

Elżbieta Gołata

Uniwersytet Ekonomiczny w Poznaniu

INTEGRACJA DANYCH Z RÓŻNYCH ŹRÓDEŁ DLA POTRZEB SPISU WIRTUALNEGO – NSP 2011

1. Wstęp

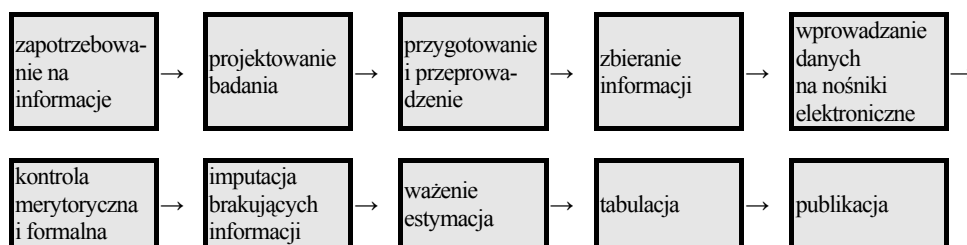
Opracowanie poświęcone jest problemom integracji danych dla potrzeb NSP 2011, którego realizacja zaplanowana jest w formie tzw. spisu wirtualnego. Określenie „spis wirtualny” stosowane jest w literaturze zachodniej w odniesieniu do spisu opartego na rejestrach (por. [Linder 2003]). W obliczu nieuchronnych, zdaniem autorki, zmian w organizacji statystyki publicznej, wynikających z wykorzystania na coraz szerszą skalę źródeł administracyjnych, zdefiniowano problemy związane z integracją rejestrów oraz spójnością danych. Wskazano możliwości ich rozwiązania oraz zidentyfikowano potencjalne zagrożenia i korzyści. Rozważania zilustrowano dyskusją wybranych zagadnień z zakresu badania aktywności ekonomicznej ludności w ramach spisu powszechnego.

2. Idea integracji danych – reorganizacja procesu badania statystycznego

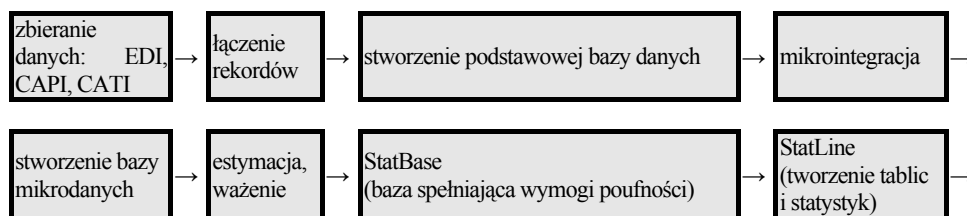
Problem integracji różnych źródeł informacji dotyczy nie tylko spisów ludności. Jest on przedmiotem dyskusji statystyków w wielu krajach, m.in. w Holandii, Danii, Szwecji, Niemczech, Austrii, Włoszech czy w Finlandii. P. Al i B. Bakker [2000] zmiany zachodzące w Statistics Netherland nazywają procesem reorganizacji statystyki społecznej w wyniku utworzenia bazy mikro danych na podstawie rejestrów administracyjnych. Wskazują oni przyczyny zachodzących przemian, takie jak: wzrost popytu na informacje i kosztów badań statystycznych, ale także rozwój informatyki i metod gromadzenia danych. Artykułując zapotrzebowanie na informacje, władze coraz większy nacisk kładą na problemy rozwoju gospodarczego, dystrybucji środków zapewniającej równomierny wzrost dobrobytu, na poprawę warunków życia, walkę z ubóstwem i z wykluczeniem społecznym.

