacze opierają swoje dociekania naukowe. Brak danych, lub ich nieadekwatność do możliwości opisu w kontekście zachodzących zjawisk nie pozwala na obiektywne przedstawienie rzeczywistości, jej analizę, formułowanie wniosków, tez, budowę modeli, identyfikację prawidłowości w otaczającym nas świecie. Na problem ten należy również spojrzeć z perspektywy ujęć dynamicznych, czy też opisujących relacje, przepływy pomiędzy miejscami. Dynamika zmian społeczno–gospodarczych i fakt pojawiania się nowych zjawisk oraz procesów wymaga ich analizowania w ujęciach czasowych, od momentu ich pojawienia się. Jednak zmiany w świecie rzeczywistym są szybsze niż opis i publikowanie opracowań na ich temat. Można stwierdzić, że w chwili opublikowania wyniki mają walor „historyczny”. Często dla nowych zjawisk, mamy do czynienia z problemem braku dostępu do danych je opisujących.

W przypadku geografii społeczno–ekonomicznej i gospodarki przestrzennej, tak jak w większości dyscyplin, szczególnie istotne są i będą zmiany spowodowane przez *big data*, sztuczną inteligencję (SI), automatyzację wielu procesów, „oddanie” istotnych decyzji algorytmom. Pojawienie się oraz upowszechnienie tych technologii wpływa na powstawanie nowych problemów badawczych i nowych sposobów rozwiązywania starych. Cyfryzacja „wszystkiego” zmienia przedmiot badań, sposób

prowadzenia badań, jak i (pożądane) kompetencje badaczy. Cyfryzacja to „zmiany” zachowań przestrzennych, zmiany podmiotów kształtujących przestrzeń. To również zmiana w zakresie, wolumenie dostępnych danych o świecie, jak i sposobach ich pozyskiwania, przechowywania i eksploracji.

Odnosząc się do zarysowanej powyżej problematyki celem opracowania jest zwrócenie uwagi na źródła danych, pozwalające chociaż w części wypełnić deficyty poznawcze i zaangażowanie badaczy w rozpoznaniu przestrzennych aspektów zjawisk społeczno-gospodarczych w skali lokalnej oraz aplikacyjny ich charakter [zob. Kostrzewski i in., 2011; Strykiewicz, 2011]. Cel ten realizowany jest poprzez odwołanie się do zagadnienia ruchliwości i zachowań przestrzennych, czyli w konsekwencji tych źródeł danych, które umożliwiają ich opisanie. Opracowanie dotyczy zasadniczo kontekstu polskiego, jednak również z odwołaniami do literatury zagranicznej. Ma ono częściowo charakter przeglądowy, gdyż znajdują się w nim odwołania do przykładów zastosowań poszczególnych źródeł danych, jak i *de facto* samego przeglądu tych źródeł. W dużej części opracowanie zawiera interpretacje oraz refleksje nad stanem, możliwością wykorzystywania przywoływanych źródeł danych.

„Tradycyjne” dane – instytucje publiczne

Do tradycyjnych źródeł służących analizie ruchliwości przestrzennej należy zaliczyć, te które od kilkudziesięciu lat są wykorzystywane do opisu niektórych aspektów funkcjonowania mieszkańców Polski. Są to bazy Głównego Urzędu Statystycznego (GUS), będące efektem, konsekwencją programu badań statystycznych, czy stanowiące kompilacje innych, różnych źródeł; dane Systemu Informacji Oświatowej SIO II, czy dane dotyczące szkolnictwa wyższego. Są to dane gromadzone przez instytucje publiczne.

Po 1988 r. zagadnienia dotyczące przemieszczeń i ruchliwości przestrzennej ludności straciły na znaczeniu w polskich badaniach naukowych i społecznym dyskursie. Jednak w rzeczywistości społeczno-gospodarczej wciąż są istotne, pomimo braku zainteresowania badaniami. GUS zaprzestał publikacji danych dotyczących ruchliwości przestrzennej Polaków, co utrudniło prowadzenie badań, gdyż brak dostępu do danych uniemożliwia rzetelne badania. Problemатyczne jest również utrzymanie tajemnicy statystycznej w przypadku migracji wewnętrznych na pobyt stały, gdzie dane dla „jednostkowych przepływów” nie są publikowane. Dodatkowo, niepełność danych dotyczy dojazdów do pracy, zwłaszcza dla przepływów poniżej 10 osób.

