

**Marek CIERPIAŁ-WOLAN, Grażyna MARCINIAK,
Włodzimierz OKRASA, Janusz WITKOWSKI**

Współczesne problemy statystyki — teoria i praktyka w perspektywie globalnej

59 Światowy Kongres Statystyki, Hongkong 2013

Międzynarodowy Instytut Statystyczny (MIS) jest jedną z najstarszych naukowych organizacji na świecie. Został powołany w 1885 r. i przez cały okres istnienia stanowi autonomiczne towarzystwo (organizację), które poszukuje możliwości rozwoju i doskonalenia metod statystycznych oraz ich aplikacji poprzez promocję międzynarodowej współpracy. Kolejne kongresy światowe (wcześniej sesje plenarne) MIS stanowią duże wydarzenia naukowe dla środowiska statystyków. Organizowane co dwa lata, gromadzą naukowców z różnych dziedzin wiedzy statystycznej oraz praktyków reprezentujących urzędy statystyczne, a także prywatne instytucje statystyczne. Jest to zawsze doskonała okazja do zaprezentowania najnowszych osiągnięć metodycznych, badawczych i organizacyjnych w dziedzinie statystyki. Dzięki temu wypełniana jest misja tej organizacji w zakresie wymiany profesjonalnej wiedzy. Od wielu lat coraz większe zainteresowanie pracami MIS wykazują praktycy reprezentujący urzędy statystyki publicznej. Wzrost aktywności tego środowiska jest bardzo widoczny. Z tego też względu w obradach sesji plenarnej kongresów statystyki, a także innych wydarzeniach organizowanych przez MIS dość powszechnie uczestniczą pracownicy urzędów statystycznych z większości krajów.

59 Światowy Kongres Statystyki zgromadził 2,5 tys. uczestników reprezentujących 114 krajów ze wszystkich kontynentów oraz różne instytucje związane bezpośrednio lub pośrednio ze statystyką. W tym wydarzeniu udział wzięli — jak zwykle — przedstawiciele nauki, statystyki oficjalnej, stowarzyszeń i organizacji międzynarodowych, a także sektora prywatnego. Program Kongresu obejmował sesje naukowe, spotkania administracyjne, krótkie kursy i spotkania satelitarne.

Komitet Naukowy MIS zorganizował zarówno sesję, na które zaprosił prelegentów (IS — *Invited Sessions*), jak też specjalne sesje dotyczące określonych dziedzin statystyki (STS — *Special Topic Sessions*) oraz sesje plakatowe (PS — *Poster Sessions*). Omówiono na nich nie tylko dorobek statystyki jako dyscypliny naukowej, ale także jej praktyczne zastosowania w wielu dziedzinach życia, a przede wszystkim wyznaczały perspektywę rozwoju.

Na program naukowy Kongresu złożyło się 1400 referatów prezentowanych podczas 243 sesji. Delegaci GUS zostali zaproszeni do opracowania i przedstawienia referatów w dwóch sesjach. W pierwszej STS, poświęconej *Roli spisów ludności i mieszkań w systemach statystyki krajowej* Grażyna Marciniak wygłosiła referat pt. *Nowoczesne podejście do spisów. Przykład Polski — zalety i ograniczenia* (*Modern approach to censuses in the case of Poland — advantages and constraints*). Podsumowania całości sesji dotyczącej spisów, składającej się z prezentacji prowadzonych przez delegatów z Włoch, Polski, Hiszpanii, Danii i Węgier, dokonał prezes GUS Janusz Witkowski, zaproszony do pełnienia roli dyskutanta.

Wystąpienia polskich prelegentów spotkały się z żywym zainteresowaniem uczestników sesji. Oceniono, że doświadczenia krajów, które przeprowadziły spisy z wykorzystaniem źródeł administracyjnych stanowią doskonałą podstawę do dyskusji nad przyszłą formą spisów.

W drugiej sesji nt. *Statystyka operacji celnych i transgranicznych: perspektywa dla współpracy* (*Intra-customs and transborder statistics: Prospects for cross-cooperation*) Marek Cierpiał-Wolan przedstawił referat pt. *Procesy w obszarach transgranicznych — istotny wpływ na wzrost gospodarczy* (*Processes in transborder areas — significant impact on the economic growth*). W ramach tej sesji Włodzimierz Okrasa zaprezentował referat, opracowany z M. Cierpiał-Wolanem i Pawłem Markockim, pt. *Transgraniczna nieformalna działalność gospodarcza: wielopoziomowe oszacowanie efektów dla dobrostanu społeczności lokalnych — przypadek Polski i krajów sąsiedzkich* (*Transborder Non-Institutional Economic Activities: Multilevel Assessment of their Impact on Community Well-Being — case of Poland and Neighboring Countries*). W podsumowaniu sesji dotyczącej statystyki transgranicznej — którego dokonał Józef Oleński jako zaproszony dyskutant — uznano, że stanowi ona ważny element rozwoju statystyki regionalnej i że niezbędne byłoby włączenie tych badań do aktywności międzynarodowej w ramach ESS (Europejski System Statystyczny).

