

Elżbieta Gołata^{*}
Maciej Beręsewicz^{**}

Przyszłość badań statystycznych wobec potrzeb społecznych i gospodarczych na przykładzie spisu ludności

Streszczenie: Spis ludności jest podstawowym źródłem informacji o stanie i strukturze ludności oraz jej rozmieszczeniu terytorialnym. Badanie to stanowi zasadniczą część narodowego i międzynarodowego systemu statystyki oficjalnej przede wszystkim dlatego, że dostarcza szacunków liczby ludności. Spis jest jednym z najstarszych badań statystycznych, ale... czy to badanie jest nam dzisiaj nadal potrzebne? Czy statystyka oficjalna, której zadaniem jest zapewnienie rzetelnego systemu informacji, jest w stanie sprostać tym oczekiwaniom? W rozdziale podjęto próbę odpowiedzi na powyższe pytania poprzez omówienie zmian organizacyjnych badań statystycznych, metod przeprowadzania oraz transformacji spisów ludności, a zwłaszcza zagrożeń i perspektyw na przyszłość związanych z wykorzystaniem pozastatystycznych źródeł danych (m.in. rejestrów administracyjnych czy *big data*), jako alternatywnych metod przeprowadzania spisu ludności.

Słowa kluczowe: statystyka publiczna, spisy powszechne, rejestry administracyjne, *big data*

1. Wstęp

Podejmując wyzwanie udziału w dyskusji nt. ewolucji nauk ekonomicznych i relacji do innych nauk, odniesiemy się do przyszłości badań statystycznych, zwłaszcza spisu ludności jako badania statystycznego dotyczącego populacji osób zamieszkujących terytorium danego państwa. Związków demografii i statystyki jest wiele¹. Wiele jest też związków obu tych nauk z ekonomią, nie tylko z racji klasyfikacji dziedzin nauki i wyodrębnionych dyscyplin. To praca – czyli ludzie, są obok kapitału i ziemi podstawowym czynnikiem produkcji. Statystyka też odgrywa służebną rolę wobec

^{*} Uniwersytet Ekonomiczny w Poznaniu.

^{**} Uniwersytet Ekonomiczny w Poznaniu.

¹ John Graunt, autor słynnego dzieła pt. *Natural and political observation made upon the bills of mortality* (<http://www.hetwebsite.net/het/profiles/graunt.htm>) wydanego w 1662 roku uznawany jest za ojca demografii i statystyki [Kurkiewicz, 2010].

innych nauk, m.in. jako odpowiedzialna za metody pozyskiwania wiarygodnych danych opisujących otaczającą nas rzeczywistość: gospodarkę i społeczeństwo.

Spis ludności jest podstawowym źródłem informacji o stanie i strukturze ludności oraz jej rozmieszczeniu terytorialnym. Badanie to stanowi zasadniczą część narodowego i międzynarodowego systemu statystyki oficjalnej przede wszystkim dlatego, że dostarcza szacunków liczby ludności [Kordos, 2017; Kotowska, 2009]. Informacje te są wykorzystywane przez różne instytucje i organizacje samorządowe, krajowe i międzynarodowe, w rachunkach narodowych, w badaniach porównawczych, czy w celu konstrukcji prognoz ludności. Spis jest jednym z najstarszych badań statystycznych, ale... czy to badanie jest nam dzisiaj nadal potrzebne? Czy spis dostarcza informacji pozwalających na badanie zmian w zakresie reprodukcji ludności, jakie ukształtowały się w wyniku przejścia demograficznego² czy kolejnej fali gwałtownych przeobrażeń społecznych określanych mianem drugiego przejścia?³ Czy dane spisu ukazują rzeczywisty obraz przemian procesu tworzenia rodzin i ich destabilizacji, upowszechniania związków nieformalnych, skomplikowaną strukturę rodzin? Jak dane spisu ukazują rozmieszczenie ludności, zwłaszcza wobec intensywnej mobilności i suburbanizacji, wielu miejsc zamieszkania, różnorodności form zatrudnienia. Czy spis odzwierciedla obserwowane zmiany na rynku pracy, zaawansowanie starzenia się populacji, czy spełnia oczekiwania samorządowców, przedsiębiorców i polityków społecznych, czy może stanowić podstawę konstrukcji wskaźników makroekonomicznych, alokacji środków państwowych na dotacje dla gmin, systemu edukacji czy opieki zdrowotnej, na zasiłki opiekuńcze, pomoc dla bezrobotnych, aktywizację zawodową itp. Czy statystyka oficjalna, której zadaniem jest zapewnienie rzetelnego systemu informacji, jest w stanie sprostać tym oczekiwaniom?

W ostatniej rundzie spisów około 2010 roku byliśmy świadkami historycznych przeobrażeń o niespotykanym zasięgu. Ogromne zmiany i modyfikacje dotyczyły każdego etapu spisu. Dążąc do redukcji kosztów, ale także poprawy jakości szacunków, terminowości i formy upowszechniania wyników, poszczególne państwa decydowały się na wprowadzanie nowoczesnych rozwiązań technologicznych (ICT) czy alternatywnych metod przeprowadzania spisu [UN, 2013a]. Szeroko wykorzystywano technologie Geograficznych Systemów Informacyjnych (Geographical Information Systems, GIS), urządzenia mobilne, kwestionariusze internetowe, mapy cyfrowe, systemy nawigacji GPS (Global Positioning System), telefony komórkowe,

² Przejściem demograficznym określa się proces zmian reprodukcji ludności na tle przeobrażeń społeczno-ekonomicznych opisany przez teorię przejścia demograficznego, za której twórców uznaje się W.S. Thompsona i F. Notesteina [Kędelski, Paradysz, 2006; Okólski, Fihel, 2012].

³ Intensywne zmiany w sferze małżeńskiej i rozrodczości prowadzące do niemal jednoczesnego w wielu krajach spadku płodności poniżej poziomu prostej zastępowalności pokoleń, określane są mianem koncepcji drugiego przejścia demograficznego [Lesthaeghe, van de Kaa, 1986; Okólski, Fihel, 2012; van de Kaa, 1987].

umożliwiający monitorowanie prac w terenie. Coraz większe zainteresowanie skupiały alternatywne metody przeprowadzania spisu wykorzystujące zasoby rejestrów administracyjnych. We Francji zastosowano metodykę spisu kroczącego, a w Stanach Zjednoczonych, tradycyjne badanie reprezentacyjne gromadzące dane tzw. długiego formularza spisowego, zastąpiono badaniem gospodarstw domowych – *American Community Survey* (ACS).

2. Zmiana organizacji badań statystycznych

Zmiana w zakresie organizacji badań statystycznych dotyczy nie tylko spisu ludności, ale także innych badań społecznych i ma charakter fundamentalny. Przede wszystkim obejmuje ona nie tylko sposób gromadzenia danych, ale modelowe ujęcie procesu badania statystycznego, rozumianego w sposób kompleksowy [Wallgren, Wallgren, 2014]. Przez wieki spis powszechny był podstawowym sposobem pozyskiwania informacji o całej populacji danego kraju. Rozwój metody reprezentacyjnej wprowadził możliwość ograniczenia badania do losowo wybranych z tej populacji jednostek. W drugiej połowie XX wieku powszechne spisy ludności przeprowadzano zazwyczaj raz na 10 lat⁴, gdyż badanie to było zbyt kosztowne i czasochłonne. Obecnie jednak poszukując informacji o populacji, zanim zaprojektowane zostanie badanie reprezentacyjne, w pierwszej kolejności dokonuje się kwerendy istniejących źródeł, przede wszystkim rejestrów administracyjnych, ale także innych, pozastatystycznych źródeł⁵.