Statystyka dojazdów do pracy jest zjawiskiem, które w polskiej statystyce ulega odtwarzaniu, a właściwie powstawaniu na nowo. Do chwili transformacji systemu społeczno-gospodarczo-politycznego przełomu lat 80. i 90. XX w. dojazdy do pracy były integralnym elementem sprawozdawczości i płaszczyzną badań obecną

w Narodowych Spisach Powszechnych (NSP). Ostatnie badanie z tego zakresu, które można określić mianem pełnego, miało miejsce w Narodowym Spisie Powszechnym z 1988 r. Natomiast kolejne spisy powszechne (2002, 2011) nie uwzględniały w badaniu istotnych zagadnień dla okresu przekształceń społeczno-gospodarczych, w tym i powiązań funkcjonalnych, rozpatrywanych między innymi, przez pryzmat dojazdów do pracy⁵¹. W pierwszej dekadzie XXI w. opublikowano dane o dojazdach do pracy dla: 2006, 2011, 2016 oraz 2021 r.⁵². Statystyki te mają swoją specyfikę interpretacyjną [Ilnicki, Michalski, 2015; Śleszyński, 2013; Wiśniewski, 2013]. Wynika ona z faktu, że dane te (2006, 2011, 2016) pozyskano nie w bezpośredni, ale pośredni sposób, opierając się na tzw. źródłach administracyjnych. Podstawowym źródłem danych o wielkości i kierunkach dojazdów do pracy były dane zaczerpnięte z systemu podatkowego zgromadzone w bazie POLTAX. Głównymi informacjami były te zawarte w formularzach podatkowych PIT-11/8B oraz PIT-40 lub PIT 11B. W 2011 i 2016 dane pochodzące z rejestrów podatkowych zostały wzbogacone o informacje pochodzące z Zakładu Ubezpieczeń Społecznych (ZUS) oraz Kasy Rolniczego Ubezpieczenia Społecznego (KRUS). Jednak należy zwrócić uwagę, że dla wspomnianych wcześniej lat, głównym rejestrem w ustalaniu zbiorowości dojeżdżających był rejestr ZUS, a rejestr podatkowy miał charakter pomocniczy. Pozwoliło to między innymi na dokładniejszą identyfikację terytorialną miejsca pracy. Natomiast dane o dojazdach z 2021 r. posiadają odmienny charakter. Po pierwsze – dojazdy do pracy najmniej zostały pozyskane (następnie udostępnione) metodą mieszaną, tj. z wykorzystaniem danych zebranych od respondentów (NSP 2021) oraz danych pochodzących ze źródeł administracyjnych⁵³. Po drugie – w opublikowanych danych próg ich prezentacji, przepływów pomiędzy gminami, to nie 10 i więcej, a 3 i więcej osób. Obniżenie progu publikowanych danych z jednej strony spowodowało, że według przyjętej metodologii jest możliwym opisanie blisko 96% wszystkich faktycznych dojazdów do pracy. Z drugiej strony można ukazać powiązania w mniejszej skali przestrzennej, jak również wskazać, że dojazdy do pracy, nie tyle co nie mają tendencji do regionalnego domykania się, co wykazują się widocznym ich występowaniem na stykach, granicach pomiędzy regionami [por. Ilnicki, Michalski, 2015; Ilnicki, Janc, 2021].

Obecna statystyka dojazdów do pracy posiada dwa znaczące mankamenty. Pierwszy to ograniczenie zbiorowości pracujących do pracowników najemnych, a drugi to brak informacji, czy te dojazdy mają charakter dojazdów codziennych, wahałowych, cyklicznych [Ilnicki, Janc, 2021].

Migracje wewnętrzne na pobyt stały niezmiennie są elementem dostępnej publicznie statystyki. Najlepiej udokumentowany aspekt ruchliwości przestrzennej

⁵¹ W NSP 2021 zbierano dane które mogą pozwolić na powrót do danych o dojazdach do pracy bazujących na pełnych danych. Jak natomiast będzie miejmy nadzieję że niedługo się przekonamy.

⁵² Dane o dojazdach do pracy są publikowane, udostępniane przez GUS od 2006 r. w cyklach 5-letnich.