Większość informacji w sferze statystyki społecznej pochodzi z badań specjalnych przeprowadzanych metodą reprezentacyjną. Są one kosztowne i stanowią duże obciążenie respondentów skutkujące coraz większym odsetkiem odmów i braków odpowiedzi¹, a w rezultacie pogorszeniem jakości danych statystycznych². Pod presją polityków domagających się aktualnej, rzetelnej i spójnej informacji przy jednoczesnej redukcji kosztów badań, personelu i błędów wykorzystano rozwój technologii informatycznych oraz istniejące źródła danych. P. van der Laan [2000] przedstawia ideę podejścia zastosowanego w programie Statistics Netherland nazwanym *Sociaal Statistisch Bestand* (SSB). Jako istotę konstrukcji SSB, czyli Bazy Danych Statystyki Społecznej, wskazuje on integrację rejestrów administracyjnych oraz wyników badań reprezentacyjnych na poziomie mikro, podkreślając spójność uzyskanych na tej podstawie statystyk. Podstawę konstrukcji SSB stanowi rejestr ludności, do którego dołączano informacje z innych rejestrów oraz badań specjalnych.

A. Tradycyjna organizacja badań w statystyce społecznej



B. Reorganizacja procesu badań w statystyce społecznej wskutek integracji baz danych



Uwaga: skróty zamieszczone na rysunku mają następujące znaczenie:

EDI (*Electronic Data Interchange*) – system elektronicznej wymiany danych,

CAPI (*Computer Assisted Personal Interview*) – wspomagany komputerowo wywiad osobisty,

CATI (*Computer Assisted Telephone Interview*) – wspomagany komputerowo wywiad telefoniczny.

Rys. 1. Proces badań w statystyce społecznej w ujęciu tradycyjnym i jego reorganizacja

Źródło: opracowanie własne na podstawie [van der Laan 2000 s. 9-10].

¹ W Holandii w latach 1977-1995 odsetek gospodarstw biorących udział w badaniu siły roboczej zmniejszył się z 90 do 60%. W badaniu jakości życia odsetek odpowiedzi spadł z 72% w 1974 r. do 50% w 1995 r.

² Większość braków odpowiedzi (ok. 65%) spowodowana jest odmową udziału w badaniu. Odmowy te mają charakter selektywny, dotyczą przede wszystkim osób samotnych, szczególnie jednoosobowych gospodarstw domowych ludzi młodych, mniejszości etnicznych, mieszkańców miast.

Tradycyjne i aktualne podejście do organizacji badania statystycznego P. van der Laan [2000] przedstawia w formie schematycznej (por. rys. 1). Na podkreślenie zasługuje zmiana kolejności tradycyjnych etapów oraz nowe zadania, takie jak tworzenie wieloźródłowej podstawowej bazy danych jako wynik procesu integracji baz danych. Kolejnym krokiem jest tworzenie spójnej bazy mikrodanych zawierającej informacje po etapie „czyszczenia”, tj. eliminacji (powtarzających się rekordów i zmiennych), imputacji, porządkowania klasyfikacji itp. Spełnienie wymogu poufności danych wymaga agregacji danych do poziomu jednostek statystycznych. Utworzona na tym etapie baza danych StatBase stanowi podstawę konstrukcji wiarygodnych tablic, zestawień i statystyk na potrzeby użytkowników.

W nowym podejściu podkreślić należy przede wszystkim efektywność wykorzystania istniejących źródeł informacji, wynikającą stąd redukcję kosztów oraz obciążeń społecznych. Nowym zadaniem statystyki jest więc opracowanie strategii integracji istniejących baz danych oraz estymacji zapewniającej numeryczną zgodność publikowanych informacji w różnych przekrojach, klasyfikacjach i wymiarach. Baza mikrodanych statystycznych powstaje jako rezultat pracy statystyków, informatyków, ekonometryków i specjalistów od badań reprezentacyjnych. Charakteryzuje się ona zwiększoną „wartością”, jaką daje efekt synergii wskutek integracji źródeł statystycznych nie tylko w sensie zakresu informacji, ale także poprawy ich jakości wynikającej z harmonizacji badań.