W sesji poświęconej metodologii i teorii statystyki wystąpili Krzysztof Szajowski i Aleksandra Ochman-Gozdek z Politechniki Wrocławskiej oraz Daniel Kosiorowski z Uniwersytetu Ekonomicznego w Krakowie. Wyniki prac o charakterze teoretycznym zaprezentowali także przedstawiciele Uniwersytetu Łódzkiego. Czesław Domański i Katarzyna Bolonek-Lasoń przygotowali pracę na temat uogólniania pewnych rozkładów prawdopodobieństwa. Z kolei Alina Jędrzejczak przedstawiła temat związany z własnościami asymptotycznymi estymatorów, a prezentacja Jerzego Korzeniewskiego dotyczyła analiz przestrzennych. Zastosowania praktyczne przedstawili Kamil Kruszyński i Wacława Starzyńska. Ich prezentacja dotyczyła wpływu funduszy europejskich na standard życia w ujęciu regionalnym. W tej części wystąpili również przedstawiciele Uniwersytetu Ekonomicznego w Katowicach, m.in. Justyna Majewska i Grażyna Trzpiot z prezentacją o zastosowaniach testów niezależności na rynku papierów wartościowych, a także Sebastian Twaróg, Grażyna Trzpiot, Anna Ojrzyńska i Jacek Szołtysek — na temat zastosowania statystyki w medycynie.

W sesji plakatowej zaprezentowano jeszcze jeden polski referat przygotowany przez Włodzimierza Okrasę na podstawie wstępnych wyników badań pilotażowych pt. *Przestrzenne aspekty dobrostanu lokalnego — analiza kontekstowych i indywidualnych źródeł zmienności za pomocą modelowania wielopoziomowego* (*Spatial Aspects of Community Well-Being. Analyzing Contextual and Individual Sources of Variation using Multilevel Modeling*), którego innowacyjnym elementem było odwołanie się do metod statystyki przestrzennej w analizie danych geokodowanych na poziomie indywidualnych gospodarstw domowych i jednostek lokalnych.

PROBLEMATYKA NAUKOWA — WYBRANE NURTY TEMATYCZNE

Zakres tematyczny części Kongresu był bardzo szeroki. W prezentacjach i dyskusjach występowały wszystkie najważniejsze aktualnie problemy teoretyczne i praktyczne statystyki. W sposób syntetyczny można wyróżnić cztery dziedziny problemowe, które były najczęściej dyskutowane, a mianowicie:

- teoria statystyki,
- nauczanie statystyki,
- metodologia badań statystycznych,
- statystyka oficjalna.

Ten ostatni problem był szczególnie mocno akcentowany na Kongresie dzięki dużej aktywności Międzynarodowego Stowarzyszenia Statystyki Publicznej (IAOS), koncentrującego się w znacznym stopniu na problemach statystyki publicznej w wymiarze globalnym. W tym zakresie dyskutowane były kwestie dotyczące: reorganizacji badań w krajowych urzędach statystycznych, jakości statystyki oficjalnej, edukacji statystycznej, wykorzystywania administracyjnych źródeł danych oraz *Big Data*, łączenia informacji z różnych źródeł, przyszłości spisów i innych specyficznych zagadnień metodologicznych (np. problemów pokrycia i estymacji) oraz pomiaru jakości życia i dobrostanu subiektywnego. Niektóre z tych zagadnień stają się nowymi dziedzinami statystyki oficjalnej, jak np. subiektywne aspekty dobrostanu.

IAOS to międzynarodowa organizacja, która skupia przedstawicieli statystyki (producentów) i użytkowników danych, a jej głównym zadaniem jest promowanie zrozumienia i postępu statystyki publicznej, w tym praktycznych wzorów usprawnień instytucjonalnych i metodologicznych. Stowarzyszenie wykazuje dużą aktywność na forum międzynarodowym, korzystając także z okazji Światowego Kongresu Statystyki. Wśród bardziej szczegółowych tematów dotyczących statystyki oficjalnej znalazły się zagadnienia niezwykle istotne dla wiarygodności i użyteczności danych statystyki oficjalnej, jak:

- rola i znaczenie statystyki, a w ślad za tym kierunki jej rozwoju (wyzwania wobec statystyki),
- jakość produkowanych danych oraz wiarygodność instytucji statystycznych,

- modernizacja funkcjonowania statystyki na różnym poziomie organizacji,
- przyszłość spisów ludności i mieszkań,
- korzystanie z alternatywnych źródeł danych, w tym wykorzystanie *Big Data*, dla celów statystycznych,
- współpraca z użytkownikami oraz popularyzacja informacji i wiedzy statystycznej,
- integracja badań, wdrażanie nowych metod pomiaru rzeczywistości społeczno-ekonomicznej i syntetyzacja wiedzy.