Wzrost wykorzystania rejestrów administracyjnych przez statystykę publiczną Wallgren i Wallgren [2014] określają mianem transformacji badań statystycznych w kierunku systemu opartego na rejestrach z systemu bazującego na gromadzeniu danych w wyniku indywidualnych wywiadów. Dane z tych źródeł są dostępne natychmiast, ale charakteryzuje je duża różnorodność, odmienna jakość i właściwości. Dlatego szczególnego znaczenia nabiera ich wnikliwa ewaluacja, edycja i integracja. Wykorzystanie rejestrów przez statystykę publiczną wymaga rozwiązania wielu problemów, m.in. integracji danych, estymacji pośredniej, imputacji, kalibracji, zastosowania technologii GIS, niestandardowych konstrukcji operatu losowania, badań internetowych, upowszechniania danych i ochrony ich poufności [Baffour

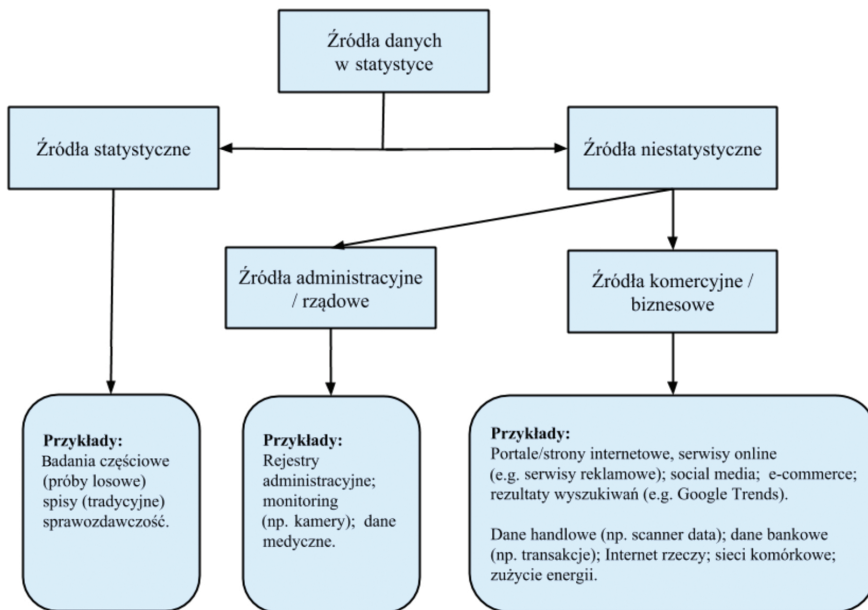
⁴ W niektórych państwach spisy sporządzano co 5 lat, tak było np. we Francji (por. rozdz. 1.1) i w Kanadzie, gdzie ta praktyka nadal się utrzymuje. W Polsce w okresach międzyspisywanych trzykrotnie przeprowadzono tzw. mikrospisy. Były to spisy przeprowadzone metodą reprezentacyjną w latach: 1974, 1984 i 1995.

⁵ Przykładem mogą być spisy abonentów telefonicznych, wykazy odbiorców energii elektrycznej, dane geodezyjne. Ponadto nowoczesne technologie informacyjno-komunikacyjne dostarczają niewiarygodnych ilości danych z dotychczas niedostępnych źródeł. *Big data* są rzeczywistością, stanowiącą wyzwanie dla statystyki publicznej, przede wszystkim pod względem oceny reprezentatywności [Beręsewicz, 2017].

i in., 2013]. Oznacza to reorganizację statystyki publicznej, co może być postrzegane w kategoriach zmiany paradygmatu badań statystycznych [Zhang, 2015a; 2015b].

Podjęcie sugerujące poprzedzenie decyzji o przeprowadzaniu badania oceną istniejących źródeł, wymaga również uwzględnienia ich różnorodności oraz wyróżnienia ich odmiennych właściwości. Jednym z częściej proponowanych rozwiązań – klasyfikacji, jest podział na źródła statystyczne i pozastatystyczne [Nordbotten, 2010; Wallgren, Wallgren, 2014]. Przyjmuje się przy tym, że dane administracyjne mają charakter wtórny, są one gromadzone przede wszystkim w celach niestatystycznych, ale mogą być wykorzystane w procesie tworzenia statystyk. Natomiast dane statystyczne mają charakter pierwotny, są gromadzone w trakcie badania statystycznego i zasadniczo nie są dostępne do wykorzystania w żadnym innym celu. Jedną z częstszych klasyfikacji źródeł danych wykorzystywanych w statystyce zakłada właśnie podział na statystyczne i pozastatystyczne (por. rys. 1).

Rysunek 1. Klasyfikacja źródeł danych w statystyce z uwzględnieniem nowych źródeł i ich gestorów



Źródło: Beręsewicz [2016, s. 28].

Źródła administracyjne, choć najczęściej wskazywane, nie są ani jedynymi, ani najobszerniejszymi pozastatystycznymi źródłami danych wykorzystywanymi w badaniu statystycznym. Coraz częściej statystyka publiczna sięga po źródła komercyjne/biznesowe, które powstały w innym celu – wspomnienia działalności gospodarczej lub wręcz oparcia działalności określonych przedsiębiorstw na zbieraniu i wyko-

rzystaniu danych. W tym przypadku można wskazać różnego rodzaju informacje zbierane w sposób zależny (np. dokonanie transakcji) lub niezależny (np. łączenie telefonu komórkowego ze stacją bazową) od użytkownika.

Należy jednak zaznaczyć, że pojęcie *big data* może dotyczyć zarówno źródeł administracyjnych (np. monitoringu ruchu), jak i komercyjnych. Kwestia ustalenia gestora określonych źródeł może być kluczowa, aby te dane pozyskać i wykorzystać na potrzeby statystyki publicznej.

3. Metody przeprowadzania spisów ludności

W ostatniej dekadzie sposób organizacji spisu ludności uległ zasadniczym zmianom. W Rekomendacjach Konferencji Statystyków Europejskich z 2006 roku [UNECE, 2006], czy w Rekomendacjach ONZ z 2015 roku [UN, 2015, s. 18] przedstawiono klasyfikację metod przeprowadzania spisów wyróżniając dwa kryteria: sposobu gromadzenia danych i wykorzystanych źródeł danych, w tym przede wszystkim rejestrów administracyjnych. Można wskazać wiele metod przeprowadzania spisu ludności od tradycyjnego spisywania w terenie, po spis określany mianem „wirtualnego”, który wbrew nazwie dostarcza jak najbardziej realnych informacji.

Spis tradycyjny

W spisie tradycyjnym informacje dotyczące osób fizycznych i gospodarstw domowych zbierane są przy użyciu kwestionariusza spisowego bezpośrednio w terenie. Możliwe jest zastosowanie takich metod jak wywiad bezpośredni, CAPI, CATI, CAII⁶, wywiad telefoniczny, pocztowy, poprzez Internet lub kombinacja różnych form. Informacje spisowe mogą być gromadzone za pomocą jednego bądź kilku formularzy. Formularz tzw. krótki obejmujący pytania tzw. minimum spisowego stosowany jest w badaniu pełnym, natomiast długi formularz w doraźnym badaniu reprezentacyjnym (lub ciągłym – spis kroczący). Zwyczajowo długi i krótki kwestionariusz wykorzystuje się razem w ramach jednego spisu w tym samym czasie. Wyniki badania reprezentacyjnego obejmującego tzw. tematy uzupełniające (długi formularz) traktowane są także jako wyniki spisu.

Spis kroczący

Informacje o osobach i gospodarstwach domowych są gromadzone w wyniku badania ciągłego i kumulowane, obejmując stopniowo obszar całego państwa w czasie zazwyczaj kilku lat, w przeciwieństwie do jednego dnia tzw. momentu krytycznego

⁶ Skrót CAPI, CATI czy CAII oznaczają odpowiednio wspomagany komputerowo wywiad osobisty, telefoniczny czy internetowy.

lub krótkiego okresu. Wyróżnia się dwa podstawowe parametry spisu kroczącego: długość okresu gromadzenia informacji i wielkość próby. Pierwszy parametr związany jest z wymaganą częstością aktualizacji, a drugi z wielkością najmniejszych jednostek podziału terytorialnego, dla których wymagane są szacunki spisowe. Od roku 2004 metodyka spisu kroczącego stosowana jest we Francji.

Spis w pełni oparty na rejestrach, wykorzystujący także inne źródła administracyjne

Informacje o osobach i gospodarstwach domowych są gromadzone z istniejących źródeł administracyjnych, przede wszystkim rejestrów, z których najważniejsze znaczenie mają rejestry ludności i mieszkań. Pierwsze spisy wykorzystujące dane rejestrów administracyjnych przeprowadzono w Finlandii i Norwegii już w 1970 roku [Tønder, 2008]. Początkowo wykorzystano jedynie dane rejestru ludności, a następnie w kolejnych spisach stopniowo wprowadzano dalsze rejestry dotyczące zagadnień społecznych i gospodarczych⁷. Podstawowe znaczenie mają rejestry ludności i mieszkań, których dane łączone są na poziomie jednostkowym z danymi rejestru ubezpieczeń, służby zdrowia, podatkowego, edukacyjnego itp., tworząc tzw. „złoty rekord”. Łączenie rejestrów przeprowadzane jest na podstawie klucza identyfikacyjnego. W Polsce takim identyfikatorem może być np. numer PESEL⁸.