3. Spis wirtualny – idea i źródła zasilania

W rekomendacjach międzynarodowych wskazuje się trzy metody przeprowadzania spisów ludności: tradycyjną, opartą na rejestrach administracyjnych, oraz złożoną, będącą kombinacją dwóch poprzednich. Metoda tradycyjna polega na przeprowadzaniu wywiadów w terenie. Spis oparty na rejestrach sprowadza się do gromadzenia informacji spisowych z istniejących źródeł administracyjnych, z tego powodu określany jest mianem „wirtualnego”. Trzecia metoda jest kombinacją dwóch poprzednich. W założeniach do spisu powszechnego w 2011 r. (por. [*Narodowy Spis Powszechny...* 2007]) przyjęto zastosowanie metody złożonej opartej na rejestrach, badaniu pełnym (krótki formularz) i badaniach reprezentacyjnych. Przyjęte rozwiązania są zgodne z tendencjami obserwowanymi w innych krajach europejskich, wśród których przodują państwa skandynawskie³.

Podstawowym źródłem danych w NSP 2011 będzie istniejący w Polsce rejestr ludności – PESEL. Ponadto rozważa się możliwość wykorzystania 28 innych rejestrów administracyjnych, co wymagać będzie ścisłej współpracy między GUS-em i jego gestorami (łącznie ok. 88 różnych instytucji) – por. [*Narodowy Spis Powszechny...*, 2007]. Informacje niedostępne w rejestrach przewiduje się zebrać w spisie opartym na tzw. krótkim formularzu osobowym. Badanie to pozwoli także

³ W Szwecji od ponad 30 lat przeprowadza się spis oparty na rejestrach. W latach 2000-2001 metodę tę zastosowano także w Danii, Finlandii, Norwegii, Holandii, Niemczech, Islandii (por. [Linder 2003]).

ocenić jakość i spójność struktur ludności oszacowanych na podstawie rejestru PESEL oraz za pomocą metody tradycyjnej. Pozostałe informacje pochodzą będą z szacunków na bazie badania reprezentacyjnego według tzw. długiego formularza o rozszerzonym zakresie tematycznym.

Przy podjęciu realizacji spisu opartego na rejestrach administracyjnych podstawowym warunkiem jest dostęp do danych jednostkowych, możliwość integracji poszczególnych źródeł oraz pokrycie informacyjne w zakresie tzw. tematów podstawowych. Jednym z często podnoszonych mankamentów spisów opartych na rejestrach jest brak informacji zgodnych z zaleceniami organizacji międzynarodowych w zakresie pojęć, definicji, klasyfikacji itp. Na przykład w zakresie charakterystyki ekonomicznej ludności do tematów podstawowych w rekomendacjach międzynarodowych zaliczono: bieżącą aktywność ekonomiczną, zawód wykonywany, rodzaj działalności zakładu pracy oraz status zatrudnienia (por. [*United Nations Principles...* 2007]). Lista tematów dodatkowych jest nieco dłuższa i obejmuje: stałą aktywność ekonomiczną, wolontariat, sektor instytucjonalny, czas pracy, niepełne zatrudnienie, okres pozostawiania bez pracy, rodzaj miejsca pracy, liczbę osób w nim pracujących, lokalizację zakładu pracy, odległość i czas dojazdu, dochód, zatrudnienie nieformalne (szara strefa). Żadne ze źródeł przewidzianych do wykorzystania w NSP'2011 nie zawiera wszystkich cech określonych jako podstawowe. Niektóre są w ogóle niedostępne, różnią się od spisowych definicją albo są rozproszone po różnych rejestrach (np. system podatkowy Ministerstwa Finansów, centralny rejestr ZUS, system informacji o ubezpieczonych i o świadczeniobiorcach KRUS, system informacyjny PFRON i inne). Zawarte w nich informacje różnią się pod względem definicji, klasyfikacji oraz dotyczą różnych grup ludności. Problem integracji informacji z tych źródeł oraz wyników badania reprezentacyjnego jest więc niezmiernie istotny.

