

Modernizacja statystyki działalności gospodarczej w krajach Unii Europejskiej

Zrównoważony i ciągły rozwój gospodarczy Europy wymaga harmonizacji istniejących oraz nowych źródeł danych statystycznych charakteryzujących zachodzące w tej dziedzinie procesy, jak również stosowania nowoczesnych metod uzyskiwania, przetwarzania oraz analizy zgromadzonych informacji. Owe źródła i metody umożliwiają stałe obserwowanie różnorodnych zjawisk, formułowanie wniosków i na ich podstawie wdrażanie stosownych strategii ekonomicznych zarówno na poziomie rządowym, jak i samorządowym czy też poszczególnych przedsiębiorstw.

Aby efektywny monitoring był możliwy (co stanowi kluczowy warunek skuteczności podejmowanych działań strategicznych), konieczne jest zapewnienie odpowiedniego wsparcia metodologicznego i informatycznego, a przede wszystkim naukowego. Osiągnięciu tych celów służy zestaw projektów badawczych i ujednolicejających rozwiązania metodologiczne realizowanych w ramach kompleksowego programu *Modernisation of European Enterprise and Trade Statistics* (MEETS) ustanowionego mocą decyzji nr 1297/2008/WE Parlamentu Europejskiego i Rady z 16 grudnia 2008 r. w sprawie programu modernizacji europejskiej statystyki przedsiębiorstw i handlu. Realizacja tego ogromnego przedsięwzięcia rozpoczęła się w 2009 r. Główną platformą wykonawczą stała się tutaj sieć współpracy w Europejskim Systemie Statystycznym (*European Statistical System* — ESS) — tzw. sieć ESS (ESSnet) — stworzona w celu sprawnej wymiany wiedzy i doświadczeń specjalistycznych, monitorowania postępów prac oraz udostępniania wyników badań.

Działania MEETS obejmują prace w ramach projektów realizowanych w niektórych państwach członkowskich, jak również inicjatywy podejmowane w drodze indywidualnych umów grantowych zawieranych z krajowymi urzędami statystycznymi oraz poprzez zlecenie prowadzenia badań zewnętrznych. Wyniki tych projektów są przekazywane do ESS. Towarzyszy temu wsparcie techniczne i administracyjne, np. organizacja warsztatów i spotkań grup zadaniowych z udziałem krajowych i międzynarodowych ekspertów. O tym, jak ogromny zakres zadań mieści w sobie program MEETS, niech świadczy to, że Unia Europejska (UE) przeznaczyła na jego realizację łącznie ok. 5 mln euro w 2009 r. i ok. 10 mln euro w 2010 r., obejmując łącznie ponad 25 projektów¹.

¹ Sprawozdanie Komisji dla Parlamentu Europejskiego i Rady w sprawie wdrożenia decyzji nr 1297/2008/WE Parlamentu Europejskiego i Rady z dnia 16 grudnia 2008 r. w sprawie programu modernizacji europejskiej statystyki przedsiębiorstw i handlu (MEETS), Bruksela, 11 stycznia 2011 r. (dostępne pod adresem <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0813:FIN:PL:PDF>).

W niektórych inicjatywach aktywny udział bierze również strona polska. Przykładem może tu być badanie możliwości wykorzystania źródeł administracyjnych w statystyce przedsiębiorstw, koordynowane ze strony polskiej przez prof. dr hab. Grażynę Dehnel z Uniwersytetu Ekonomicznego w Poznaniu oraz Urzędu Statystycznego w Poznaniu. Celem tego projektu jest zbadanie użyteczności dostępnych źródeł dla rocznej i krótkookresowej statystyki przedsiębiorstw w odniesieniu do średnich i dużych podmiotów gospodarczych, zmniejszenie obciążeń średnich i dużych podmiotów gospodarczych wynikających ze sprawozdawczości statystycznej, zwiększenie efektywności prowadzonych szacunków, zastosowanie wzorcowania do szacunku braków danych, wykorzystanie nowych metod estymacji pośredniej (statystyka małych obszarów) w celu poprawy precyzji szacunku, jak również rozszerzenie zakresu przekrojów, w których publikowane są wyniki badania.