⁵³ Opis przetwarzania danych przedstawiono w publikacji [<https://stat.gov.pl/spisy-powszechne/nsp2021/nsp-2021-wyniki-ostateczne/ludnosc-na-ryнку-pracy-w-swietle-wynikow-narodowego-spisu-powszechnego-ludnosci-i-mieszkan-2021,11,2.html>].

społeczeństwa polskiego można uzyskać z Banku Danych Lokalnych (BDL) oraz z bazy danych „Demografia” GUS. Choć istnieje wiele baz danych, tabela 2g w bazie Demografia prezentuje wielkość migracji między gminami w układzie macierzowym. Niemniej jednak, ten układ nie jest efektywny w graficznym przedstawieniu przepływu migracji. Korzystnym byłoby udostępnianie tych danych w sposób analogiczny do tych o dojazdach do pracy. Obecnie dostępne dane o migracjach na pobyt stały obejmują lata 2002–2022.

Jednym z interesujących źródeł danych jest System Informacji Oświatowej tzw. baza SIO II. Baza ta jest w gestii przedmiotowego resortu i dostęp, bezpłatny i „otwarty”, do niej zawarty jest w zapisach aktów prawnych [*O systemie...*, 2011]. Pomijając listę podmiotów które mogą mieć w różnym zakresie do niej dostęp, nie ma w nim przedstawicieli szeroko rozumianej nauki. Baza ta zawiera między innymi jednostkowe informacje o miejscu zamieszkania i pobierania nauki od poziomu miejscowości statystycznej, poprzez kolejne poziomy podziału administracyjnego. Poza tym zawiera ona dane o kształceniu na poziomie przedszkolnym, szkoły podstawowej i średniej. Z tego też względu jest idealnym źródłem, w oparciu o które w różnych skalach przestrzennych, dla różnych typów ośrodków, można zidentyfikować wielkość i strukturę przemieszczeń wahałowych. Pozwala to na dokonanie delimitacji obszarów oddziaływania, ośrodków, placówek w różnych ujęciach. Co istotne, przemieszczeniom w tym zakresie towarzyszą przemieszczenia do pracy i innych usług. Można je zestawiać z dojazdami do pracy, czy też z migracjami wewnętrznymi na pobyt stały. Siłą, tych danych jest to, że dzięki nim można prowadzić analizę zróżnicowania zjawisk na poziomie miejscowości.

Dane te mogą być wykorzystywane przy tworzeniu koncepcji, polityk transportowych, rozwoju sieci transportu, połączeń komunikacyjnych, czy to dla obszarów metropolitalnych, aglomeracji, zespołów miejskich, czy też w układach wewnątrz-miejskich. Warto zaznaczyć, że GUS po raz pierwszy wykorzystał te dane opracowując raport dotyczący terytorialnego pochodzenia uczniów szkół podstawowych, gimnazjów oraz szkół ponadpodstawowych, szkół dla dorosłych [*Dojazdy uczniów do szkół...*, 2018]. Ciekawą konstatacją w oparciu o wyżej wspomniany raport, jak i informacje sygnałną [*Dojazdy uczniów do szkół...*, 2023], jest stwierdzenie o znacznym potencjale drzemącym w tego typu danych i ich utylitarnym charakterze.

Podobnymi w swojej naturze są dane o terytorialnym pochodzeniu studentów. Dane te nie mają, jakby się wydawało, charakteru rozproszonego. Od 2011 r. wszystkie uczelnie wyższe mają obowiązek raportowania informacji o studentach. Jednak są one pozbawione informacji o stałym miejscu zamieszkania, pozwalających na określenie relacji pomiędzy miejscem pobierania nauki, a miejscem pochodzenia studentów. W przypadku tych danych wydaje się, że najniższym poziomem odniesienia przestrzennego do ich analizy jest poziom miasta i gminy. Wydaje się, że poza samą wartością poznawczą dla badań naukowych mogą one posłużyć do oceny, weryfikacji i optymalizacji funkcjonującego podziału administracyjnego, terytorialnego. W tym

przypadku koniecznym byłaby łączna analiza tych danych z tymi dotyczącymi kształcenia na poziomie ponadpodstawowym (średnim), czy też kształcenia dorosłych.