4. Metody integracji – spójność wyników

W literaturze wyróżnia się dwie grupy procedur w zakresie integracji danych, określając je mianem procedur mikro i makro (por. [Denk, Hackl 2003]). Procedury mikro dotyczą łączenia danych na poziomie rekordów, tj. tych samych indywidualnych jednostek rzeczywistych pochodzących z różnych zbiorów. Natomiast procedury makro odnoszą się do agregacji zintegrowanych danych. Do kategorii mikro zaliczane są procedury dokładnego (jednoznacznego) i statystycznego⁴ łączenia rekordów, jak również metody imputacji danych. Natomiast wśród procedur makro wymienia się różne metody ważenia, polegające np. na dostosowaniu wyników badań reprezentacyjnych do odpowiednich struktur populacji, oraz procedury łączenia danych syntetycznych. Wśród metod integracji wyróżnić można m.in.: łączenie jednoznacznego, klasy jakościowe, komparatory łańcuchowe, probabilistyczne łączenie re-

⁴ W parowaniu statystycznym nie ma wymogu łączenia rekordów dla tej samej realnej jednostki. Połączone rekordy mogą być raczej określane mianem jednostek „syntetycznych” aniżeli rzeczywistych.

kordów, metody taksonomiczne, klasyfikację obiektów, metody delimitacji, metody statystyczne oraz imputację danych i metody geostatystyczne. Przegląd metod integracji danych statystycznych zawiera opracowanie [Denk i Hackl 2003].

Integracja danych w spisie opartym na rejestrach może mieć różnorodny wymiar w zależności od charakteru i zakresu łączonych źródeł. Rozważyć tutaj można następujące sytuacje: łączenie baz danych pełnych (np. rejestru administracyjnego ze spisem „krótki formularz” czy z innym rejestrem), łączenie źródeł pełnych z wynikami badania reprezentacyjnego (np. rejestru czy spisu „krótki formularz” z badaniem reprezentacyjnym) oraz łączenie wyników różnych badań reprezentacyjnych (np. danych BAEL z danymi BBGD). Innym problemem, oprócz zakresu badania, może być kwestia jednoznacznych identyfikatorów rekordów. Rozróżnić można: łączenie baz posiadających identyfikator, baz bez identyfikatorów oraz tworzenie „klucza identyfikacyjnego”.

Uwzględniając potrzeby NSP’2011, zaproponować można następującą koncepcję integracji wykorzystującą metodę klas jakościowych oraz komparatora łańcuchowego dla danych adresowych. Propozycja dotyczy sytuacji łączenia źródeł pełnych oraz wyników badania reprezentacyjnego przy braku jednoznacznych identyfikatorów rekordów indywidualnych. Dla celów jej testowania mogą posłużyć wyniki NSP’2002 oraz BAEL’2002. Etap pierwszy sprowadza się do podziału obu zbiorów na rozłączne klasy według cech adresowych (powiat, gmina), klasę miejscowości i płci. Następnie w wyodrębnionych klasach określa się zmienne parujące, np.: wiek, wykształcenie oraz stan cywilny. Łączenie rekordów w ramach wyróżnionych grup jednorodnych przeprowadzić można metodą probabilistycznego łączenia rekordów lub parowania statystycznego (por. [Kadane 1978; 2001]).