W artykule zajmiemy się projektem modernizacji statystyki działalności gospodarczej *Methodology for Modern Business Statistics* (w skrócie *MeMoBuSt*) realizowanym przez Urząd Statystyczny w Poznaniu. Wpisuje się on w ogólne cele MEETS. Jego zamierzeniem jest wspieranie wdrażania coraz efektywniejszych metod uzyskiwania, przetwarzania i analizy danych statystycznych z zakresu wytwórczości i handlu. W tym kontekście występuje potrzeba uogólnienia i rozprzeczowania najlepszych myśli praktycznych (co powinno umożliwić stworzenie odpowiedniej pomocy metodologicznej) oraz dokumentowania rozwiązań powszechnie stosowanych w zakresie prowadzenia statystyki działalności gospodarczej w UE.

Kolejny cel projektu polega na redukcji obciążenia odpowiedzi, doskonaleniu jakości poprzez optymalizację projektowania badań i przetwarzania danych, wykorzystaniu istniejących źródeł informacji (rejstry), zastosowaniu nowoczesnych środków gromadzenia danych (zwłaszcza w kontekście rozwoju technologii informacji i komunikacji) oraz nowych metod estymacji. Należy przy tym utrzymywać i rozwijać rzetelność oraz spójność procesów tworzenia produktu statystycznego. Powinno to przyczynić się do zwiększenia stopnia porównywalności danych dla różnych dziedzin i krajów członkowskich oraz efektywności ich integracji. Dzięki zastosowaniu nowoczesnych narzędzi badawczych i partycypacji w kosztach ich rozwoju uzyskuje się wyższą efektywność prac statystycznych, a bardziej efektywne wykorzystanie zewnętrznych źródeł danych obniży obciążenie odpowiedzi. W projekcie realizowanym pod nadzorem Eurostatu bierze udział 8 narodowych instytucji statystycznych z: Grecji, Niemiec, Norwegii, Polski, Szwecji, Szwajcarii, Węgier i Włoch. Głównym koordynatorem jest dr Leon Willenborg z Centralnego Biura Statystycznego Niemiec (CBS). Projekt ma trwać 3 lata (2011—2013) i składać się z dwóch części — każda po 1,5 roku realizacji.

GLÓWNE CELE OPERACYJNE PROJEKTU MODERNIZACJI STATYSTYKI DZIAŁALNOŚCI GOSPODARCZEJ *MeMoBuSt*

Zasadniczym zadaniem programu jest opracowanie podręcznika metodologicznego w zakresie pozyskiwania, gromadzenia, analizy i rozpowszechniania danych statystycznych z zakresu statystyki działalności gospodarczej. Ma on

powstać w miejsce niewystarczającego i częściowo już zdezaktualizowanego poradnika pod redakcją A. Willeboordse (1997). Do najistotniejszych ułomności tej publikacji należą:

- podręcznik pisany był niemal wyłącznie przez stronę niderlandzką i dlatego też dotyczy na ogół realiów statystyki tego kraju, pozbawiony jest zatem porównań międzynarodowych;
- sporo informacji na temat aktów prawnych obowiązujących w UE dotyczących statystyki działalności gospodarczej jest nieaktualnych, brak też informacji o nowych uregulowaniach;
- nie wyspecyfikowano kluczowych zmiennych statystycznych służących do klasyfikacji podmiotów gospodarczych celem efektywnego prowadzenia badań statystycznych, jak również zasadniczych klas zmiennych w nich stosowanych;
- brak jest aktualizacji rejestrów gospodarczych oraz różnic metodologicznych pomiędzy nimi, a także najlepszych schematów losowania prób;
- nie omówiono problemów dotyczących rozpoznawania statusu LEA (ang. *Legally and Economically Active* — aktywny prawnie i gospodarczo) podmiotów gospodarczych, który stanowi kluczowe kryterium włączania ich do badań, w tym do operatu losowania prób reprezentacyjnych;
- nie sprecyzowano powszechnie występujących błędów odpowiedzi i ich przyczyn oraz zmiennych, które najczęściej wymagają imputacji z tego właśnie powodu;
- narzędzia elektronicznego przetwarzania i analizy danych nie odpowiadają współczesnemu stanowi rozwoju tej dziedziny techniki.