„Nowe” dane – *big data*

Do momentu powszechnego dostępu do technologii cyfrowych, w tym głównie Internetu, mieliśmy do czynienia z sytuacją, w której występował deficyt informacji o otaczającym świecie. Geografowie, planiści przestrzenni opisując, poznając przestrzeń, wypełniali misję zbierania, przetwarzania, przekazywania danych, informacji do społeczeństwa. Obecnie wobec powszechnego dostępu do informacji w społeczeństwie pełnią oni rolę osób je „filtrujących”. Mają oni za zadanie nie tyle blokowanie zalewu nierzetelnych, nieprawdziwych informacji, ich interpretacji, co obiektywizować przekaz płynący z danych. Aby móc to czynić, świat nauki musi bazować na zobiektywizowanych danych, których pozyskanie posiada jasno zdefiniowane kryteria i uwarunkowania ich pozyskania – czyli „dobre” dane. Dostęp do informacji cyfrowych można określić mianem natychmiastowego i niemal nieograniczonego, jednak nie jest to regułą, gdyż większość danych jest trudna lub niemożliwa do pozyskania na rzecz badań naukowych. Szczególnie dotyczy to danych gromadzonych przez instytucje komercyjne (np. operatorów usług telekomunikacyjnych, instytucje finansowe) – danych związanych ze specyfiką ich działania.

W okresie ostatnich kilku lat można dostrzec wzrost wykorzystania, w celu identyfikacji i poznania cech procesów przestrzennych, zwłaszcza związanych z ruchliwością, danych będących tzw. „cyfrowym śladem” [González–Bailón, 2013]. Są one efektem powszechnego korzystania przez ludzi z urządzeń podłączonych do Internetu, w tym smartfonów. W efekcie tego procesu tworzona jest reprezentacja świata – efekt wielowymiarowej geografii codziennego życia użytkowników [Shelton i in., 2015]. Tego typu dane są jedną ze składowych tzw. *big data*. Big data pochodzą m.in. z: różnego rodzaju czujników, sieci WWW, danych wytworzonych w wyniku interakcji użytkownik – Internet [Anselin, Williams, 2016]. Duża część *big data* może być zaliczona do danych przestrzennych (około 80%). Zawierają one informacje o współrzędnych geograficznych, geograficzne metadane, adres, czy inne odwołania do przestrzeni rzeczywistej [Leszczynski, Crampton, 2016]. Dane te umożliwiają odtworzenie, czy nawet obserwowanie w czasie rzeczywistym codziennych zachowań ludzi, ich działań w przestrzeni. Z perspektywy geografii społeczno–ekonomicznej i gospodarki przestrzennej *big data* stosowane są między innymi do poszukiwania odpowiedzi na klasyczne problemy (m.in. wyznaczanie granic obszarów i regionów oddziaływania; dominacji rynkowej; mobilności przestrzennej, a szczególnie tej codziennej – wadliwej) [Rzeszewski, Rodak, 2019]. Mamy więc do czynienia z poszerzeniem

zakresu danych będących podstawą dociekań odnośnie ruchliwości, zachowań, procesów przestrzennych.

W kontekście ruchliwości przestrzennej z zakresu nowych źródeł danych na pewno należy wymienić: transakcje, wypłacanie pieniędzy z bankomatów; transakcje bezgotówkowe z wykorzystaniem kart płatniczych oraz innych sposobów autoryzacji transakcji płatniczych – płatności; logowań do sieci telefonii komórkowej. Wydaje się, że mają one zbliżoną, jeśli nie podobną wartość wyjaśniającą.

Źródłem danych o dużym potencjale analitycznym są transakcje dokonywane przez użytkowników bankomatów. Same bankomaty w polskiej literaturze geograficznej były analizowane w pierwszej dekadzie XXI w. Przedmiotem analiz naukowych były przestrzenne zróżnicowania lokalizacji bankomatów, tak jak i transakcji bankomatowych [Ilnicki, 2009, 2010]. W większym stopniu były to analizy dotyczące posadowienia i zróżnicowania w przestrzeni ich liczby. Liczba ta w danej lokalizacji była traktowana jako relatywny poziom zamożności, poziom przepływów finansowych. Dane o dokonywanych transakcjach bankomatowych w połączeniu z ich lokalizacją przypisaną do danego ośrodka, jednostki osadniczej pozwalają wyznaczyć ich obszary oddziaływania. Powiązanie miejsca zamieszkania z miejscem realizacji danej transakcji mówi o relacjach pomiędzy miejscami, przemieszczeniami które mają charakter rutynowy, powtarzalny i incydentalny. Interesującą kwestią w przypadku poboru pieniędzy z bankomatów jest ich sezonowość, w układzie miesięcy, dni tygodnia i cyklu dobowego.