Korzystanie z różnych źródeł informacji wiąże się z brakiem ich spójności. Porównując dane dotyczące aktywności ekonomicznej ludności z NSP’2002 oraz BAEL z II kwartału 2002 r. (por. tab. 1), rozbieżności można zauważyć już na poziomie całego kraju. Różnica w liczbie ludności w wieku 15 lat i więcej nie przekraczała 1%, ale w odniesieniu do kategorii ludności wyróżnionych ze względu na status na rynku pracy była już znaczna, sięgając 19% w przypadku ludności aktywnej zawodowo czy ponad 23% w odniesieniu do pracujących. Jeszcze większe rozbieżności pojawiają się wraz z uwzględnieniem podziału terytorialnego. Przy czym największe rozbieżności w wartościach syntetycznych wskaźników wcale nie dotyczą tych województw, w których względne błędy szacunku były największe. Dodatkowo różna jest skala rozbieżności oraz błędów estymacji w tym samym województwie w zależności od szacowanych parametrów. Największe różnice odnotowano w województwie świętokrzyskim, gdzie BAEL w ok. 8% niedoszacowuje współczynnika aktywności zawodowej, oraz w podkarpackim, w którym współczynnik ten jest w ok. 8% przeszacowany. Natomiast najmniejsza precyzja szacunku dotyczy województw opolskiego (REE = 2,43%) i warmińsko-mazurskiego (REE = 2,32%). W celu zapobieżenia takim konsekwencjom korzystania z różnych źródeł statystyka proponuje stosowanie metod ważenia, kalibracji

czy poststratyfikacji. W Danii np. zaproponowano metodę podwójnego ważenia (por. [Renssen i in. 2001]).

Tabela 1. Relacja pomiędzy informacjami o aktywności ekonomicznej ludności w przekroju województw na podstawie NSP'2002 i BAEL, II kwartał 2002

Województwo	NSP'2002			BAEL II'2002			Relacja NSP/BAEL		
	współczynnik aktywności zawodowej	wskaznik zatrudnienia	stopa bezrobocia	współczynnik aktywności zawodowej	wskaznik zatrudnienia	stopa bezrobocia	współczynnik aktywności zawodowej	wskaznik zatrudnienia	stopa bezrobocia
Ogółem	55,5	43,7	21,2	55,6	44,6	19,9	0,998	0,980	1,065
Dolnośląskie	54,2	40,3	25,6	54,6	40,8	25,3	0,993	0,988	1,012
Kujawsko-pomorskie	55,8	42,3	24	55,4	43	22,3	1,007	0,984	1,076
Lubelskie	55,5	45,2	18,5	59,1	49,4	16,4	0,939	0,915	1,128
Lubuskie	54,7	39,9	27,1	55,1	40,8	25,9	0,993	0,978	1,046
Łódzkie	56,2	44,9	20,2	56,7	45,2	20,2	0,991	0,993	1,000
Małopolskie	54,1	43,7	19,1	57,5	48,5	15,6	0,941	0,901	1,224
Mazowieckie	58,3	48,1	17,4	57,3	47,7	16,7	1,017	1,008	1,042
Opolskie	54,7	42,9	21,6	55,3	43,7	20,9	0,989	0,982	1,033
Podkarpackie	54,3	42,7	21,4	58,9	48,2	18,2	0,922	0,886	1,176
Podlaskie	56,3	45,5	19,1	56,8	47,6	16,1	0,991	0,956	1,186
Pomorskie	56,1	43,6	22,3	55,9	44,2	20,9	1,004	0,986	1,067
Śląskie	52,9	41,9	20,8	51	40,7	20,3	1,037	1,029	1,025
Świętokrzyskie	56,3	43,8	22,2	52,2	42,3	18,9	1,079	1,035	1,175
Warmińsko-mazurskie	55,1	39,6	28,2	53,4	38,9	27,3	1,032	1,018	1,033
Wielkopolskie	56,9	46,3	18,6	56,7	46,1	18,7	1,004	1,004	0,995
Zachodnio-pomorskie	55,4	40,3	27,3	53,6	39,3	26,6	1,034	1,025	1,026

Źródło: opracowanie własne na podstawie danych NSP'2002, www.stat.gov.pl oraz raportu *Aktywność ekonomiczna ludności Polski*, II kwartał 2002 r., GUS, Warszawa.