Spis wirtualny⁹

Jest to zastosowany w latach 2001 i 2011 w Holandii złożony system zintegrowanych mikrobaz danych o spójnych definicjach, w którym informacje o osobach i gospodarstwach domowych są zbierane z istniejących źródeł administracyjnych. Są one powiązane na poziomie jednostkowym z informacjami pochodzącymi z istniejących badań reprezentacyjnych takich jak badanie siły roboczej oraz badanie warunków i jakości życia. Utworzona Statystyczna Baza Danych Społecznych łączy rejestry ludności, zatrudnionych i samozatrudnionych, rejestr podatkowy, emerytur i ubezpieczenia społecznego, pomocy społecznej, rejestr budynków. Połączenie danych z różnych źródeł przeprowadzone jest na poziomie jednostkowym przy wykorzystaniu różnych metod statystycznej integracji danych, w tym za pomocą unikatowego klucza identyfikacyjnego.

⁷ W rundzie spisów roku 1980 w Danii, roku 1990 w Finlandii i roku 2000 w Norwegii w pełni „wygenerowano” dane spisu ludności z danych rejestrów administracyjnych.

⁸ Powszechny elektroniczny system ewidencji ludności PESEL, rejestr ludności prowadzony w Polsce od 1973 r.

⁹ W przyjazny sposób idea „spisu wirtualnego” prezentowana jest w Internecie: <https://www.youtube.com/watch?v=SLpDkcyenf0>

Warto zauważyć, że ostatni spis metodą tradycyjną przeprowadzono w Holandii w 1971 roku. W roku 1981 spis został przesunięty z powodu ochrony prywatności, a następnie odwołany. W roku 1991 ponownie protesty ludności i niechęć do udziału w spisie spowodowały odstąpienie od projektu. W następstwie tej sytuacji uchylono zapis prawny, że spis powinien być przeprowadzony co 10 lat, a Statistics Netherlands zaproponował nowe rozwiązanie znane jako „spis wirtualny” [Baffour i in., 2013; Gerards, 2012; SN, 2011].

Spis złożony, łączący źródła administracyjne z badaniami próbkowymi [UNECE, 2006]

Spisowe dane o osobach i gospodarstwach domowych mogą być gromadzone poprzez łączne zastosowanie różnych metod: przeprowadzenie badania pełnego (spisywanie tradycyjne), wykorzystanie danych jednego lub kilku badań reprezentacyjnych, z danymi rejestrów administracyjnych lub innych pozastatystycznych źródeł. Dane z rejestrów są wykorzystywane nie tylko jako operat losowania czy dla wsparcia działań operacyjnych w terenie, ale również bezpośrednio jako źródło danych dla niektórych informacji spisowych. W niektórych przypadkach dane rejestrowe są wykorzystywane do wstępnego wypełniania kwestionariuszy, które są weryfikowane lub poprawione podczas gromadzenia danych. W Polsce spis ludności w 2011 roku przeprowadzono metodą złożoną.

Nawiązując do źródeł danych wykorzystywanych w spisach ludności, obok rezultatów badania i danych administracyjnych, podkreślić trzeba coraz częściej wykorzystywane źródła pozaadministracyjne. Są to dane takich systemów informacyjnych, jak np. dotyczących odbiorców energii elektrycznej, wody, systemów informacyjnych dostawców usług telekomunikacyjnych zawierających wykaz odbiorców oraz inne, szczególnie zasoby danych tworzone w ramach działalności przedsiębiorców.

W większości państw, które posiadają rejestr ludności, jest on wykorzystywany w trakcie spisu. Z badania przeprowadzonego przez M. Poulain i A. Herm [2013], w 2010 roku wynika, że wśród objętych badaniem 30 państw europejskich, cztery nie posiadały rejestru ludności. Są to Wielka Brytania, Portugalia, Irlandia i Francja. W kolejnych trzech państwach (Cypr, Grecja i Malta), nie było rejestru centralnego, ale na poziomie lokalnym funkcjonowały bazy danych zawierające dane dotyczące pojedynczych osób.

Wyniki badań opublikowanych przez Europejską Komisję Ekonomiczną ONZ dotyczących źródeł informacji wykorzystywanych w spisach 2010 roku [UN, 2013b] wskazują, że blisko 16% państw zastosowało inne aniżeli tradycyjne badania w terenie. Rozkład terytorialny państw według wykorzystywanych w spisie źródeł informacji nie był równomierny. Wśród państw europejskich ponad 43% wykorzystało

głównie administracyjne źródła informacji¹⁰. Podobne podejście zastosowało 10% państw Azji i 25% Ameryki Północnej. W Afryce i Ameryce Południowej wszystkie państwa pozostały przy tradycyjnej metodzie przeprowadzania spisu.

Tabela 1. Państwa* według metody przeprowadzania spisu ludności w latach 1985–2014

Liczba państw według metody przeprowadzenia spisu ludności	Czas przeprowadzania spisu		
	Runda 1990 (1985–1994)	Runda 2000 (1995–2004)	Runda 2010 (2005–2014)
Liczba państw, które nie przeprowadziły spisu**	31 (13%)	34 (14%)	12 (5%)
Liczba państw, które przeprowadziły spis	208 (86%)	207 (86%)	227 (94%)
Metoda tradycyjna	200 (83%)	194 (81%)	188 (78%)
Metoda alternatywna	8 (3%)	13 (5%)	39 (16%)
Inna metoda	2 (1%)	0 (-)	2 (1%)
Razem	241	241	241

Uwaga:

* Pojęcie „country” użyte w publikacji obejmuje suwerenne państwa, terytoria zależne i inne obszary ujęte w Kodeksie Handlu oraz Lokalizacji Transportowych Organizacji Narodów Zjednoczonych (Code for Trade and Transport Locations). Wydany przez Głównego Geodetę Kraju *Urzędowy wykaz nazw państw i terytoriów niesamodzielnych* obejmuje 195 państw uznawanych przez Rzeczpospolitą Polską (tj. 193 państwa członkowskie Organizacji Narodów Zjednoczonych, Kosowo i Watykan) i 69 terytoriów niesamodzielnych [Zych, Czerny, 2015].

** Pozycja ta uwzględnia państwa, które zgromadziły dane z rejestrów, ale formalnie nie wygenerowały spisu ludności.

Źródło: Kukutai, Thompson i McMillan [2015, s. 12].

W Europie spis oparty na rejestrach obok państw skandynawskich Danii, Finlandii, Norwegii i Szwecji, przeprowadziły Austria, Belgia, Słowenia, Andora. Natomiast na metodę mieszaną zdecydowały się: Czechy, Estonia, Hiszpania, Holandia, Islandia, Izrael, Lichtenstein, Litwa, Łotwa, Niemcy, Polska, Szwajcaria, Turcja i Włochy.

4. Wybrane powody transformacji spisów

Określenie wyzwań i zagrożeń związanych z wykorzystaniem przez statystykę publiczną wielu dostępnych źródeł informacji jest zadaniem w zasadzie niemożliwym. Wynika to zarówno z powodu wielości źródeł, ich różnorodnego, jak i nie-

¹⁰ Jako główne źródło informacji przyjęto to, które było podstawą szacunku łącznej liczby ludności.

ustrukturyzowanego charakteru. Dlatego poniższe rozważania są jedynie kolejnym głosem w dyskusji, różniącym się od innych prób tym, że staramy się na każdą z omawianych kwestii spojrzeć z co najmniej dwóch stron, dostrzegając zarówno wyzwania, jak i zagrożenia wynikające z zastosowania alternatywnych źródeł informacji (tab. 2).