Główną trudnością poza pozyskaniem danych odnośnie transakcji bankomatowych, jest fakt obecności wielu operatorów bankomatów. Jednak w przypadku większości badań wystarczające powinno być dla możliwości formułowania generalnych wniosków pozyskanie danych od operatora bankomatów z dużym wolumenem dokonywanych operacji, a w szczególności zlokalizowanych na terenie całego kraju. Kolejnym ograniczeniem może być pozyskanie informacji o miejscu zamieszkania, zameldowania właściciela karty płatniczej. Poziomem „minimum” powinno być uzyskanie informacji o gminie, powiecie i województwie zamieszkania. Jednak, przy niechęci udostępnienia lokalizacji o takim poziomie „szczegółowości”, elementem definiującym miejsce zamieszkania może być kod pocztowy. W przypadku tego źródła danych można by pozyskać również dwie kluczowe informacje o osobie przemieszczającej się, a mianowicie o jej wieku i płci. Pomimo że mamy do czynienia, ze spadkiem liczby bankomatów oraz liczby transakcji związanych z poborem pieniędzy to w dalszym ciągu w źródle tym odnaleźć można potencjał analityczny, wyjaśniający i użyteczny.

Kolejne dwa źródła to te pochodzące z rejestracji bezgotówkowych transakcji płatniczych oraz pochodzącej z logowań telefonii komórkowej. Trudność pozyskania tych danych, poza wcześniej wspomnianymi, polega na tym, że dysponenti danych są niezależni od państwa, a konkretnie od sprawozdawczości statystycznej. Kolejnym mankamentem tych danych może być konieczność płatności za pozyskane dane, a które to kwoty są poza zasięgiem, „jednego” badacza (nie)posiadającego projekt badawczy. Tak jak wcześniej sygnalizowano, kluczową informacją są dane definiujące

użytkownika, osobę dokonującą płatności bezgotówkowej, czy to posiadacza telefonu komórkowego. Jakkolwiek dane użytkownika można uznać za stałe, to zmienne i żmudne w geokodowaniu byłyby informacje o miejscu realizacji danej płatności. Wydaje się, że w przypadku tego typu danych, podobnie jak przy transakcjach bankomatowych, warto zwrócić uwagę na jedną istotną wielkość, która często umyka z pola widzenia, a mianowicie przypisanie określonych interakcji do danego użytkownika. Taka informacja może bazować na numerze PESEL, czy numerze karty płatniczej, czy też numerze telefonu, dla których zostaną przypisane kody bazujące na tych wielkościach⁵⁴. Wcześniej sygnalizowano, że czas zajścia interakcji, zdarzenia jest ważny. Są to wielkości o wiele bardziej istotne niż niepowtarzalna informacja o użytkowniku. Należy też zwrócić uwagę na fakt, że kluczowym dla opracowywania danych, poza tezami, hipotezami, pytaniami badawczymi, które formułuje się przed badaniem, jest wiedza jak prezentuje się struktura danych które są możliwe do uzyskania. Bez względu na wcześniej poruszane kwestie ważny jest też horyzont czasowy pozyskanych danych, jak również powtarzalność cykliczność ich pozyskiwania. W przypadku danych pochodzących od operatorów telefonii komórkowej zasadniczym problemem jest brak wypracowanej metodologii metodyki opracowywania tych danych, u podłoża których leży sam proces (s)tworzenia algorytmu pozyskania danych które są w danym badaniu niezbędne.

Jednym z najbardziej obiecujących zbiorów danych – cyfrowych śladów, są te pochodzące z telefonów komórkowych, konkretnie z logowań do stacji przekaźnikowych (bazowych) [Kung i in., 2014; Louail i in., 2014]. Badania tego typu w polskiej rzeczywistości można określać mianem pionierskich. W tym miejscu można przywołać jedno z badań dla Wrocławia [*Kompleksowe badanie ruchu...*, 2018]. Na jego podstawie można sformułować dwa spostrzeżenia, a mianowicie: wysoka ruchliwość, powiązanie otoczenia z Wrocławiem zamyka się w granicach byłego województwa wrocławskiego; najwyższe wartości powiązań dotyczą bezpośredniego otoczenia Wrocławia oraz drugiego pasa gmin. W obraz tak identyfikowanej ruchliwości przestrzennej bardzo dobrze wpisuje się obszar funkcjonalny Wrocławia [Ilnicki, Janc 2021]. Inne z tego typu badań umożliwiły określenie liczby odwiedzających lasy na pograniczu polsko-czesko-niemieckim – ich czasowo-przestrzenne wzorce zachowań w przestrzeni [Ciesielski, Tkaczyk, 2023]. Wskazuje to na potencjał tych danych do określania procesów przestrzennych.