W sytuacji dostępności różnych źródeł informacji ważnym problemem jest spójność i porównywalność danych statystycznych. Wagę tego zagadnienia odzwierciedlają programy badań podejmowane przez urzędy statystyczne różnych krajów. W Danii prace w zakresie spójności informacji prowadzono w ramach projektu *One Figure For One Phenomenon* (por. [Bakker, van Rooijen 2000]), a w Wielkiej Brytanii realizowano projekt *One Number Census* (por. [Brown i in. 1999]).

5. Korzyści i zagrożenia

Rozważając nowe podejście metodologiczne, podjęto próbę porównania jego wad i zalet (por. tab. 2). Po stronie potencjalnych zagrożeń podkreślić należy administracyjny charakter. Ważnym problemem jest uzyskanie akceptacji społecznej dla korzystania z informacji zawartej w rejestrach i obawa przed inwigilacją. Wśród innych problemów wskazać można zależność statystyki publicznej od organów administracji państwowej, wymóg ścisłej współpracy między wieloma gestorami i GUS-em. Kwestię zgodności „momentu krytycznego” spisu można rozwią-

zać za pomocą odpowiednich regulacji administracyjnych. W tym zakresie istnieje także możliwość zastosowania odpowiednich procedur demograficznych, takich jak np. postarzanie ludności. Wiele niebezpieczeństw wiąże się ze stosowaniem imputacji na poziomie jednostkowym oraz integracji statystycznej dla małych jednostek terytorialnych.

Tabela 2. Korzyści i zagrożenia zastosowania podejścia opartego na rejestrach w NSP 2011

Korzyści	Zagrożenia
Efektywne wykorzystanie istniejących źródeł	Brak dostępności wszystkich zalecanych informacji
Redukcja kosztów	Zależność statystyki publicznej od organów administracji państwowej
Redukcja obciążeń społecznych	Wymóg ścisłej współpracy między gestorami rejestrów i GUS-em
Poprawa bezpieczeństwa danych	Zmiany regulacji prawnych skutkują jakością otrzymanych statystyk
Gwarancja harmonizacji badań	Zgodność momentu krytycznego spisu
Dostępność informacji spisu opartego na rejestrach każdego roku	Problemy powstałe wskutek imputacji oraz integracji statystycznej danych z różnych źródeł
Dostępność danych rejestrów administracyjnych dla dowolnego poziomu dezagregacji terytorialnej	
Możliwość identyfikacji błędów „podwójnego spisywania”	
Utworzenie mikrobazy danych wspomagających estymację pośrednią – modelowanie na poziomie jednostkowym	
Poprawa estymacji dla małych obszarów dzięki zastosowaniu metod integracji wykorzystujących silnie skorelowane zmienne	
Poprawa spójności i wiarygodności danych statystycznych	

Źródło: opracowanie własne.

Jednak zdecydowanie więcej jest korzyści wynikających z integracji rejestrów. Najważniejsza ze społecznego punktu widzenia jest efektywność wykorzystania istniejących źródeł, redukcja kosztów i obciążeń społecznych oraz poprawa bezpieczeństwa danych. Dodać należy także zwiększenie dostępności danych dla odbiorców w sensie czasu i przestrzeni, tj. częstotliwości oraz szczegółowości zakresu terytorialnego. Za podejściem bazującym na łączeniu rejestrów oraz badań specjalnych przemawiają także korzyści o charakterze merytorycznym. Obejmują one przede wszystkim poprawę struktur ludności w wyniku redukcji błędów podwójnego spisywania, utworzenie bazy mikrodanych, która może być wykorzystana dla poprawy estymacji dla małych obszarów, poprawę spójności oraz wiarygodności publikowanych statystyk.