Tabela 2. Wyzwania i zagrożenia alternatywnych metod przeprowadzania spisu ludności

Kryterium	Wyzwania	Zagrożenia
1. Dostępność różnych źródeł danych	<ul style="list-style-type: none"> – Wiek XX wiekiem informacji, wiek XXI wiekiem informacji generowanej samoczynnie, niezależnie, nieświadomie (bez wiedzy użytkownika) – Zagrożenie traktowania danych pozyskanych od respondenta w trakcie wywiadu jako bezwzględnie wiarygodnych 	<ul style="list-style-type: none"> – Wtórny charakter informacji – Zagrożenie traktowania danych administracyjnych jako bezwzględnie wiarygodnych – bo urzędowych – Rozproszenie źródeł i różni gestorzy – Niektóre źródła mogą być niestabilne lub obserwowane od określonego czasu (np. POLon)
2. Aktualność	<ul style="list-style-type: none"> – Możliwa większa częstotliwość publikacji danych spisu niż raz na 10 lat – Dane rejestrów administracyjnych i systemów automatycznej rejestracji na bieżąco aktualizowane 	<ul style="list-style-type: none"> – Konieczność synchronizacji danych różnych rejestrów na tzw. moment krytyczny spisu – Niebezpieczeństwo opóźnień – Trudność w wykorzystaniu danych spisu w celu realizacji bieżącej polityki społecznej
3. Pomiar mobilnej populacji	<ul style="list-style-type: none"> – Określenie ludności rezydującej – Możliwość pomiaru dojazdów do pracy, podróży służbowych i prywatnych 	<ul style="list-style-type: none"> – Trudność właściwej definicji populacji – Trudność poprawnej identyfikacji jednostek
4. Zakres informacji	<ul style="list-style-type: none"> – Dostęp tylko do tych informacji, które są zawarte w rejestrze – Dostęp do informacji dla populacji celu określonej przez obowiązujące regulacje prawne, dla której tworzony jest rejestr (por. plik JPK) – Możliwość pozyskania informacji o dochodach – rejestr POLTAX 	<ul style="list-style-type: none"> – Brak jednostek statystycznych (np. gospodarstw domowych, rodzin) – Brak możliwości pomiaru cech nieujętych w rejestrach administracyjnych – Brak możliwości pomiaru cech tzw. drażliwych (wyznanie, niepełnosprawność, przynależność etniczna, język ojczysty)

Tabela 2. (cd.)

Kryterium	Wyzwania	Zagrożenia
5. Poufność informacji	<ul style="list-style-type: none"> – Regulacje prawne dotyczące wykorzystania informacji osobowych i ich udostępniania statystyce publicznej – Konieczność zabezpieczenia numerów identyfikacji osobowej typu PESEL, NIP itp. w numery identyfikacji statystycznej uniemożliwiającej identyfikację konkretnych osób – Jednostronny obieg informacji <ul style="list-style-type: none"> – przekazywanych przez gestorów organom statystyki publicznej – po uzyskaniu informacji niemożność kontaktu z gestorem w celu wprowadzenia korekt 	<ul style="list-style-type: none"> – Europejski Kodeks Statystyczny określa zasady upowszechniania informacji z zachowaniem tajemnicy statystycznej – Niebezpieczeństwo wykorzystania informacji niezgodnie z intencją osób je generujących – Niebezpieczeństwo wykorzystania informacji niezgodnie z intencją prawodawcy – W państwach o przeszłości totalitarnej szczególnie trudne uzyskanie aprobaty społecznej dla wykorzystania danych administracyjnych w celach statystycznych
6. Zakres przestrzenny	<ul style="list-style-type: none"> – Dzięki zastosowaniu nowoczesnych technik telekomunikacyjnych istnieje możliwość „przypisania” każdej informacji do współrzędnych geograficznych – W przypadku populacji mobilnych istnieje możliwość śledzenia przemieszczania się ludności – Dane rejestrów administracyjnych oraz systemów informacji o klientach, odbiorcach energii, wody, gazu – duża szczegółowość terytorialna 	<ul style="list-style-type: none"> – Szczegółowy zakres przestrzenny informacji możliwy do określenia tylko w wyniku badania pełnego <ul style="list-style-type: none"> – np. spisu organizowanego raz na 10 lat – Ograniczona porównywalność międzynarodowa wynikająca ze stosowania danych administracyjnych określonych przez regulacje prawne danego państwa – Ograniczona porównywalność w czasie wynikająca ze zmian obowiązujących regulacji prawnych w danym państwie <ul style="list-style-type: none"> – np. regulacje dotyczące osób niepełnosprawnych
7. Wiarygodność	<ul style="list-style-type: none"> – Dane rejestrów administracyjnych oraz źródeł pozastatystycznych mają charakter określony w wyniku regulacji prawnych – Dane administracyjne mogą nie odzwierciedlać stanu faktycznego (np. miejsce zamieszkania/zameldowania) 	<ul style="list-style-type: none"> – Dane spisów ludności i badań reprezentacyjnych mają charakter deklaracyjny – Duży odsetek odmów udziału w badaniach (może związany z badaną cechą) – Błąd respondenta grupowego

Kryterium	Wyzwania	Zagrożenia
8. Jakość danych	<ul style="list-style-type: none"> – Nielosowy charakter próby – Autoselektywność – Rozwój metodologii weryfikacji jakości danych ze źródeł pozastatystycznych – Wbudowany system jakości (spójności) – Weryfikacja danych – zasada redundancji – Metainformacje i Paradane <ul style="list-style-type: none"> – konieczność precyzyjnej informacji o sposobie generowania zmiennych 	<ul style="list-style-type: none"> – Informacja złej jakości wypiera informację dobrej jakości – Reprezentatywność danych to losowy charakter próby o odpowiedniej liczebności – <i>Big data</i> – niebezpieczeństwo „dużej próby” nieupoważniającej do uogólnień – Niebezpieczeństwo jednakowego traktowania danych wiarygodnych i obciążonych – Błędy pokrycia, braki danych – Błędy pomiaru (różne definicje)
9. Stymulacja rozwoju teorii i metod badań statystycznych	<ul style="list-style-type: none"> – Integracja statystyczna (łączenie probabilistyczne) – Wielokrotne ważenie, kalibracja, imputacja – Estymacja dla małych domen – Parentyzacja – Rekonstrukcja rodzin 	<ul style="list-style-type: none"> – Brak odpowiedniej teorii badania opartego na danych administracyjnych – Redefinicja podstawowych pojęć statystycznych takich jak populacja, jednostka, cecha – Integracja stochastyczna – błędy łączenia
10. Koszt	<ul style="list-style-type: none"> – Informacja wtórna jest tańsza od pierwotnej – Koszt inwestycji w urządzenia nowoczesnej techniki komunikacyjnej, IoT, oprogramowania itp. – Koszt wysoko kwalifikowanej kadry 	<ul style="list-style-type: none"> – Informacja jest towarem – Wtórna informacja też kosztuje – Kosztuje także „dostosowanie” informacji na potrzeby badania

Źródło: opracowanie własne.

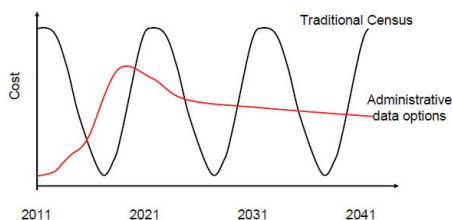
Wśród zasadniczych zalet danych ze źródeł administracyjnych wymienia się przede wszystkim to, że dane te są dostępne w zasadzie od razu bez dodatkowych nakładów finansowych. Szczególnie mocno podkreśla się aktualność danych administracyjnych, gdyż są one na bieżąco aktualizowane i jest tylko kwestią techniczną zaplanowanie odpowiednio częstego raportowania, które wręcz umożliwiłoby przygotowanie zestawień danych „spisowych” nawet z roczną częstotliwością.

Wyniki spisu tradycyjnego publikowane są średnio z około 2 letnim opóźnieniem i bardzo szybko tracą aktualność [Calder, Teague, 2012, s. 5; Vega, Argüeso, 2016]. Tymczasem zgodnie z obowiązującymi zasadami wyniki spisu są uznawane za referencyjne w okresie do kolejnego spisu – czyli przez kolejne 10 lat. Dotychczas

obowiązujące standardy przewidują korektę danych wstecz, zgodnie z informacjami otrzymanymi w bieżącym spisie.