W rezultacie dyskusji ustalono, że nowy podręcznik winien obejmować kompletną dziedzinę statystyki i docelowo być adresowany do całego ESS, z uwzględnieniem perspektywy integracyjnej jako punktu wyjścia do poprawy jakości i efektywności europejskiej statystyki działalności gospodarczej. Postanowiono rozpocząć prace od skonstruowania zasad tworzenia i ramowej struktury nowego dzieła i na tej podstawie rozpoznać możliwości ewentualnego wykorzystania dotychczasowej wersji. Podręcznik ten nie będzie jednak skompilowany w stylu „ad hoc”. Trzeba bowiem uwzględnić fakt, iż użytkownicy mają różne potrzeby.

Należy przeanalizować rozmaite rodzaje badań, np. krótkookresowe, strukturalne oraz aspekty długookresowe, metody weryfikacyjne, techniki łączenia danych próbkowych i danych administracyjnych, metody redukowania obciążenia respondentów, problemy operatu opartego na rejestrach i integracji danych. Sporo uwagi warto poświęcić nowym technologiom, projektowaniu e-kwestionariuszy czy gromadzeniu danych za pomocą sieci. Konieczna będzie ocena terminowości, dostępności, precyzji, zrównoważenia pomiędzy jakością, budżetem, potrzebami, funduszami i obciążeniami tak wewnątrz poszczególnych badań, jak również pomiędzy nimi. Ponadto pożądane wydaje się wykorzystanie wyników innych projektów realizowanych w ramach pakietu MEETS lub poza nim, jak również odpowiednich źródeł informacji.

Jednym z istotnych założeń jest to, iż podręcznik po jego opracowaniu ma być powszechnie dostępny w formie elektronicznej. Zaletą tego podejścia byłaby

także łatwość aktualizacji dzieła. Co więcej, podejmowane są starania zmierzające do precyzyjnego określenia poszczególnych tematów — tak, aby mogły zostać im poświęcone odrębne moduły (mniej lub bardziej niezależne rozdziały), które można by łatwo modyfikować bez istotnej ingerencji w pozostałe. Ma to być rodzaj portalu dostarczającego czytelnikowi nie tylko pierwszej informacji, ale także wskazującego bardziej szczegółowe źródła wiedzy.

Podręcznik będzie zestawem standardów, wykorzystywanym jako podstawowe i łatwo dostępne źródło. Pożądane jest by podręcznik był stosowany w kursach szkoleniowych dla statystyków z ESS. Dlatego też w odpowiednim czasie zostanie przeprowadzone rozpoznanie, czy podręcznik jest użyteczny dla instytucji statystycznych, jak i krajów uczestniczących w ESS. W tym kontekście przewiduje się podjęcie czynności sprawdzających, mających na celu zagwarantowanie tego, żeby zawartość podręcznika mogła być wykorzystana w cyklu szkoleń realizowanych w ramach Europejskiego Programu Szkoleń Statystycznych (*European Statistical Training Programme* — ESTP). Doświadczenia zebrane podczas takich szkoleń posłużą do ulepszania użyteczności podręcznika.

W przypadku gdy dany temat poruszany w publikacji jest nowy i rozwiązania praktyczne na nim oparte nie zdołały się jeszcze dostatecznie wykształcić, możliwe będzie przeprowadzenie badań z tego zakresu. Wyniki takich studiów winny zostać zapisane w raportach, które mogą posłużyć jako konieczna podstawa informacyjna (motywacja, rozumowanie implikacyjne, materiał techniczny itp.) i znajdą swoje odzwierciedlenie w odpowiednim module podręcznika. Należy podkreślić, że ewentualne studia badawcze wykonane w ramach tego projektu dotyczyć będą tylko określonego modułu podręcznika. W przypadku gdy okaże się, że badanie takie zostało już przeprowadzone w ramach innego projektu, w podręczniku znajdzie się odniesienie do niego. Można też włączyć wykonawców pokrewnego projektu do współpracy przy redagowaniu podręcznika albo nawet zaproponować im napisanie odpowiedniego modułu w całości. Każdy moduł podręcznika będzie podlegał wewnętrznej recenzji dokonywanej przez kraje biorące udział w przedsięwzięciu z możliwym współuczestnictwem recenzentów zewnętrznych.