ZNACZENIE STATYSTYKI PUBLICZNEJ

Wymienione zagadnienia są ważne dla sprawności funkcjonowania statystyki publicznej, ale ich rozwiązanie ma przede wszystkim odpowiadać na potrzeby użytkowników. Z tego względu w trakcie wielu spotkań i sesji przewijał się problem znaczenia statystyki dla decydentów i innych użytkowników danych. Podkreślano rolę statystyki oficjalnej w kreowaniu postępu i pomocy w podejmowaniu decyzji, co świadczy o użyteczności statystyki w procesie kierowania i zarządzania. Nie ulega wątpliwości, że początkiem sukcesu jest dobra jakość informacji. Z przebiegu dyskusji wynikała jednoznacznie teza o potrzebie bardziej aktywnej roli statystyki oficjalnej w zarządzaniu rozwojem na różnym poziomie organizacji społeczeństwa. Podkreślano także inspirującą rolę statystyki i to nie tylko w zakresie poznawania rzeczywistości społeczno-gospodarczej, ale też wdrażania najnowszych rozwiązań technologicznych. Ten wątek obrad Kongresu nawiązywał do nowych oczekiwań dotyczących statystyki oficjalnej, wynikających z kryzysu finansowego i gospodarczego, który ujawnił kluczową rolę danych statystycznych jako podstawy diagnozowania i monitorowania sytuacji społeczno-gospodarczej we współczesnym świecie.

Z większą dostępnością danych statystycznych łączy się także ich dobra jakość. Te problemy są obecnie w zakresie zainteresowań niemal wszystkich urzędów statystycznych i większości organizacji międzynarodowych koordynujących bądź korzystających z danych statystycznych. Coraz więcej użytkowników informacji oczekuje bowiem lepszej jakościowo statystyki: bardziej precyzyjnych danych, o większym znaczeniu dla oceny zjawisk społecznych i ekonomicznych, bardziej wiarygodnych wyników badań oraz szybciej udostępnianych informacji.

W trakcie Kongresu przedstawiano interesujące rozwiązania praktyczne w urzędach statystycznych Szwecji, Holandii i Kanady w zakresie poprawy jakości badań statystycznych. Kilka sesji poświęcono problematyce braku wiedzy, która jest szczególnie ważna dla statystyki w naszym kraju.

Ciekawy przykład podali przedstawiciele Urzędu Statystycznego Szwecji. Ze względu na fakt, że w badaniu zdrowia (HES) mieszkańców Szwecji (próba 20000 osób) zaobserwowano stale zmniejszający się odsetek odpowiedzi (z 61%

w 2004 r. do 49% w 2011 r.), postanowiono przeprowadzić badanie eksperymentalne. Głównym celem było określenie kilku alternatywnych strategii kontaktu z respondentami oraz zmniejszenie kosztów¹.

W ramach prezentacji przygotowanej przez przedstawicieli Urzędu Statystycznego Holandii także nawiązano do kwestii braku odpowiedzi w badaniach. Podkreślono, że taka sytuacja wpływa nie tylko na zmniejszenie ilości dostępnych danych, ale także na obiektywność wyników badania. Jednakże zwiększenie wielkości próby nie jest dobrym rozwiązaniem, ponieważ prawdopodobieństwo nieuzyskania odpowiedzi pozostaje bez zmian. Zatem wymaga to nowego podejścia do projektowania badań, w których nie każda wybrana do badania osoba jest traktowana w ten sam sposób. Zakłada się, że różne osoby objęte badaniem powinny być traktowane w rozmaity sposób, określany przed rozpoczęciem badania. Sposób ten zależy od informacji dostępnych na etapie przygotowywania badania. Takie informacje można znaleźć w operacie (np. zmienne demograficzne) lub w źródłach administracyjnych, które mogą być związane z próbką do badania. W trakcie prezentacji podkreślono również, że coraz częściej organizacja badań jest konfrontowana z cięciami budżetowymi i presją obniżenia kosztów badań. Powoduje to przejście od badań wspomaganych terminalem przenośnym (CAPI) i telefonicznie (CATI) do badań samodzielnie zarządzanych, w szczególności badań internetowych (CAII). Doświadczenie pokazuje, że odsetek odpowiedzi w przypadku ankiet internetowych jest nie większy niż 40%.

Statystyka publiczna musi zatem wdrożyć metody badań statystycznych z zastosowaniem nowych narzędzi zbierania danych, ich przetwarzania oraz udostępniania informacji. Aby spełnić te oczekiwania statystyka oficjalna powinna szerzej korzystać z nowoczesnej technologii oraz dostosowywać metodologię i organizację badań do nowych możliwości technologicznych. W tym kontekście podkreślano potrzebę modernizacji statystyki oficjalnej pod względem organizacji badań statystycznych, wykorzystania nowoczesnej techniki realizacji badań oraz komunikowania się ze społeczeństwem, jak również zarzą-

¹ Standardowa strategia (S, wielkość grupy 2500) zakładała wysłanie kwestionariusza papierowego, jak i instrukcji do logowania się do ankiety internetowej. Alternatywna strategia 1 (A1, wielkość grupy 2500) zakładała wysłanie zaproszenia do uczestnictwa w badaniu oraz instrukcji do logowania się do ankiety internetowej. Zaproszenie zawierało również informacje o kwestionariuszu papierowym wysłanym do osób, które nie wypełniły ankiety *on-line* w przeciągu tygodnia. Alternatywna strategia 2 (A2, wielkość grupy 2500) zakładała wysłanie zaproszenia do uczestnictwa w badaniu. Następnie wysłano drugi list z zaproszeniem do uczestnictwa w badaniu, instrukcję do logowania się do ankiety internetowej oraz kwestionariusz papierowy. Wszystkie trzy strategie kontaktu generowały niemal równy odsetek odpowiedzi. Nawet w obrębie podgrup nie było różnic między trzema strategiami. Tak więc, w odniesieniu do wskaźników odpowiedzi, żadna z dwóch alternatywnych strategii nie dała lepszych wyników niż standardowa. Warto podkreślić, że standardowa strategia wymagała wysyłania większej liczby kwestionariuszy papierowych. Jeśli chodzi o odpowiedzi *on-line*, strategia A1 dała lepsze wyniki niż pozostałe dwie, zatem, jak podkreślano, będzie wykorzystana w dużych badaniach reprezentacyjnych.