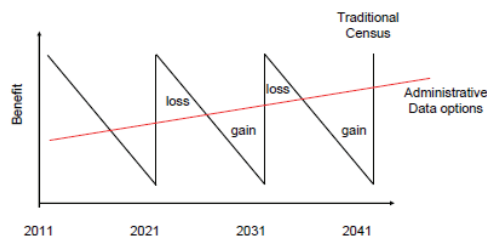
Z powodu opóźnień aktualność tradycyjnych danych spisowych jest dyskusyjna nawet w chwili publikacji wyników. Natomiast dane administracyjne charakteryzuje stała aktualność. Podejście bazujące na rejestrach może zapewnić częstsze aktualizacje, co pozwala na redukcję spadku dokładności w okresie międzyspisowym. Potencjalny zysk na dokładności szacunków spisowych w ujęciu tradycyjnym i opartym na danych rejestrów administracyjnych autorzy programu Beyond 2011¹¹ przedstawili w ujęciu dziesięcioletniego cyklu badania spisowego (rys. 2 i 3). Na podstawie przeprowadzonych analiz Calder i Teague [2012] przedstawili potencjalne profile kosztów i aktualności danych administracyjnych wobec podejścia spisu tradycyjnego.

Rysunek 2. Potencjalne profile kosztów spisu tradycyjnego i wykorzystującego dane administracyjne



Źródło: Calder i Teague [2012, s. 4].

Rysunek 3. Potencjalne profile aktualności danych tradycyjnego spisu i bazującego na rejestrach



Źródło: Calder i Teague [2012, s. 5].

Przygotowując badanie oparte na rejestrach, które z natury są na bieżąco aktualizowane, można tak zaplanować i zautomatyzować szereg zadań, by skrócić czas potrzebny do udostępnienia danych i publikowania ich stosunkowo bliżej daty referencyjnej. Uwzględnienie możliwości automatyzacji procedur procesu przekazywania danych oznacza ich większą aktualność i skrócenie czasu od momentu referencyjnego. Dodatkowo wykorzystanie danych ze źródeł administracyjnych umożliwia podejście zapewniające publikowanie danych o ludności z większą częstotliwością – np. roczną.

Szczególnie szybka jest utrata aktualności danych spisowych w populacji doświadczającej gwałtownych zmian procesów demograficznych, w tym intensywnych migracji przyspieszających zaawansowanie starzenia czy obniżenie dzietności.

¹¹ Beyond 2011 to projekt realizowany w Wielkiej Brytanii po spisie 2011 roku w celu wypracowania rekomendacji dla kolejnego spisu.

Aktualne szacunki stanu i struktury ludności są w takich sytuacjach niezbędne do realizacji bieżącej polityki społecznej.

Wśród często podnoszonych ograniczeń danych administracyjnych wskazuje się konieczność korzystania z tych informacji, jakie zostały zebrane w rejestrze, a nie koniecznie tych, jakie są oczekiwane. Czy dane rejestru administracyjnego zawierają minimum spisowe, czy spełniają potrzeby osób fizycznych, ministerstw i urzędów centralnych, przedsiębiorców, naukowców itp.

Porównując tematy uwzględniane w polskich spisach ludności od 1921 roku, należy zauważyć wyraźne i systematyczne poszerzanie zakresu zagadnień [Krywulł-Albańska, 2012]. Zmiany rzeczywistości gospodarczej oraz sytuacji społecznej powodowały zmiany tematów: eliminowanie jednych i dodawanie innych. Na tle wcześniejszych spisów przeprowadzonych metodą tradycyjną NSP 2011, jest najbardziej wyczerpujący.

Wśród podstawowych ograniczeń spisów opartych na rejestrach często podaje się argument niemożności pozyskania wszystkich potrzebnych informacji, gdyż nie są one dostępne w źródłach administracyjnych. Jednym z najczęściej przytaczanych jest brak informacji dotyczących rodzin. Pojawia się wówczas potrzeba rekonstrukcji rodzin, która prowadzi do opracowywania metod tzw. parentyzacji¹². Do zmiennych, jakie można pozyskać tylko w badaniach specjalnych, zalicza się np.: stan cywilny faktyczny – związki nieformalne, rodzina – typ biologiczny, wykształcenie, narodowość, język ojczysty, wyznanie. W znacznej mierze są to tzw. tematy uzupełniające.

Spis korzystający z większej liczby źródeł spełnia kryterium przydatności. Podkreślmy też pełną zgodność z zasadami spisu: powszechność, czyli objęcie całej populacji oraz indywidualne spisywanie – informacji dla każdej osoby niezależnie. Jest to niewiarygodnie pozytywna cecha. Dzięki pozyskaniu danych z rejestru bezpośrednio o każdej indywidualnej osobie niweluje się wielkość błędu danych z drugiej ręki. W tym miejscu trzeba też nawiązać do charakteru spisywanych informacji, w tym ludności i jej różnych charakterystyk. Pomijamy tu niezwykle istotne z perspektywy szacunku demograficznego kategorie ludności faktycznej i rezydującej¹³, zwrócimy jedynie uwagę na jakość danych, wynikającą z ich deklaratywnego charakteru.

Informacje uzyskiwane za pośrednictwem rachmistrza spisowego mają charakter deklaratywny i nie podlegają żadnej weryfikacji. W większości przypadków dane pozyskiwane są od jednej bądź kilku osób obecnych w mieszkaniu, a nie bezpośrednio od każdego z respondentów. W efekcie pojawia się tzw. obciążenie „drugiej ręki”. Nie jest znany ani prawny, ani faktyczny stan cywilny badanych osób, podobnie

¹² Mogą one czerpać z metod demografii historycznej, w której rekonstrukcję rodzin przeprowadzano wykorzystując różne źródła, np. na podstawie ksiąg metrykalnych [Kukło, 2009]. Ciekawą alternatywą może być metoda własnych dzieci (Own Children Method), która umożliwia szacunek cząstkowych współczynników płodności [Feeney, 1975; Krapf, Kreyenfeld, 2015].

¹³ Dyskusja tych pojęć była przedmiotem różnych opracowań, m.in. M. Kupiszewskiego [Bijak i in., 2007; Kupiszewski, Bijak, 2012] czy J. Paradysza [2007; 2010].

miejsce zamieszkania, poziom wykształcenia czy status na rynku pracy. Z kolei informacje dotyczące kwestii drażliwych, z zasady mogą być gromadzone jedynie za zgodą zainteresowanych i według ich deklaracji, przy czym odpowiedź nie jest obowiązkowa¹⁴. Pozyskiwanie takich danych za pośrednictwem osób trzecich narusza prawo wolności sumienia i wyznania oraz nieujawniania informacji. Na przykład informacje zebrane o osobach niepełnosprawnych mogą być uzyskane wyłącznie bezpośrednio od zainteresowanych i bez weryfikacji orzeczeń o niepełnosprawności. Podobnie pytania dotyczące kwestii etnicznych, wyznania, narodowości czy języka ojczystego są sprawą osobistą każdego respondenta i mogą być pozyskane jedynie w sposób bezpośredni za zgodą i według indywidualnych deklaracji.

Koszty spisu rosną z dekady na dekadę w niemal geometrycznym tempie. Jeśli w USA w 1960 roku całkowite, realne koszty spisu szacowano na 995 mln USD, to 50 lat później wynosiły one prawie 15 razy więcej – 14,7 mld USD (tab. 3). Podobnie sytuacja kształtuje się w Wielkiej Brytanii, gdzie w latach 1991–2011 koszty spisu wzrosły ze 140 mln GBP do 481,7 mln GBP [Brown i in., 2010].

Koszty spisu są potęgowane w wyniku mobilizacji co dekadę dużej rzeszy ankietowanych zatrudnionych w celu dostarczenia i zebrania formularzy ankietowych. Według wyliczeń ekspertów ONZ, w spisach 2010 roku średni koszt w przeliczeniu na mieszkańca w państwach europejskich wyniósł 8,84 USD, podczas gdy w USA koszt ten sięgał 48,9 USD. Generalnie najwyższe są koszty w przypadku spisów tradycyjnych, natomiast spisy oparte na rejestrach bądź przeprowadzone metoda mieszaną są zdecydowanie tańsze. W takich państwach jak Austria, Belgia, Dania, Finlandia, Holandia, Norwegia, Słowenia czy Szwajcaria, koszt spisu na osobę oszacowano na poziomie 1,3 USD. Najtańszy okazał się spis duński – z kosztem na osobę 0,3 USD [Coleman, 2013]. Oczywiście porównanie kosztów spisu tradycyjnego i bazującego na danych administracyjnych nie jest proste, gdyż wymaga uwzględnienia kosztów utrzymania rejestru i odpowiednich prac metodycznych.