Podczas przygotowywania podręcznika szczególna uwaga ma być zwrócona na następujące zagadnienia:

- projektowanie: organizacja, schemat losowania i ogólna zgodność, sposoby mieszane, koordynacja badań reprezentacyjnych i losowania prób celem umożliwienia wykorzystania powiązań danych, zakres obciążenia odpowiedzi, wykorzystanie technik tworzenia hurtowni danych;
- gromadzenie danych: nowe techniki gromadzenia informacji (kwestionariusze sieciowe, elektroniczne) — jak wykorzystać te techniki do redukcji kosztów i optymalizacji jakości; wtórne wykorzystanie źródeł administracyjnych, zastosowanie odpowiednich zakresów metadanych² do monitorowania procesów twórczych;

² Metadane — definicje lub opisy zmiennych, które umożliwiają użytkownikom odnalezienie potrzebnej informacji wraz z ustaleniem jej ewentualnych powiązań formalnych, logicznych czy metodologicznych z innymi danymi.

— estymacja: kiedy i jak używać odpowiednich modeli do zastąpienia luk w zgromadzonych danych, prognozowania, predykcji informacji przeszłych, jak i systematycznego raportowania błędów nielosowych.

Zakres zagadnień objętych treścią podręcznika jest szeroki, przenika wszystkie rodzaje działalności statystycznej. Wymaga to zaangażowania zespołu ekspertów z poszczególnych dziedzin. Jest także rzeczą ważną, żeby specjaliści z różnych krajów współpracowali ze sobą w celu wykorzystania doświadczeń i rozwiązań praktycznych stosowanych w różnych państwach członkowskich objętych ESS. Do wykonania takich ekspertyz będzie też można zaangażować inne osoby i instytucje.

TECHNICZNE I METODOLOGICZNE ZAŁOŻENIA OPRACOWANIA NOWEGO PODRĘCZNIKA ZAWIERAJĄCEGO METODOLOGIĘ BADAŃ DZIAŁALNOŚCI GOSPODARCZEJ

Od strony merytorycznej i redakcyjnej rzecz ujmując, finalny produkt projektu będzie miał charakter podręcznika metodologii badań gospodarczych, obejmujących wszystkie fazy przewidziane schematem GSBPM (*Generic Statistical Business Process Model* — Podstawowy Model Prowadzenia Statystyki Działalności Gospodarczej), ale niekoniecznie wszystkie metody. Zgodnie ze stanowiskiem przedstawionym w *Generic...* (2009), GSBPM został stworzony w celu stosowania do działań podejmowanych przez badaczy zajmujących się gromadzeniem, analizą i publikowaniem danych statystycznych w statystyce publicznej zarówno na poziomie krajowym, jak i międzynarodowym. Ma on być niezależny od źródeł danych i winien być używany do opisu i oceny jakości procesów opartych na badaniach statystycznych, spisach powszechnych, rejestrach administracyjnych oraz innych niestatystycznych czy mieszanych źródłach informacji. W odróżnieniu od typowych badań statystycznych, GSBPM może być także wykorzystany do przypadków, gdy istniejące dane zostały uprzednio zweryfikowane bądź gdy dokonano powtórnego przeliczenia szeregów czasowych — czy to w wyniku zastosowania szerszych lub lepszych źródeł danych czy też zmiany metodologii. Jego twórcy uważają, że model ten da się również zastosować do rozwijania i utrzymywania rejestrów statystycznych, których dane są podobne do tych uzyskiwanych w badaniach i analizach statystycznych, a rezultaty tych działań stanowią typowe operaty losowania lub wyciągi informacyjne wykorzystywane w innych procesach. Schemat GSBPM obejmuje wszystkie etapy i procedury związane z projektowaniem, przeprowadzeniem, kontrolą jakości, analizą danych i publikowaniem rezultatów badania statystycznego z zakresu statystyki działalności gospodarczej.

W przypadku rozpatrywanego podręcznika szczególną uwagę poświęca się modelowi GSBPM w kontekście projektowania, przeprowadzania, oceny efektów i rozpoznania reakcji respondentów badań statystycznych oraz przechodzenia od dokonywania samej oceny do weryfikacji tejże oceny. Oprócz modelu GSBPM w trakcie redakcji podręcznika brać się będzie pod uwagę techniczny standard SDMX (*Statistical Data and Metadata eXchange*) z Powszechnym

Słownikiem Metadanych (MCV — *Metadata Common Vocabulary*). SDMX stanowi powszechne techniczne i statystyczne standardy oraz wskazówki (w tym strukturę i narzędzia IT³) stworzone przez konsorcjum składające się z BIS⁴, ECB⁵, Eurostat, MFW, OECD⁶, UN⁷ oraz Banku Światowego, celem wykorzystania do efektywnej wymiany i współdzielenia danych oraz metadanych statystycznych. Uniwersalny format plików danych oraz ich standaryzowana zawartość według SDMX są wstępnymi warunkami zautomatyzowanego gromadzenia, przetwarzania i wymiany danych pomiędzy krajowymi i międzynarodowymi instytucjami statystycznymi. Więcej szczegółów na ten temat można znaleźć na oficjalnej stronie projektu: <http://sdmx.org/>.