dzania statystyką. Kluczową jednak sprawą jest wdrożenie takiego sposobu myślenia i działania, aby — poza modyfikacją metodologii badań, która jest realizowana sprawnie przez wszystkie instytucje statystyczne — przebudować statystykę tak, aby szybciej, lepiej i skuteczniej reagowała na nowe potrzeby użytkowników, aby przyczyniła się do lepszej harmonizacji danych oraz lepszej porównywalności informacji jako warunku do właściwego wnioskowania w celu spożytkowania wyników badań do podejmowania konkretnych działań programowych. Modernizacja jest zatem ściśle powiązana z poprawą aktywnej roli statystyki oficjalnej.

POMIAR JAKOŚCI ŻYCIA I DOBROSTANU SUBIEKTYWNEGO

Wśród nurtów tematycznych zasługujących na uwagę z racji rosnącego znaczenia w statystyce oficjalnej wielu krajów wyróżniały się też zagadnienia związane z doskonaleniem pomiaru jakości życia i dobrostanu oraz ich roli w badaniach poziomu i warunków życia całych populacji, wybranych grup socjodemograficznych czy społeczności lokalnych. Szczególnie silny akcent położono na uzupełnienie tradycyjnych ujęć — odwołujących się do miar obiektywnych (monetarnych, materialnych i innych, w tym wskaźników społecznych) — o miary o charakterze subiektywnym. Opierając się na sondażach opinii w zakresie rozmaicie konceptualizowanych i operacjonalizowanych pojęć pokrewnych — takich jak szczęście, satysfakcja, postęp społeczny czy dobrostan danej kategorii osób (np. dzieci czy osób starszych) — badania te należą do jednych z najbardziej dynamicznie rozwijających się na świecie tak z powodu ich rosnącej atrakcyjności i znaczenia w szeroko pojętej polityce redystrybucyjnej (w tym w polityce społecznej), jak też z racji ich nie do końca doprecyzowanego statusu metodologicznego (brak jednorodności ujęć) oraz zróżnicowanego miejsca, jakie zajmują one w systemach statystyki oficjalnej poszczególnych krajów. Najbardziej zaawansowane pod tym względem programy badań prowadzone są przez Urząd Statystyczny Wielkiej Brytanii (*The Measuring National Well-being Programme*, realizowany od 2010 r.). Intensywne prace badawczo-wdrożeniowe podejmowane są także przez statystyków we Włoszech.

Większość krajów próbuje jednak wdrażać rozwiązania pośrednie, łączące wykorzystanie danych z istniejących sondaży z eksperymentalnymi modułami tematycznymi, np. INSEE we Francji, by wspomnieć tylko te kraje, których przedstawiciele prezentowali na Kongresie wyniki dokonań swoich urzędów statystycznych. Występująca odmiennność stosowanych skal pomiarowych odzwierciedla specyfikę krajowych warunków i potrzeb, uwzględniając z jednej strony cele polityki wewnętrznej, z drugiej zaś postawy i opinie w zakresie dziedzin istotnych dla subiektywnych aspektów dobrostanu (jak poziom akceptacji ryzyka rozwojowego, np. dla zdrowia, czy awersję do nierówności dochodów).

Obiecujący jest fakt bezpośredniego zaangażowania się w tę problematykę organizacji międzynarodowych, na czele z OECD i Eurostatem, które również

pokazały programy poświęcone metodologii i koordynacji badań w zakresie subiektywnych aspektów dobrostanu w skali międzynarodowej i międzykulturowej. Na specjalną uwagę zasługuje (popularyzowany w kilku prezentacjach na Kongresie) dorobek OECD, a zwłaszcza opublikowane w połowie ub. roku opracowanie pt. *Wytyczne dotyczące pomiaru odczuwanego dobrostanu (Guidelines on Measuring Subjective Well-being)*. Jest to najbardziej ambitna, jak dotąd, próba standaryzacji pojęć i podejść wraz z rekomendacjami dla statystyki oficjalnej, z intencją umożliwienia dokonywania porównań międzynarodowych w tej trudno uchwytnej, ale ważnej materii. Nie jest to wszakże propozycja jedyna — stosunkowo najbliższe jej rekomendacjom są badania jakości życia i dobrostanu subiektywnego w Australii i Szkocji.

Z kolei Eurostat, respektując w zasadzie podejście OECD do samego pomiaru dobrostanu subiektywnego, rekomenduje tworzenie systemu wskaźników jakości życia i dobrostanu jako komplementarnego do systemu rachunków narodowych, sugerując m.in. wyjście poza pytania stosowane dotychczas w EU-SILC (np. włączając pytania o subiektywne aspekty dobrostanu do *Badań budżetów gospodarstw domowych* oraz do *Badań budżetu czasu*), postulując zarazem równoległe uwzględnianie wskaźników dotyczących środowiska naturalnego i rozwoju zrównoważonego. Jeszcze innego przykładu dostarcza opublikowane niedawno przez panel ekspertów Amerykańskiego Komitetu Statystyki opracowanie, które niestety nie było przedstawione na Kongresie.