Komentując wysokie koszty spisów tradycyjnych, warto spojrzeć na nie z jeszcze jednej, innej perspektywy. Mianowicie D. Coleman [2013] podkreśla marnotrawstwo kapitału ludzkiego. W jego opinii większy byłby pożytek dla statystyki publicznej z trwałego zatrudnienia nielicznej grupy osób regularnie zdobywających doświadczenie i wiedzę ekspercką, zamiast intensywnego szkolenia ogromnej rzeszy osób, które zatrudnione są przez krótki okres realizacji spisu.

Zmiana metody przeprowadzania spisu spowodowała, że łącznie koszt NSP 2011 był o ponad 100 mln PLN niższy aniżeli w 2002 roku. Przeprowadzenie spisu metodą mieszaną w Polsce w 2011 roku przełożyło się na szereg oszczędności (tab. 4). Za najważniejsze wskazuje się wykorzystanie narzędzi komunikacji elektronicznej,

¹⁴ Zgodnie z zapisami ustawy z dnia 4 marca 2010 r. o narodowym spisie powszechnym ludności i mieszkań w 2011 r. udzielanie odpowiedzi na pytania dotyczące niepełnosprawności odbywało się na zasadzie dobrowolności.

a tym samym ograniczenie kosztów przeprowadzania spisu w terenie, druku formularzy spisowych oraz 10-krotne obniżenie liczby rachmistrzów spisowych, których wynagrodzenie jest jednym z najkosztowniejszych elementów budżetu [Dygaszewicz, 2014, s. 353]. Duże oszczędności poczyniono przede wszystkim dzięki wprowadzeniu nowoczesnych narzędzi telekomunikacyjnych.

Tabela 3. Koszty spisu ludności w USA (w latach 1960–2010)

Spis	Koszty realne w 2009 r.		Koszty nominalne
	pełny koszt spisu (mln USD)	koszt spisu (USD/gosp.dom.)	pełny koszt spisu (mln USD)
1960	995	16,89	120
1970	1206	17,06	231
1980	2947	32,71	1136
1990	4424	42,54	2600
2000	8060	69,54	6600
2010	14700	114,93	14700

Źródło: Brown i in. [2010].

Tabela 4. Struktura wydatków NSP 2002 i NSP 2011

Grupa wydatków	NSP 2002 (koszt w PLN)	(%)	NSP 2011 (koszt w PLN)	(%)
Przygotowanie operatów spisowych, w tym map	1 564 150,5	0,3	1 574 545,2	0,4
Szkolenia wyspecjalizowanego personelu	407 983,9	0,1	1 683 764,9	0,4
Przeprowadzenie badań pilotażowych i spisu – wydatki osobowe i bezosobowe z funduszu wynagrodzeń	402 877 214,9	81,1	220 236 302,8	55,7
Formularze spisowe i inne materiały (opracowanie, druk)	20 214 467,1	4,1	166 669,8	0,04
Modernizacja infrastruktury, inwestycje, prace projektowo-programowe	53 118 864,7	10,7	159 621 327,9	40,4
Popularyzacja spisów	1 245 041,8	0,25	2 903 737,7	0,7
Opracowanie, analiza i udostępnianie	828 283,4	0,17	45 500,0	0,01
Inne prace i wydatki	16 634 841,7	3,3	9 052 152,9	2,3
Razem	496 890 847,8	100	395 284 001,2	100

Źródło: Dygaszewicz [2014, s. 361].

Porównując lokatę Polski na tle innych państw należy podkreślić, że spis tradycyjny NSP 2002 był w przeliczeniu na mieszkańca o ponad 80% droższy w ujęciu nominalnym, czy o 60% według parytetu PPPs, w porównaniu z medianą kosztów spisów tradycyjnych oszacowanych przez ONZ na rundy roku 2000 (tab. 5). Z kolei porównania dla ostatniej rundy spisów 2010 roku wskazują, że koszty per capita podejścia mieszanego są o ponad 40% niższe aniżeli w przypadku spisu tradycyjnego. Dodatkowo w Polsce koszty te były o ponad 10% niższe od mediany kosztów dla metody mieszanej na świecie. Polska w wielu zestawieniach zyskała bardzo pozytywne oceny zwłaszcza w zakresie wdrożenia nowatorskich rozwiązań metodycznych, organizacyjnych i logistycznych, w tym wykorzystanie danych z rejestrów oraz zastosowanie tańszych metod zbierania danych dla tych podmiotów, dla których nie udało się wykorzystać danych rejestrowych (CAII, CATI).

Tabela 5. Mediana kosztów w zależności od metody przeprowadzania spisu 2010

Metoda	Łączny koszt (mln USD)	Koszt nominalny (USD/per capita)	Koszt wg parytetu (USD PPPs/per capita)
Spis tradycyjny 2000		2,1	5,2
NSP 2002 – spis metodą tradycyjną		3,8	8,5
Łącznie spis 2010	18,7	3,94	6,01
Spis metodą tradycyjną 2010	34,1	5,57	7,31
Spis metodą mieszaną 2010	18,7	3,94	6,01
Spis oparty na rejestrach 2010	1,9	0,24	0,18
NSP 2011 – spis metodą mieszaną		3,54	5,59

Źródło: Dygaszewicz [2014, s. 360; UNECE, 2013].

5. Perspektywy spisów ludności

W końcu sierpnia 2014 roku odbyło się spotkanie Dyrektorów Statystyki Społecznej urzędów statystycznych państw Unii Europejskiej (DSS UE). Tematem spotkania był program spisów ludności i mieszkań w państwach UE w kolejnej rundzie – roku 2020 oraz w dalszej perspektywie. Dyskutowano trzy grupy zagadnień: (i) jakość spisów rundy 2010, (ii) sposób przeprowadzenia spisów w 2021 roku (iii) oraz modernizację europejskiej statystyki ludności i mieszkań po roku 2021 (Beyond 2021). Echo tej dyskusji i kolejnych spotkań powtarza wśród statystyków wieść

o końcu spisów. W rzeczywistości sprawa dotyczy nie tyle przysłowiowego „końca spisów”, co modernizacji całego systemu statystyki demograficznej i społecznej [EU, 2014; 2016].

Przed wszystkim podczas spotkania DSS UE ustalono, że w kolejnej rundzie spisy zostaną przeprowadzone na podstawie dotychczas obowiązujących w UE regulacji prawnych [UE, 2008]. Przyjęto jednak, że konieczne są nowe regulacje dotyczące definicji, przekrojów, klasyfikacji, tabel krzyżowych (cross-tabulacji), metadanych i jakości raportowania. Podkreślono, że kwestie zasadniczych zmian dotyczących podejścia, stosowanych metod i wykorzystywanych źródeł danych podejmowane są na poziomie krajowym. Eurostat z kolei kontynuuje prace w zakresie modernizacji systemu statystyki społecznej. Zwrócono uwagę na potrzebę równoważenia istotnych i zmieniających się potrzeb związanych ze spisem w państwach UE w kontekście aspektów praktycznych i obciążeń wynikających z tworzenia statystyk. Jako istotny podjęto temat integracji danych statystycznych o ludności i danych przestrzennych w celu dostarczenia danych demograficznych według miejsca zamieszkania w przekroju terytorialnym dla siatki gridowej (1 km², geo-grid: dyrektywa INSPIRE). Na marginesie warto zaznaczyć, że takie badania dotyczące rozwoju demograficznego Poznania w kontekście starzenia się ludności prowadzone są w Ośrodku Statystyki Miast Urzędu Statystycznego w Poznaniu¹⁵.