Integralną częścią SDMX jest Powszechny Słownik Metadanych (MCV) zawierający idee i odpowiednie definicje, do których może odnosić się terminologia stosowana w metadanych krajowych i międzynarodowych organizacji zajmujących się statystyką. Najnowsza wersja słownika (*SDMX Content-Oriented...*, 2009) obejmuje:

- ogólne koncepcje metadanych pochodzących głównie z dokumentów ISO⁸ i UNECE⁹, użyteczne dla tworzenia ram zarządzania metadanymi;
- pojęcia metadanych opisujące metodologię statystyczną (częstotliwość, okres odniesienia, gromadzenie danych, źródło, dopasowania itp.);
- metadane niezbędne do oceny jakości informacji (precyzja, terminowość itp.);
- pojęcia odnoszące się szczególnie do wymiany danych i metadanych (terminologia pochodząca z modelu informacyjnego SDMX oraz z istniejących definicji itp.).

Słownik MCV powstał na podstawie wielu źródeł i nie jest jeszcze całkiem wewnętrznie spójny, ale stanowi dobry punkt wyjścia do badań i analiz, ewentualnie dalszej jego rozbudowy. Powiązanie MCV i GSBPM daje też możliwości efektywniejszej komunikacji pomiędzy autorami i użytkownikami podręcznika, a doświadczenia wyniesione z redakcji publikacji mogą posłużyć także do ulepszenia treści słownika MCV, np. w kontekście paradanych¹⁰.

³ IT — *Information Technology* — technologia informacyjna, czyli ogół elektronicznych narzędzi i technik przetwarzania oraz publikowania informacji.

⁴ BIS — *Bank for International Settlements* — Bank Rozliczeń Międzynarodowych.

⁵ ECB — *European Central Bank* — Europejski Bank Centralny.

⁶ OECD — *Organization for Economic Cooperation and Development* — Organizacja Współpracy Gospodarczej i Rozwoju.

⁷ UN — *United Nations* — Organizacja Narodów Zjednoczonych.

⁸ ISO — *International Organization for Standardization* — Międzynarodowa Organizacja Standaryzacyjna.

⁹ UNECE — *United Nations Economic Commission for Europe* — Komisja Gospodarcza Organizacji Narodów Zjednoczonych do spraw Europy.

¹⁰ Paradane — informacje objaśniające naukowe procesy poznawcze i interpretację danych statystycznych, a zatem np. sformalizowany opis metod wykorzystania owych danych liczbowych czy charakterystyka metodologicznych podstaw ich analizy. Paradane są ściśle powiązane z metadanymi i na nich się opierają. Stanowią przede wszystkim dokumentację procesu badawczego, który doprowadził do osiągnięcia danego rezultatu.

Model GSBPM stanowił więc podstawę konstrukcji struktury podręcznika, choć brano tutaj też pod uwagę specyfikę dziedziny, której ma on być poświęcony, jak również potencjalne zainteresowanie jego użytkowników poszczególnymi zagadnieniami. Między innymi opracowano następujący schemat:

ZESTAWIENIE PROJEKTOWANEJ STRUKTURY PODRĘCZNIKA

Nazwa rozdziału	Moduły charakteryzujące zawartość rozdziałów
Wprowadzenie	1. Główne wprowadzenie do podręcznika, schemat zawartości
Obserwacje ogólne	1. Metody i jakość 2. Rejestracja danych 3. Różne rodzaje badań 4. System statystyczny 5. Model GSBPM
Potrzeby użytkowników	1. Potrzeby użytkowników w zakresie systemów danych oraz specyficznych typów badań 2. Interpretacja różnych potrzeb i wymogów użytkowników 3. Ramy statystyki działalności gospodarczej w ESS 4. Specyfikacja planowanego produktu statystycznego 5. Ustalenie celów badań
Ogólny projekt badań	1. Główny moduł tematyczny (<i>Inne moduły do uzgodnienia</i>)
Jakość	1. Jakość statystyki 2. Kompleksowe zarządzanie jakością (TQM) 3. Ramy zabezpieczenia jakości (Quality Assurance) 4. Metody i narzędzia oceny jakości danych
Badania powtarzalne	1. Główny moduł tematyczny (<i>Inne moduły do uzgodnienia</i>)
Projektowanie koncepcji statystycznych	<i>Moduły do uzgodnienia</i>
Projektowanie kwestionariuszy	1. Główny moduł tematyczny 2. Kwestionariusz zbiorowy / moduły <i>ad hoc</i> 3. Projektowanie formularzy elektronicznych 4. Edycja podczas gromadzenia danych 5. Testowanie formularza
Rejestry i operaty statystyczne	1. Główny moduł tematyczny 2. Jednostki statystyczne i rejestr działalności gospodarczej 3. Jakość rejestrów i operatów
Dobór prób	1. Główny moduł tematyczny (w tym próbkowanie zbilansowane i podpróbki dla estymacji wstępnej) 2. Wprowadzenie do koordynacji próbkowania (koordynacja oparta na permutacji liczb losowych oraz próbkowaniu Poissona, koordynacja ze względu na obciążenia odpowiedzi, aspekty praktyczne, projektowanie metod próbkowania)
Gromadzenie danych wtórnych	1. Główny moduł tematyczny 2. Pomiar użyteczności rejestrów 3. Scenariusze zapasowe
Gromadzenie danych	1. Główny moduł tematyczny 2. Projektowanie metodologii gromadzenia danych (sposoby mieszane, sekwencyjne, równoległe, w tym z brakami odpowiedzi, definicja respondenta, strategie kontaktu itp.) 3. Zasady gromadzenia danych (dane administracyjne, alokacja tematów CATI ^a między ankierami, planowanie, aktualizacja i zarządzanie CATI, zarządzanie CAPI ^b , kwestionariusze papierowe i elektroniczne, transfer plików, obserwacje)

^a CATI — *Computer Assisted Telephone Interview* — wywiad telefoniczny wspomagany komputerowo.

^b CATI — *Computer Assisted Personal Interview* — wywiad osobisty wspomagany komputerowo.

ZESTAWIENIE PROJEKTOWANEJ STRUKTURY PODRĘCZNIKA (cd.)

Nazwa rozdziału	Moduły charakteryzujące zawartość rozdziałów
Odpowiedzi	1. Proces przetwarzania odpowiedzi 2. Obciążenie odpowiedzi
Mikrointegracja	1. Główny moduł tematyczny 2. Łączenie tych samych obserwacji z różnych źródeł 3. Łączenie różnych obserwacji z różnych źródeł 4. Rozwiązywanie problemów konfliktowych i dopasowania
Kodowanie	1. Główny moduł tematyczny 2. Jak zbudować bazę informacyjną 3. Różne metody i techniki kodowania 4. Różne strategie kodowania 5. Jak mierzyć jakość kodowania
Edycja	1. Główny moduł tematyczny 2. Korekcja dedukcyjna 3. Edycja selektywna 4. Edycja automatyczna 5. Edycja interaktywna 6. Makroedycja 7. Wykrywanie obserwacji odstających 8. Edycja danych administracyjnych 9. Edycja danych długookresowych 10. Projektowanie edycji
Imputacja	1. Główny moduł tematyczny 2. Imputacja dedukcyjna 3. Imputacja oparta na modelach zależności 4. Imputacja oparta na dawcach („gorąca”/„zimna”) 5. Imputacja oparta na dawcach przy użyciu mechanizmu „ruletki statystycznej” 6. Imputacja danych długookresowych 7. Imputacja w obliczu ograniczeń edycyjnych: podejście bezpośrednie 8. Imputacja w obliczu ograniczeń edycyjnych: metody dopasowania 9. Projektowanie imputacji
Pochodzenie jednostek statystycznych	<i>Moduły do uzgodnienia</i>
Ważenie i estymacja	1. Główny moduł tematyczny 2. Projektowanie estymacji 3. Ważenie, kalibracja, GREG ^c 3. Problem obserwacji odstających 4. Estymacja oparta na modelu 5. Estymacja z wykorzystaniem danych panelowych 6. Estymacja statystyk krótkookresowych 7. Estymacja dla małych obszarów (estymatory syntetyczne, złożone, EBLUP ^d)
Ocena jakości	1. Jakość statystyk 2. Metody estymacji wariancji 3. Błędy badań w przypadku losowania nieprobabilistycznego 4. Efekty błędów nielosowych a ogólny wynik badania 5. Projektowanie efektów dla losowania zespołowego 6. Analiza rewizji 7. Terminowość i punktualność 8. Spójność i porównywalność
Makrointegracja	1. Główny moduł tematyczny 2. Metoda RAS ^e 3. Metoda Stone’a 4. Metoda Dentona dla benchmarkingu 5. Metoda Dentona dla tymczasowej dezagregacji