Wszystkie z opracowanych dotychczas metod pomiaru dobrostanu subiektywnego zasługują na wnikliwą uwagę. Nie tylko w celu spełnienia kryteriów metodologicznych, ale też opracowania rekomendacji praktyki krajowych badań statystycznych w sposób uwzględniający zarazem specyficzne potrzeby polityki w sferze ekonomicznej i społecznej.

SPISY LUDNOŚCI I MIESZKAŃ

Ważne miejsce w dyskusji kongresowej zajmowała problematyka spisów ludności i mieszkań, a także spisów rolnych. Dotyczyła ona głównie pytania o przyszłość spisów ludności i mieszkań w warunkach ograniczeń finansowych, coraz większych trudności w realizacji badań na podstawie wywiadów oraz niezwykle skomplikowanej organizacji spisów tradycyjnych. Stąd próba poszukiwania odpowiedzi, czy spisy tradycyjne mają szanse być kontynuowane nawet w warunkach, kiedy w ostatniej rundzie badań spisowych ponad 85% krajów realizowało spis tradycyjny. Z tego względu zainteresowanie nowymi rozwiązaniami w zakresie realizacji spisów było duże. Te kwestie podejmowano na kilku sesjach, w tym również na zorganizowanej przez Węgierski Urząd Statystyczny, z udziałem statystyków polskich. Doświadczenia zaprezentowane przez kilka krajów cieszyły się dużym zainteresowaniem. W dyskusji dokonywano porównań korzyści i niedogodności różnych rozwiązań. Poza kilkoma krajami, które realizują spisy ludności wyłącznie na podstawie źródeł administracyjnych, naj-

bardziej popularne są rozwiązania mieszane. Dotyczy to również doświadczeń polskich, w ramach których wprowadzono nową metodologię realizacji spisu (połączenie danych administracyjnych z reprezentacyjnym badaniem pogłębionym) oraz nowe źródła uzyskiwania danych. Jest to rozwiązanie, które ma największe szanse na szersze zastosowanie w kolejnej rundzie spisów na świecie. Z tego względu polskie rozwiązania i doświadczenia spotkały się z dużym zainteresowaniem.

Ostatnia runda spisów zaowocowała zarówno wzrostem zastosowań geo-przestrzennych metod i technologii w zbieraniu danych, jak i ogromnym przyrostem geokodowanych danych jednostkowych. Dlatego obserwowano na Kongresie olbrzymie zainteresowanie rozwojem i zastosowaniem nowego podejścia „przestrzennego” zarówno w wyznaczaniu prób (nie tylko w badaniach związanych z rolnictwem, ekologią, epidemiologią itp.), jak i uzyskiwaniu danych. Dotyczyło to analiz angażujących metody statystyki przestrzennej na najniższym poziomie oszacowań, co znakomicie uzupełniało modelową technikę estymacji małych obszarów.

BIG DATA JAKO POTENCJALNE ŹRÓDŁO DANYCH DLA STATYSTYKI PUBLICZNEJ

W ramach modernizacji statystyki oficjalnej, szczególnie w zakresie organizacji badań statystycznych, dużo miejsca zajmowała kwestia korzystania z alternatywnych źródeł danych, które nie ograniczają się tylko do danych administracyjnych, ale dotyczą także innych źródeł występujących w formie elektronicznej i nowoczesnej technologii. Mamy tu na myśli tzw. *Big Data*, które od kilku lat stanowią przedmiot zainteresowania statystyków oficjalnych w kontekście możliwości ich wykorzystania do potrzeb statystyki. *Big Data* (ogromne zasoby informacyjne/wielkie zbiory danych, nazywane też danymi masowymi, w szczególności w zastosowaniu do danych z badań naukowych) zostały zdefiniowane przez Europejską Komisję Gospodarczą (*United Nations Economic Commission for Europe* — UNECE) jako duża ilość szybkich i różnorodnych danych, które wymagają wydajnych, innowacyjnych form przetwarzania. Obecnie są one coraz częściej wykorzystywane do tworzenia polityki informacyjnej.

W trakcie Kongresu koncentrowano się na różnych aspektach merytorycznych oraz na prezentacji prób zastosowania tego typu danych w praktyce, omawiano metodologiczne innowacje uwzględniające specyfikę tak ogromnych zbiorów informacji, charakteryzujących się zarazem wielką różnorodnością — od społecznych i gospodarczych, poprzez klimatyczne, astrofizyczne, medyczne, środowiskowe, po eksperymentalne i mikrosymulacyjne, generowane w ramach praktycznie każdej dziedziny działalności, w tym naukowej.

Wszyscy autorzy i uczestnicy specjalistycznych sesji zgodnie stwierdzili, że w ciągu najbliższych lat wykorzystanie *Big Data* wpłynie na funkcjonowanie

statystyki publicznej, podkreślali związane z tym zarówno możliwości, jak i zagrożenia. Przypominali artykuł opublikowany w „Le Monde” pod kontrolującym tytułem *Google wieźą więcej lub mogą dowiedzieć się więcej na temat Francji niż INSEE*.