Wykorzystanie wielu źródeł w spisie nowej generacji wynika z ich dostępności i pozytywnej weryfikacji jakości. Nie byłoby to możliwe, gdyby nie rozwój nowoczesnych technologii i elektroniczna forma tych źródeł. Ale uwzględnienie wielości źródeł wymaga też zastosowania odpowiedniej metodyki. Statystycy muszą odpowiedzieć na wiele problemów związanych z integracją źródeł i oceną ich wiarygodności, co motywuje bezwzględnie konieczny, intensywny rozwój metod statystycznych. W większości zgodnej opinii wielu statystyków, to rozwój metod statystycznych nie nadąży za rozwojem technologii informatycznych [Szreder, 2009; 2017]. Nie mniej prace metodyczne, także prowadzone w Polsce są coraz bardziej zaawansowane. Odwołując się jedynie do prac prowadzonych w ośrodku poznańskim [Beręsewicz i in., 2018], warto zwrócić uwagę na kwestie integracji danych z różnych źródeł, kalibracji, imputacji, *repeated weighting* i innych metod korekty wag oraz metody estymacji określane mianem statystyki małych obszarów.

Ze względu na rosnące oczekiwania i potrzeby odbiorców informacji statystycznej, doskonałe technologie informatyczne oraz zmieniające się wyzwania statystyki publicznej, wynikające z jej nowej roli w XXI wieku, coraz więcej uwagi przywiązuje się do alternatywnych źródeł informacji w postaci tzw. *big data* (niestatystycznych źródeł danych) [Beręsewicz, Szymkowiak, 2015].

¹⁵ Można tutaj wymienić następujące prace: Basarbowicz i in. [2015], Filas-Przybył [2017].

Wykorzystanie *big data* przez długi czas odbywało się poza głównym nurtem statystyki publicznej, w której preferuje się w dalszym ciągu informacje pochodzące ze spisów, badań reprezentacyjnych czy rejestrów administracyjnych. Ze względu jednak na rosnące znaczenie tego nowego źródła informacji niektóre urzędy statystyczne – na przykład w Holandii, zdecydowały się przetestować ich użyteczność na potrzeby statystyki oficjalnej. Jest to zadanie niezwykle trudne, a dotyczy zarówno problematycznych kwestii prawnych, dostępności danych i sposobu ich przechowywania, zagadnień związanych z koniecznością zachowania tajemnicy statystycznej czy, co ważniejsze, problemów dotyczących obciążenia będącego konsekwencją charakteru „masowych” źródeł. Istnieje zatem ogromne zapotrzebowanie na zaadaptowanie danych pochodzących z nowych źródeł (*big data*) na potrzeby i cele statystyki oficjalnej, tak aby zmaksymalizować korzyści i zminimalizować zagrożenia wynikające z ich wykorzystania w statystyce publicznej [Beręsewicz, Szymkowiak, 2015].

Problemy z *big data*, pomimo relatywnie niedługiej popularności tego pojęcia, nie są nowe w statystyce i możemy wskazać podobieństwa do:

- badań opartych na rejestrach (*register surveys*) – wykorzystanie danych niestatystycznych do celów statystycznych,
- badań internetowych (*Internet surveys*) – dane są zbierane przy wykorzystaniu Internetu (w szerokim ujęciu, włączając IoT),
- badań wymagających wyłącznie zgody respondenta (*opt-in surveys*) – ponieważ udział/dobór podlega nielosowemu mechanizmowi,
- badań panelowych (*panel surveys*) – ponieważ obserwujemy jednostki w czasie (wraz z wycieraniem się próby).

Wykorzystanie nowych źródeł danych na potrzeby spisów powszechnych będzie się wiązać z automatycznym zbieraniu informacji bez konieczności ingerencji człowieka. Opiera się to głównie na wykorzystaniu technologii Internetu rzeczy, która umożliwia kontaktowanie się urządzeń między sobą oraz rejestrowanie informacji określonych przez twórców danej technologii/urządzenia.

W tym miejscu możemy sobie wyobrazić, że ludzie, jak i ich nieruchomości lub inne dobra (np. samochody) dokonają samospisu. Wizja ta nie jest bardzo odległa. Już teraz w tym kontekście obserwujemy określone jednostki populacji generalnej (zwykle osoby młodsze, pochodzące z miast i charakteryzujące się wyższymi dochodami), które przekazują informacje o swojej aktywności poprzez korzystanie z różnego typu aplikacji mobilnych czy urządzeń typu smart (np. telewizory czy zegarki). Jako przykład może posłużyć niedawno uruchomiony sklep *Amazon Go*, w którym nie ma obsługi, kamery na podstawie algorytmów głębokich sieci neuronowych rozpoznają zachowania klientów i wybrane produkty, a płatność dokonywana jest automatycznie w momencie opuszczenia sklepu.

Należy jednak zaznaczyć, że pomimo automatyzacji procesów zbierania i przetwarzania danych, jakość zebranej informacji będzie miała kluczowe znaczenie.

Wyzwaniem będzie określenie podstawowych pojęć używanych w statystyce i spisach powszechnych, czyli populacji, jednostki i badanych cech. Definicje wykorzystywane w nowych źródłach rzadko odzwierciedlają istniejące koncepty używane w spisach, a liczba potencjalnych źródeł będzie powodować problemy z jednoznacznym określeniem, kogo i co badamy. W tym miejscu spisy powszechne powinny również dopasowywać się do zmieniającego się otoczenia (m.in. poprzez rozszerzanie zakresu informacyjnego o dane zbierane w sposób automatyczny).

Niemniej, *big data* w całości nie będzie pokrywać badanej populacji [Hand, 2018]. Kluczowe może się stać wykorzystanie bieżących badań reprezentacyjnych, które pozwolą skorygować błąd pokrycia, autoselekcji czy pomiaru. Omawiane powyżej przykłady samospisów, niezależnie od ich wielkości, będą próbami nie-losowymi, obciążonymi selektywnością wyboru, które będą wymagać szczególnej uwagi, zwłaszcza, jeżeli chcemy dostarczać informacji o populacji generalnej na niskich poziomach agregacji przestrzennej.

Bibliografia

- Baffour B., King T., Valente P. [2013], *The Modern Census: Evolution, Examples and Evaluation*, „International Statistical Review”, vol. 81, no. 3, s. 407–425. <https://doi.org/10.1111/insr.12036>
- Basarbowicz K., Dąbrowski A., Filas-Przybył S., Kaźmierczak M., Klimanek T., Kowalewski J., Kruszka K., Pawlikowski D., Stachowiak D., Stawikowska M. [2015], *Identyfikacja obszarów specjalnych wewnątrz miast wojewódzkich oraz na ich obszarach funkcjonalnych uwzględniających sytuację demograficzną i ekonomiczną ich mieszkańców na podstawie analiz przestrzennych z wykorzystaniem Geographic Information System (GIS)*, GUS, Warszawa, Poznań, https://stat.gov.pl/files/gfx/portalinformacyjny/pl/defaultstronaopisowa/5850/1/1/raport_obszary_specjalne_gis.pdf (21.03.2019).
- Beręsewicz M. [2016], *Internet data sources for real estate market statistics*, University of Economics and Business, Poznań, <http://www.wbc.poznan.pl/dlibra/doccontent?id=393454> (21.03.2019).
- Beręsewicz M. [2017], *A Two-Step Procedure to Measure Representativeness of Internet Data Sources*, „International Statistical Review”, vol. 5, s. 1–21.
- Beręsewicz M., Pawlikowski D., Filas-Przybył S., Józefowski T., Klimanek T., Szymkowiak M., Kaźmierczak M. [2018], *Poprawa wykorzystania administracyjnych źródeł danych*, Poznań.
- Beręsewicz M., Szymkowiak M. [2015], *Big Data w statystyce publicznej – nadzieje, osiągnięcia, wyzwania i zagrożenia*, „Ekonometria”, nr 2(48), s. 9–22, <https://doi.org/10.15611>
- Bijak J., Kicing A., Kupiszewski M. [2007], *Studium metodologiczne oszacowania rzeczywistej liczby ludności Warszawy*, „CEFMR Working Paper”, no. 2, Warszawa, http://www.cefmr.pan.pl/docs/cefmr_wp_2007-02.pdf (21.03.2019).
- Brown L.D., Cohen M.L., Cork D.L., Citro C.F. (red.) [2010], *Envisioning the 2020 Census*, National Academies Press, Washington, <https://doi.org/10.17226/12865>
- Calder A., Teague A. [2012], *Beyond 2011 – the future of population statistics?* UNECE Seminar on New Frontiers for Statistical Data Collection, Geneva, s. 1–9.