Wyposażenie niemal każdego urządzenia przenośnego w moduł GPS, konieczny do nawigacji, lokalizowania obiektów w przestrzeni, powoduje, że istotnym źródłem informacji o przemieszczaniu się ludności stają się *big data* pochodzące z urządzeń rejestrujących zmiany lokalizacji. Smartfony, czy inne urządzenia podłączone do Internetu, mogą być wykorzystywane do określania przemieszczeń ich użytkowników w przestrzeni fizycznej na kilka sposobów. Między innymi prowadzone były badania nad przemieszczeniami, w których do pozyskania danych wykorzystywano specjalne

⁵⁴ Chodzi o tak zwane rekodowanie albo „haszowanie”.

aplikacje na telefony komórkowe – instalowane przez ochotników biorących udział w badaniach [Gadziński, 2017]. Innym źródłem danych są media społecznościowe [m.in. Curry i in., 2019]. W części badań skorzystano, w celu określenia wzorców przemieszczania się, z danych oferowanych przez Google Trends [Chan, 2020]. Innym źródłem danych mogą być te pochodzące z serwisów (aplikacji) umożliwiających wyszukiwanie obiektów, oglądanie map, zdjęć lotniczych powierzchni Ziemi, takich jak Mapy Google, czy Mapy Baidu [m.in. Wei i Wang, 2020].

Podsumowanie

Uwzględniając potrzeby prowadzenia badań naukowych, to pomijając dane pochodzące z oficjalnej statystyki, pozostałe dane są możliwe do uzyskania, ale bez poniesienia wysokich kosztów ich pozyskania (i opracowania), praktycznie są niedostępne. Wydaje się, że w większym stopniu należy wywierać nacisk na GUS, aby „pozyskiwał” i udostępniał dane do bezpośredniego przetworzenia, użycia oraz aby były one publikowane cyklicznie, a nie były tylko efektem badań eksperymentalnych. Poza tym konieczne jest umożliwienie pozyskiwania danych GUS z BDL (Bank Danych Lokalnych) nie tylko w ich podziale na poszczególne poziomy podziału administracyjnego, terytorialnego, ale również dla miejscowości statystycznych.

T. Strykiewicz [2011] wskazuje na niedostatecznie rozpoznane przez geografów zagadnienia, wśród których wymienia między innymi: konieczność identyfikacji nowych czynników lokalizacji działalności gospodarczej; kształtowanie się nowych form i układów komunikacyjnych oraz popytu na usługi komunikacyjne; migracje, nowe formy mobilności przestrzennej i ich zmieniające się uwarunkowania; dyfuzję przestrzenną. Wszystkie te niedoskonałości dotychczasowych badań, szczególnie w skali lokalnej dla całego kraju, czy też szerszej perspektywie, można chociaż częściowo zniwelować wykorzystując źródła o których wspomniano w niniejszym opracowaniu. Środowisko naukowe powinno wykazywać się większą determinacją w odkrywaniu i pozyskiwaniu nowych źródeł danych, a otoczenie instytucjonalne powinno być otwarte na stworzenie przestrzeni, kiedy nauka nie tylko będzie się spotykać, stykać z rzeczywistością społeczno-gospodarczą, ale będzie odpowiadać na jej zapotrzebowanie w opisie, wyjaśnianiu, optymalizacji przestrzeni społecznej, gospodarczej, infrastrukturalnej. Wówczas nauka zacznie mieć faktyczny wpływ na otoczenie społeczno-gospodarcze.

Dostęp danych na poziomie jednostkowym (dotyczących zachowań pojedynczych osób) prowadzi do powstania bardzo dużych zbiorów danych, nie zawsze łatwo podających się analizie, czy posiadających jasną, uporządkowaną strukturę. Mają one natomiast tą istotną przewagę, nad danymi pochodzącymi z oficjalnych statystyk, że mogą opisywać zjawiska niemal w czasie rzeczywistym. Umożliwiają również analizy

zachowań przestrzennych w skali godzin, pór dnia, dni czy okresów kilkudniowych lub kilkutygodniowych.