Możliwości *Big Data* wynikają przede wszystkim z ogromnej skali produkcji ogólnodostępnych danych, które jednocześnie wspierane są przez rozpowszechnianie w coraz większym stopniu urządzeń cyfrowych. Dane pochodzące z różnych źródeł, takich jak portale społecznościowe, telefonia komórkowa, transakcje *on-line* charakteryzują się dużą różnorodnością i częstotliwością. Warto podkreślić, że obejmują one ludzkie działania, doświadczenia, pragnienia, zamiary i oczekiwania. Mogą być wykorzystane np. do monitorowania inflacji, a także do tworzenia wskaźników wzrostu gospodarczego czy badania migracji w czasie rzeczywistym. Nowe rodzaje danych, pochodzące od jednostek i społeczności z ich codziennego życia, można wykorzystać do stworzenia alternatywnych miar ubóstwa i dobrobytu.

Równocześnie innowacje i udoskonalenia w zakresie techniki obliczeniowej, a także coraz większa moc obliczeniowa urządzeń pozwalają porządkować te ogromne i złożone strumienie danych.

Autorzy referatów zwrócili uwagę, że w wielu krajach służby statystyki publicznej borykają się z niedoborem środków finansowych, technicznych i ludzkich i przez to często mają trudności w dostarczaniu podstawowej statystyki. Zastanawiano się zatem, czy *Big Data* stanowią szansę dla statystyki czy być może są swoistego rodzaju zagrożeniem dla niej. Zależy to, według prelegentów, od skali efektywnego wykorzystania *Big Data* w systemach statystyki publicznej.

Tak czy inaczej korzystanie z tego typu danych stanowi znaczne wyzwanie dla tradycyjnych form zbierania i przetwarzania danych w urzędach statystycznych ze względu na ilość, czas uzyskiwania i szybkość ich przetwarzania. Wyraźnie jednak podkreślano, że internetowe źródła danych, oparte na systematycznym monitorowaniu i rejestrowaniu transakcji dotyczących jednostki obserwacji w określonej populacji, wydają się oferować oczywiste korzyści dla oficjalnej statystyki. W rzeczywistości wymagają jednak zmiany sposobu myślenia i infrastruktury w odniesieniu do sposobu zbierania i przetwarzania danych. Struktury tych danych są zazwyczaj zestandaryzowane pomiędzy krajami, jednolite lub łatwe do znormalizowania. Mogą więc stanowić cenne i szybkie źródło informacji.

Szczególnie konieczna, w przypadku *Big Data*, wydaje się również bardziej systematyczna i skoordynowana współpraca organizacji statystycznych na forum międzynarodowym, wymiana doświadczeń pomiędzy krajami, realizacja projektów pilotażowych oraz współpraca z uczelniami.

W ramach sesji poświęconych *Big Data* przedstawiono również innowacyjny system LILIANA (*LIve LIinking for on-line statistic ANALytics*) stworzony w celu wspierania i łączenia wiedzy znajdującej się w Internecie ze statystyką *on-line*. Został on opracowany w ramach projektu KE LAWLA (*Longitudinal*

Analytics of Web Archive Data). Funkcjonowanie tego systemu jest szczególnie istotne ze względu na odmienny charakter tekstowych treści internetowych zawartych w różnych bazach wiedzy w sieci Web i w dokumentach statystycznych, co powoduje, że wykorzystanie ich jest ograniczone.

LILIANA to pierwszy system wypełniający lukę pomiędzy statystyką liczbową i tekstowymi treściami internetowymi. Identyfikuje on podobieństwa tekstowe i semantyczne pomiędzy treściami znajduwanymi w bazach wiedzy w sieci Web i wskazuje linki do najbardziej powiązanych z nimi artykułów statystycznych, a tym samym dostarcza cennych informacji kontekstowych. W ten sposób można wzbogacać różne strony internetowe, np. przez przeglądarki „plug-in” wraz z linkami do odpowiedniej statystyki, m.in. na stronach Eurostatu. Ze względu na to, że baza Eurostatu *Statistics Explained* z jej słownikiem statystycznym i dobrze zdefiniowanym tezaurem jest podobnym źródłem do Wikipedii, może być ona połączona z innymi bazami wiedzy, takimi jak np. *Freebase*, *Dbpedia* czy *YAGO*.

Do semantycznego wzbogacania treści tekstowych archiwum sieci Web treściami z bazy *Statistics Explained* wybrano bazę wiedzy YAGO, ponieważ ma także zestandaryzowaną strukturę. Interfejs użytkownika, po wpisaniu określonego tekstu i wybraniu opcji „Link do statystyki *on-line*”, umożliwia wybór podstawowych metod (hybrydowej lub z wykorzystaniem jednego z modeli) łączenia go z treściami w sieci Web. Następnie wskazuje podmiot przypisany do tekstu w formie łącza do Wikipedii lub innego porównywalnego źródła wiedzy oraz łącza do odpowiednich artykułów statystycznych.

Warto podkreślić, że jest to temat niezwykle popularny, ale nie doczekał się jeszcze uporządkowania w sensie wyboru kluczowych kwestii do rozwiązania z punktu widzenia ich użyteczności dla statystyki publicznej.