- Coleman D. [2013], *The Twilight of the Census*, „Population and Development Review”, vol. 38, no. 1, s. 334–351, <https://doi.org/10.1111/j.1728-4457.2013.00568.x>
- Dygaszewicz J. (red.) [2014], *Spisy powszechne*, GUS, Poznań, Warszawa.
- EU [2014], *Population and housing censuses*, http://www.knbs.or.ke/index.php?option=com_phocadownload&view=category&id=100:population-and-housing-census&Itemid=1176 (21.03.2019).
- EU [2016], *EU Census post-2021 programme*, Luxembourg.
- Feeney G. [1975], *The Own-Children Method of Estimating Age-Specific Fertility Rates*, Kula Lumpur.
- Filas-Przybył S. [2017], *Wpływ migracji na lokalizację klastrów o wysokim poziomie starości demograficznej w przestrzeni miasta Poznania, Szczecin*.
- Gerards E. [2012], *Dutch Virtual Census*, Voorburg, <https://unstats.un.org/unsd/.../meetings/.../s04-5-1-Netherlands.ppt> (21.03.2019).
- Hand D.J. [2018], *Statistical challenges of administrative and transaction data*, „Journal of the Royal Statistical Society: Series A (Statistics in Society)”, vol. 181, no. 3, s. 555–605, <https://doi.org/10.1111/rssa.12315>
- Kędelski M., Paradysz J. [2006], *Demografia*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.
- Kordos J. [2017], *The challenges of the population census round of 2020. Outline of the methods of quality assessment of population census data*, „Statistics in Transition – New Series”, vol. 18, no. 1, s. 115–138, <http://stat.gov.pl/en/sit-en/issues-and-articles-sit/previous-issues/volume-18-number-1-march-2017/> (21.03.2019).
- Kotowska I.E. [2009], *Spisy powszechne ludności jako źródło informacji o przemianach demograficznych i społecznych*, „Wiadomości Statystyczne”, nr 9, s. 1–15.
- Krapf S., Kreyenfeld M. [2015], *Fertility assessment with the own-children method: A validation with data from the German Mikrozensus*, „MPIDR Technical Report TR-003”, vol. 49, s. 1–13.
- Krywułt-Albańska M. [2012], *Spis powszechny jako źródło informacji o ludności*, „Studia Socjologiczne”, nr 4(207), s. 87–107.
- Kuklo C. [2009], *Demografia Rzeczypospolitej przedrozbiorowej*, Wydawnictwo DIG, Warszawa.
- Kukutai T., Thompson V., McMillan R. [2015], *Whither the census? Continuity and change in census methodologies worldwide*, „Journal of Population Research”, vol. 32, no. 1, s. 3–22, <https://doi.org/10.1007/s12546-014-9139-z>
- Kupiszewski M., Bijak J. [2012], *Ocena założeń aktualnej prognozy demograficznej GUS na lata 2003–2030. Ocena wpływu migracji zagranicznej na bilans demograficzny Polski. Zmiany i problemy demograficzne a rozwój przestrzenny*, t. IV, Warszawa.
- Kurkiewicz J. (red.). [2010], *Procesy demograficzne i metody ich analizy*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków.
- Lesthaeghe R., van de Kaa D.J. [1986], *Twee demografische transitities?* w: van de Kaa D.J., Lesthaeghe R. (red.), *Bevolking: groei en krimp*, Van Loghum Slaterus, Deventer, s. 9–24.
- Nordbotten S. [2010], *The Use of Administrative Data in Official Statistics – Past, Present, and Future – With Special Reference to the Nordic Countries*, w: Carlson M., Nyquist H., Villani M. (red.), *Official Statistics: Methodology and Applications in Honour of Daniel Thorburn*, Statistics Sweden, Stockholm, s. 205–233, http://www.nordbotten.com/articles/Adm_data.pdf (21.03.2019).
- Okólski M., Fihel A. [2012], *Demografia*, Wydawnictwo Naukowe Scholar, Warszawa.
- Paradysz J. [2007], *Rejestry administracyjne jako źródło zasilania w statystyce regionalnej*, w: Paradysz J. (red.), *Statystyka regionalna w jednoczącej się Europie*, Internetowa Oficyna Wydawnicza Centrum Statystyki Regionalnej, Poznań, s. 193–206.

- Paradysz J. [2010], *Konieczność estymacji pośredniej na użytek spisów powszechnych*, w: Gołata E. (red.), *Pomiar i informacja w gospodarce*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań.
- Poulain M., Herm A. [2013], *Central Population Registers as a Source of Demographic Statistics in Europe*, „Population”, vol. 68, no. 2, s. 183–212, <https://doi.org/10.3917/popu.1302.0215>
- SN [2011], *Dutch Census 2011*, Statistics Netherlands.
- Szreder M. [2009], *O relacjach między zaufaniem do statystyki a jakością badań*, w: Biolik J. (red.), *Dylematy ekonometrii: księga pamiątkowa dla uczczenia siedemdziesiątych urodzin Profesora Andrzeja Stanisława Barczaka Doktora honoris causa*, Wydawnictwo Akademii Ekonomicznej im. Karola Adamieckiego, Katowice, s. 211–222.
- Szreder M. [2017], *Nowe źródła informacji i ich wykorzystywanie w podejmowaniu decyzji*, „Wiadomości Statystyczne”, nr 7(674), s. 5–17.
- Tønder J.-K. [2008], *The Register-based Statistical System Preconditions and Processes*, International Association for Official Statistics Conference Shanghai October 14–18, Shanghai, www.iaos-isi.org/papers/CS_1_2_Tonder.doc (21.03.2019).
- UE [2008], Rozporządzenie Parlamentu Europejskiego i Rady nr 763/2008 z dnia 9 lipca 2008 r. w sprawie spisów powszechnych ludności i mieszkań, 7, <http://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX%3A32008R0763> (21.03.2019).
- UN [2013a], *Overview of national experiences for population and housing censuses of the 2010 round*, „Working Paper”, New York.
- UN [2013b], *Report on the Results of a Survey on Census Methods used by Countries in the 2010 Census Round*, United Nations.
- UN [2015], *Principles and recommendations for population and housing censuses – Revision 3*, „Statistical Papers Series M”, no. 67/Rev.3, United Nations, New York.
- UNECE [2006], *United Nations Economic Commission for Europe Conference of European Statisticians Recommendations for the 2010 Censuses of*, United Nations, Geneva, http://www.unece.org/fileadmin/DAM/stats/publications/CES_2010_Census_Recommendations_English.pdf (21.03.2019).
- UNECE [2013], *Costs and benefits: Key results from the UNECE Survey on National Census Practices, and first proposals about the CES Recommendations for the 2020 census round*, Geneva, http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2013/census_meeting/20_E_rev_4_.pdf (21.03.2019).
- van de Kaa D.J. [1987], *Europe's second demographic transition*, „Population Bulletin”, vol. 42, no. 1.
- Vega J., Argüeso A. [2016], *Spain 2021. Why will this Census have more quality than the previous one?*, June, s. 1–9.
- Wallgren A., Wallgren B. [2014], *Register-based Statistics. Statistical Methods for Administrative Data*, Second Ed., John Wiley i Sons.
- Zhang L.-C. [2015a], *On Modelling Register Coverage Errors*, „Journal of Official Statistics”, vol. 31, no. 3, s. 381–396, <https://doi.org/10.1515/jos-2015-0023>
- Zhang L.-C. [2015b], *Population Size Estimation Based on Multiple Lists. Uncertainty Analysis for Categorical Data Fusion*.
- Zych M., Czerny A. [2015], *Urzędowy wykaz nazw państw i terytoriów niesamodzielnych*, Główny Urząd Geodezji i Kartografii, Warszawa, http://www.gugik.gov.pl/_data/assets/pdf_file/0005/24935/Urzedowy-wykaz-panstw.pdf (21.03.2019).

The future of statistical research on social and economic needs on the example of the census

Summary: The census is the basic source of information on the condition and structure of the population and its territorial distribution. This study is an essential part of the national and international official statistics system, primarily because it provides population estimates. The census is one of the oldest statistical surveys, but ... is it still needed today? Is official statistics, whose task is to provide a reliable information system, be able to meet these expectations? The chapter attempts to answer the above questions by discussing organizational changes in statistical surveys, methods of conducting and transformation of censuses, and especially threats and future prospects related to the use of non-statistical data sources (eg administrative registers or big data) as alternative methods carrying out a census.

Keywords: official statistics, census, administrative records, big data