Źródła danych o których wspomniano w opracowaniu nie są niemożliwe do pozyskania. W swoim całościowym zakresie są danymi wrażliwymi, co nie znaczy że nie można ich anonimizować, czy też udostępniać w postaci, która nie narusza RODO lub tajemnicy statystycznej mającej swe źródło w ustawie o statystyce publicznej (Dz. U. 2022 poz. 459). Duża rola w udostępnianiu i dostępie do danych leży po stronie ich dysponenta. Wydaje się, że zasłanianie się tajemnicą handlową albo interesem klienta, przy masowości zachowań, interakcji nie jest zasadne. Wydaje się, że dysponentom, właścicielom danych, nie zależy na ich udostępnieniu. Nie mają, czy też nie chcą widzieć w tym korzyści. Zasłanianie się tajemnicą handlową, przedsiębiorstwa, czy też konieczność wskazywania na interes społeczny, przy próbie udostępniania danych w trybie zapytania o dane publiczne, tylko wydłuża sam proces, nadaje mu charakter dialogu, ale w ostatecznym rozrachunku niczego nie zmienia. Kończy się na odmowie udostępnienia danych, a w najgorszym przypadku brakiem odpowiedzi.

Jakkolwiek w przypadku danych pochodzących od „zewnętrznych operatorów” można zrozumieć ich brak uwagi do udostępniania danych, tak w przypadku danych mających swoje źródło w rejestrach instytucji państwowych trudno bronić stanowiska dotyczącego nieudostępniania danych. Wydaje się, że koniecznym jest przywrócenie „służebnej” roli statystyki publicznej w kontekście badań naukowych i ich użytecznego charakteru. Koniecznym jest dążenie do tego, żeby publikowane przez GUS dane nie były tylko i wyłącznie realizacją programu badań statystycznych, ale żeby one poza ten program wychodziły. Muszą one opisywać rzeczywistość i te jej cechy, które są istotne dla życia społeczno-gospodarczego. Poza tym nie mogą one mieć charakteru danych cyklicznych, tylko ciągłych. Wszędzie gdzie to tylko możliwe, powinny być one dostępne na możliwie najniższym poziomie odniesienia przestrzennego.

W przypadku wykorzystania *big data* w badaniach naukowych stykamy się z wieloma ograniczeniami metodologicznymi, założeniami które są zmienne w czasie i zależne od danego podejścia badawczego. Tym samym kolejne badania tego samego w „ten sam sposób” tylko z założenia opisują to samo. Czasami drobne niuanse powodują że interpretacja danych, wnioski, tezy czasami są do siebie zbliżone to akcenty na kwestie będące przedmiotem badania są już inaczej rozłożone [Cierpień-Wolan, Stateva, 2023; Daas, Maślankowski, 2023]. Na pewno w większym stopniu należy korzystać z danych które są już gromadzone, do których dostęp jest szybszy i może mieć charakter stały, a nie są odpowiednio często wykorzystywane w badaniach naukowych. Jednakże w tym przypadku jest konieczny udział służb statystycznych i dobra wola ustawodawcy, władz, aby takie dane zostały uwolnione i ostatecznie udostępnione.