Przy okazji refleksji nad zastosowaniem w statystyce nowoczesnej technologii podkreślano konieczność jej spożytkowania do celów lepszej komunikacji z użytkownikami danych statystycznych, a więc lepszego, bardziej przyjaznego i szybszego udostępniania danych. Możliwości w tym względzie są obecnie nieporównywalne w stosunku do tych sprzed lat. Jest to ważny problem doskonalenia statystyki w celu poprawy jej wizerunku, użyteczności i wypełniania aktywnej roli w zarządzaniu rozwojem.

Obok dyskutowanych intensywnie aspektów metodologicznych i technicznych, kwestia wykorzystania danych masowych w statystyce publicznej stawia też nowe wyzwania dla polityki dostępu i bezpieczeństwa, w tym poufności danych indywidualnych (szczególnie ważnej dla tzw. danych społecznych, tak w wymiarze krajowym jak i globalnym, z uwagi na np. transgraniczny zakres niektórych transakcji). Wyzwania takie wiążą się nie tylko z charakterem danych, ale też z nowymi rodzajami procedur i infrastruktury obliczeniowej typu „chmury”, pozwalających na przechowywanie, dzielenie się i procesowanie danych w systemach sieciowych, ale również narażając je na różnego rodzaju ataki i cyberterroryzm.

EDUKACJA STATYSTYCZNA

Na Kongresie bogato reprezentowany był tradycyjnie ważny temat dotyczący kwestii związanych z nauczaniem statystyki, tak w sensie uczenia jej jako przedmiotu w szkołach i uczelniach — w tym problem *curriculum* i sposobów kształcenia — jak i w sensie wprowadzania jej do świadomości i rutynowego postępowania w sferach decyzyjnych polityki i biznesu oraz podnoszenia kultury statystycznej obywateli w wielu krajach (także w najbardziej zaawansowanych). Kwestie te łączyły się w niektórych sesjach z dyskusją na temat roli statystyki, danych i metod statystycznych oraz statystyki jako profesji w szerszych, poza-badawczych kontekstach jej zastosowania. Podkreślano, iż w tym sensie statystyka wymaga ciągłej transformacji w kierunku lepszego wykorzystania nowych środków i sposobów uatrakcyjnających i ułatwiających jej obecność poprzez stałe usprawnianie technologiczne współdziałania „producentów” danych, na czele z instytucjami statystyki oficjalnej, z partnersko traktowanymi użytkownikami. W tym kontekście, niezależnie od pojawiających się wyzwań związanych z rosnącym udziałem danych generowanych poza tradycyjnym systemem statystyki publicznej, jakimi są (obok danych administracyjnych) dane transakcyjne, internetowe i inne, zwrócono uwagę na wciąż istniejącą niejednoznaczność co do podejścia metodologicznego i formalnego statusu niektórych z nowo pojawiających się źródeł danych.

Na uwagę zasługuje akcentowanie potrzeby budowania edukacji i kultury statystycznej poczynając od wczesnych etapów nauczania w szkołach, traktując dany poziom wiedzy i sposób korzystania z informacji statystycznej jako stan wynikający z wieloletniego procesu. Wizja edukacji i kultury statystycznej sugeruje zarazem potrzebę odpowiednio głębokiego, całościowego i dynamicznego ujmowania tych zagadnień.

MISCELLANEA

Warto dodać, że w ramach programu naukowego Kongresu odbywały się sesje historyczne związane z Międzynarodowym Rokiem Statystyki 2013, który był obchodzony przez ponad 2 tys. organizacji na świecie. Upamiętniono także 300-letnią rocznicę publikacji Jacoba Bernoulliego, zawierającej pierwsze poważne twierdzenie probabilistyczne oraz 250-lecie publikacji *An essay towards solving a problem in the doctrine of chances* (*Esej w sprawie rozwiązywania problemu w doktrynie szansy*) Thomasa Bayesa.

W trakcie Kongresu były także sesje, w których omawiano zastosowanie statystyki w astronomii, medycynie czy meteorologii. Niektóre wyniki badań, zarówno od strony teoretycznej jak i praktycznej, mogą być zastosowane w statystyce publicznej. Ciekawy przykład przedstawił zespół Uniwersytetu Kalifornijskiego, który zaprezentował zastosowanie modeli z nieparametrycznym rozkładem efektów losowych w analizie czasu przeżycia na podstawie da-

nych medycznych. Związek między wszystkimi obserwowanymi przypadkami i osobami, które przeżyły był uchwycony przez utajone efekty losowe. Unikalnym elementem zaproponowanego modelu była możliwość odchodzenia od założeń dotyczących normalności efektów losowych i pozostawienie rozkładu całkowicie nieokreślonego. Autorzy oszacowali losowe efekty rozkładu i parametry modelu, natomiast reprezentanci Uniwersytetu w Szanghaju i Hongkongu zaproponowali zastosowanie zmiennych współczynników w modelu regresji kwantylowej (dla danych obciążonych z tytułu czasu trwania badania). Takie podejście nie tylko pozwala na bezpośrednie oszacowanie warunkowych kwantyli czasu przeżycia na podstawie elastycznej struktury modelu, ale dodatkowo dopuszcza zależności między zmienną obserwowaną (ocenzuowania) i zmiennymi przewidywanymi na podstawie badania.