6. Podsumowanie

W artykule rozważano istotę zmian w działaniu statystyki publicznej związanych z wykorzystaniem zasobów rejestrów administracyjnych, odnosząc je do planowanego na 2011 r. spisu ludności. Przedstawiono także koncepcję integracji źródeł w zakresie aktywności ekonomicznej ludności z wykorzystaniem metod klas jakościowych oraz probabilistycznego łączenia rekordów lub parowania statystycznego. Rozważania o integracji źródeł administracyjnych dla potrzeb NSP'2011 zakończono syntetycznym zestawieniem wynikających z tego podejścia wad i zalet. Planowana na 2011 r. zmiana metodologii będzie pierwszą w Polsce próbą wykorzystania rejestrów w spisie ludności. Tym bardziej nawet tylko częściowa realizacja planowanych zamierzeń będzie sukcesem zmierzającym do poprawy efektywności wykorzystania istniejących źródeł informacji.

Literatura

- Al P.G., Bakker B.F.M. (2000), *Re-engineering social statistics by micro-integration of different sources; an introduction*, „Netherlands Official Statistics”, vol. 15, Summer 2000, Special issue: *Integrating administrative registers and household surveys*, Statistics Netherlands.
- Bakker B.F.M., van Rooijen J. (2000), *One figure for the supply and demand of services*, „Netherlands Official Statistics”, vol. 15, Summer 2000, Special issue: *Integrating administrative registers and household surveys*, Statistics Netherlands.
- Brown J.J., Buckner L., Diamond I.D., Chambers R., Teague A. (1999), *A methodological strategy for a one number census in the UK*, „Journal of the Royal Statistical Society”, Series A, 162, s. 247-267.
- Denk M., Hackl P. (2003), *Data integration and record matching: an Austrian contribution to research in official statistics*, „Austrian Journal of Statistics”, vol. 32, no 4, s. 305-321.
- Kadane J.B. (1978), *Some Statistical Problems in Merging Data Files*, [w:] *Compendium of Tax Research*, US Dept. of the Treasury, s. 159-171 (reprinted in „Journal of Official Statistics” 2001 no 17(3), s. 423-433).
- Kadane J.B. (2001), *Some statistical problems in merging data files*, „Journal of Official Statistics” no 17, s. 423-433.
- Linder F. (2003), *The Dutch virtual census 2001, a new approach by combining administrative registers and household sample surveys*, DIECOFIS Workshop Vienna, 13-14 November, http://petra1.istat.it/diecofis/App/WienWorkshop/Presentations/Frank_Linder.pdf.
- Narodowy Spis Powszechny Ludności i Mieszkań 2011. Założenia metodyczne* (2007), Materiał na posiedzenie Rady Programowej narodowego spisu powszechnego ludności i mieszkań w 2011 r.
- Renssen R.H., Kroese A.H., Willeboordse A.J. (2001), *Aligning estimates by repeated weighting*, Proceedings of Statistics Canada Symposium 2001, Achieving Data Quality in a Statistical Agency: A Methodological Perspective.
- United Nations Principles and Recommendations for Population and Housing Censuses* (2007), http://unstats.un.org/unsd/demographic/sources/census/docs/P&R2_February%2012%202007.pdf.
- van der Laan P. (2000), *Integrating administrative registers and household surveys*, „Netherlands Official Statistics”, vol. 15, Summer 2000, Special issue: *Integrating administrative registers and household surveys*, Statistics Netherlands, Voorburg/Heerlen.

INTEGRATING DATA FROM DIFFERENT SOURCES FOR VIRTUAL CENSUS 2011

Summary

The paper presents the idea of integrating different data sources to meet the needs of population census 2011, which is planned in Poland as the 'virtual' census making use of administration registers. It discusses the redesign of production process of social statistics that comes from the underlying tendency to increase efficiency by exhaustive use of all the available data sources. The author presents the international census recommendations in the field of economic activity in view of core topics and non-core topics and their prospective availability in administration records. Methods of data integration are presented together with the proposition of their implementation in the census process illustrated by economic activity data from the previous census 2002 and Labour Force Survey. Advantages and disadvantages of the register based population census are summarized.