- Anselin L., Williams S., 2016, *Digital neighborhoods*. „Journal of Urbanism: International Research on Placemaking and Urban Sustainability” nr 9(4): 305–328.
- Chan J., 2020, *Using Google Data to Understand Canadian Movement Reductions During the COVID-19 Pandemic*. „SSRN”.
- Cierpiał–Wolan M., Stateva G., 2023, *The evaluation of (big) data integration methods in tourism*. „Wiadomości Statystyczne” t. 68(12): 25–48.
- Ciesielski M., Tkaczyk M., 2023, *Visits in Forests During the Covid-19 Pandemic in the Cross-Border Area of Poland, the Czech Republic and Germany*. „Quaestiones Geographicae” nr. 42(2): 71–84.
- Churski P., Zwoliński Z., 2011, *Funkcje poznawcze i praktyczne nauki o informacji geograficznej (GISc)*, [w:] Kostrzewski A., Maik W., Brudnicki R. (red.), *Geografia wobec problemów współczesności. Funkcje poznawcze i praktyczne geografii*. Wydawnictwo Uczelniane Wyższej Szkoły Gospodarki w Bydgoszczy, Bydgoszcz: 95–106.
- Curry T., Croitoru A., Crooks A., Stefanidis A., 2019, *Exodus 2.0: Crowdsourcing geographical and social trails of mass migration*. „Journal of Geographical Systems” nr 21(1): 161–187.
- Daas P., Maślankowski J., 2023, *Current challenges and possible big data solutions for the use of web data as a source for official statistics*. „Wiadomości Statystyczne” t. 68(12): 49–64.
- Dojazdy uczniów do szkół zlokalizowanych w miastach wojewódzkich*, 2018, Prace studialne, GUS, Warszawa.
- Dojazdy uczniów do szkół podstawowych i ponadpodstawowych w roku szkolnym 2022/2023*, 2023, Informacja sygnałna, GUS.
- Ustawa z dnia 29 czerwca 1995 r. o statystyce publicznej* (Dz. U. 1995 Nr 88 poz. 439 z 29 czerwca 1995, tekst jednolity z 10 marca 2023).
- Ustawa z dnia 15 kwietnia 2011 r. o systemie informacji oświatowej* (Dz. U. 2011 Nr 139 poz. 814 z 15 kwietnia 2011, tekst jednolity z 19 października 2023).
- Gadziński J., 2017, *Wykorzystanie telefonów komórkowych w badaniach zachowań transportowych ludności*. „Prace Komisji Geografii Komunikacji PTG” nr 20(4): 7–19.
- González–Bailón S., 2013, *Big data and the fabric of human geography*. „Dialogues in Human Geography” nr 3 (3): 292–296.
- Ilnicki D., 2009, *Przestrzenne zróżnicowanie poziomu rozwoju usług w Polsce. Teoretyczne i praktyczne uwarunkowania badań*. Rozprawy Naukowe Instytutu Geografii i Rozwoju Regionalnego Uniwersytetu Wrocławskiego, Wrocław, t. 11.
- Ilnicki D., 2010, *Wybrane przestrzenne aspekty funkcjonowania bankomatów Polsce*, [w:] Ciok S., Migoń P. (red.), *Przekształcenia struktur regionalnych. Aspekty społeczne, ekonomiczne i przyrodnicze*. Instytut Geografii i Rozwoju Regionalnego, Uniwersytet Wrocławski, Wrocław: 303–320.
- Ilnicki D., Janc K., 2021, *Obszary intensywnych powiązań funkcjonalnych miast na prawach powiatu w Polsce – autorska metoda delimitacji*. „Przegląd Geograficzny” nr 93(2): 141–160.
- Ilnicki D., Michalski P., 2015, *Powiązania funkcjonalno–przestrzenne w świetle dojazdów do pracy*. „Studia Miejskie” t. 18: 55–70.

- Kompleksowe badanie ruchu we Wrocławiu i otoczeniu – KBR 2018*, 2018, Raport z realizacji etapu II.
- Kostrzewski A., Maik W., Brudnicki R. (red.), 2011, *Geografia wobec problemów współczesności. Funkcje poznawcze i praktyczne geografii*. Wydawnictwo Uczelniane Wyższej Szkoły Gospodarki w Bydgoszczy, Bydgoszcz.
- Kung K.S., Greco K., Sobolevsky S., Ratti C., 2014, *Exploring universal patterns in human home–work commuting from mobile phone data*. „PloS ONE” nr 9(6): e96180.
- Leszczynski A., Crampton J., 2016, *Introduction: Spatial big data and everyday life*. „Big Data & Society” nr 3(2): 1–6.
- Louail T., Lenormand M., Ros O.G.C., Picornell M., Herranz R., Frias–Martinez E., Ramasco J.J., Barthelemy M., 2014, *From mobile phone data to the spatial structure of cities*. „Scientific reports” nr 4: 5276.
- Rzeszewski M., Rodak O., 2019, *Czy więcej znaczy lepiej? Badania ilościowe w geografii społeczno–ekonomicznej ery Big Data*. „Rozwój Regionalny i Polityka Regionalna” nr 46: 67–89.
- Shelton T., Poorthuis A., Zook M., 2015, *Social media and the city: Rethinking urban socio–spatial inequality using user–generated geographic information*. „Landscape and Urban Planning” nr 142: 198–211.
- Stryjakiewicz T., 2011, *Rola geografii w badaniu procesów i struktur społeczno–ekonomicznych w skali globalnej i regionalnej*, [w:] Kostrzewski A., Maik W., Brudnicki R. (red.), *Geografia wobec problemów współczesności. Funkcje poznawcze i praktyczne geografii*. Wydawnictwo Uczelniane Wyższej Szkoły Gospodarki w Bydgoszczy, Bydgoszcz: 49–58.
- Śleszyński P., 2013, *Warszawa jako ośrodek dojazdów pracowniczych*. „Studia Regionalne i Lokalne” nr 1(51): 5–25.
- Wei S., Wang L., 2020, *Examining the population flow network in China and its implications for epidemic control based on Baidu migration data*. „Humanities and Social Sciences Communications” nr 7(1): 1–10.
- Wiśniewski R., 2013, *Společno–demograficzne uwarunkowania dojazdów do pracy do Białegostoku*. Prace Geograficzne IGiPZ PAN, Warszawa, nr 244.