Uwagi końcowe

59 Światowy Kongres Statystyki pozwolił na prezentację osiągnięć promotorów statystyki z wielu krajów i w różnorodnych dziedzinach badań oraz wymianę informacji dotyczących kierunków działań, ze szczególnym uwzględnieniem zagadnień porównywalności i jakości statystyki w ujęciu globalnym.

Prezentacja doświadczeń polskiej statystyki publicznej na tak szerokim forum — skupiającym zarówno metodologów, praktyków jak i użytkowników statystyki — stanowiła doskonałą okazję promocji aktywności GUS w ważnych dziedzinach statystyki międzynarodowej. Udział w tematycznych sesjach naukowych pozwolił na poznanie prac prowadzonych w ośrodkach akademickich oraz europejskich i pozaeuropejskich krajowych urzędach statystycznych, a także na przedyskutowanie możliwości wdrożenia niektórych rozwiązań do międzynarodowej statystyki oficjalnej.

Wiele uwagi poświęcono zagadnieniom związanym z nowoczesną organizacją spisów powszechnych, wieloaspektowym wykorzystaniem *Big Data*, jak też badaniu szeroko rozumianej jakości życia. Swoiste wyzwanie dla statystyki publicznej stanowi także wielowymiarowy charakter procesów transgranicznych. Wyniki badań w tych dziedzinach są źródłem inspiracji do modernizacji statystyki publicznej, a tym samym wyznaczają kierunki jej długookresowej transformacji. Należy zaznaczyć, że w zakresie tych zagadnień podjęto w polskiej statystyce publicznej szereg przedsięwzięć o charakterze metodologicznym i organizacyjnym.

Warto podkreślić, że przewodnim tematem Kongresu była „Młodość”, która jest siłą napędową i filarem w dzisiejszym społeczeństwie opartym na informacji i wiedzy. Dla młodych statystyków zorganizowano liczne sesje oraz różne projekty, m.in. Spotkania Satelitarne Młodych Statystyków, odbywające się w Korei Południowej, Japonii, Tajlandii, Chinach i w samym Hongkongu. Czwarty dzień kongresu był w całości poświęcony temu zagadnieniu. W ostatnich latach MIS dokonuje wielu wysiłków, których celem jest ułatwienie dostę-

pu do organizacji młodym i początkującym statystykom. Kongresowa Komisja Młodych Statystyków poszukiwała sposobu, aby stworzyć sieć umożliwiającą wspólne porozumiewanie się i nawiązanie kontaktów przez młodych statystyków z różnych organizacji statystycznych na całym świecie.

dr Marek Cierpiał-Wolan — *Urząd Statystyczny w Rzeszowie*, **dr hab. Włodzimierz Okrasa** — *profesor Uniwersytetu Kardynała Stefana Wyszyńskiego, GUS*, **dr Grażyna Marciniak** — *GUS*, **prof. dr hab. Janusz Witkowski** — *GUS*

SUMMARY

The 59th World Statistics Congress organized by the International Statistical Institute, which is one of the oldest scientific organization in the world, allowed to present achievements of statistics representatives from many countries and in various areas of research, as well as to exchange information on the direction of future activities. The Congress program included scientific sessions, short courses, administrative and satellite meetings which presented not only the achievements of statistics as a scientific discipline, but also its practical applications in many areas of life, and above all, determined the perspective of the development of statistics.

During the Congress also the experience of Polish statistics was presented, which was an excellent opportunity to promote the activity of the CSO and the academic community in important areas of international statistics. Much attention was paid to modern organization of censuses, multifaceted use of Big Data as well as the quality of life in its broad sense. The specific challenge for public statistics is also multidimensional nature of cross-border processes. The results of research in these areas are the source of inspiration for the modernization of public statistics, and thus define the direction of its long-term transformation.

The Congress program devoted much attention to young statisticians for whom numerous sessions and satellite conferences were organized, primarily in order to create a network for communication and contacts between young statisticians from different statistical organizations around the world.

РЕЗЮМЕ

59. Всемирный конгресс статистики был организован Международным статистическим институтом, который является одной из старейших научных организаций в мире. Встреча позволила представителям статистики многих стран представить свои достижения в разнообразных областях обследований, а также позволила обмениваться информацией касающимися направлений будущей деятельности. Программой Конгресса предусмотрены научные сессии, администра-

тивные встречи, краткие курсы и сателлитные встречи, во время которых представлялись не только достижения статистики с научной точки зрения, но и практическое использование во многих областях жизни, и прежде всего были указаны перспективы развития статистики.

Во время Конгресса был представлен опыт польской статистики, это считалось хорошей возможностью подтверждения активности ЦСУ и научных кругов в важных областях международной статистики. Большое внимание было уделено вопросам, связанным с современной организацией переписей, многоаспектному использованию Big Data, а также обследованию качества жизни. Конкретной задачей государственной статистики является также многомерный характер трансграничных процессов. Результаты исследований в этих областях являются источником инспирации для модернизации государственной статистики, и тем самым определяют направления ее долгосрочной трансформации.

Во время Конгресса было посвящено внимание молодым статистикам, для которых были организованы сессии и сателлитные конференции с целью создания условий для совместного общения и контактов молодых статистиков из разных статистических организаций.