c GREG — *Generalized Regression Estimator* — uogólniony estymator regresyjny.

d EBLUP — *Empirical Best Linear Predictor* — najlepszy liniowy predyktor nieobciążony.

e RAS — *Remote Access Service* — usługa zdalnego dostępu.

ZESTAWIENIE PROJEKTOWANEJ STRUKTURY PODRĘCZNIKA (dok.)

Nazwa rozdziału	Moduły charakteryzujące zawartość rozdziałów
Dopasowanie sezonowe	1. Wprowadzenie 2. Opis ogólny (komponenty szeregów czasowych, schematy dekompozycji, efekty kalendarzowe, wpływ obserwacji odstających) 3. Główne zasady (dopasowanie bezpośrednie i pośrednie, spójność czasowa, rewizja) 4. Metody i narzędzia wyrównań sezonowych (analiza wstępna, dekompozycja, np. X12 ARIMA, TRAMO/SEATS) 5. Zagadnienia dotyczące dopasowań (jakość, dokumentacja, prezentacja danych) 6. Zagadnienia specjalne (alternatywne metody redukcji wpływu obserwacji odstających, kryzysy, szeregi czasowe wsteczne i wyprzedzające, szeregi wielowymiarowe itp.) 7. Studium przypadków, najlepsze rozwiązania praktyczne
Kontrola ujawniania danych statystycznych	1. Kontrola ujawniania dla mikrodanych (globalne przekodowanie, lokalne obostrzenia, kodowanie górne/kodowanie dolne, dodawanie szumu do wag w badaniach reprezentacyjnych, PRAM^f) 2. Kontrola ujawniania dla tablic (restrukturyzacja, obostrzenia w zakresie komórek danych, zaokrąglenia, miary wrażliwości dla identyfikacji niebezpiecznych komórek danych)
Rozpowszechnianie	<i>Moduły do uzgodnienia</i>
Ocena	<i>Moduły do uzgodnienia (zakres i cele, zasady, wskazówki dotyczące oceny jakości danych)</i>

f PRAM — *Program for Reliability Assessment with Multiple Coders* — program oceny rzetelności z wielokrotnym kodowaniem.

Źródło: opracowanie własne na podstawie dokumentu *Work Plan for the Memobust Project* (last modified: 9 September, 2011), z wykorzystaniem późniejszych uzgodnień.

Schemat ten nie jest ostateczny i może być nieco zmodyfikowany w trakcie dalszych prac. Ewentualne zmiany dotyczyłyby wprowadzenia dodatkowych modułów do danego rozdziału. Nie wyklucza się także i tego, że niektóre zagadnienia nie pozostaną *tematami* w rozumieniu nowej koncepcji podręcznika — i mogą być podtematami — a niektóre dotychczasowe podtematy mogą stać się odrębnymi tematami. Prawdopodobne będzie ponadto wprowadzenie do podręcznika opisu fazy przygotowawczej badań. W tym kontekście istotne jest osiągnięcie funkcjonalności wypracowanych rezultatów oraz możliwości późniejszego uzupełniania niektórych pozycji przez specjalistów, a także łączenia różnych metod z rozmaitych modułów w pojedynczy moduł, który w efekcie zawierałby opis różnych wariantów tego samego narzędzia teoretycznego — i odwrotnie, tzn. dzielenia wariantów jednego modułu na kilka odrębnych modułów.

ORGANIZACJA PRAC NAD PODRĘCZNIKIEM

Ze względu na obszerność i różnorodność treści, które mają być zawarte w podręczniku, jego opracowanie będzie szczególnie złożone. Do pisania i recenzji przygotowano specjalne szablony. W trakcie opracowywania treści pod-

ręcznika stosuje się dwa typy owych wzorców: jeden dla głównego tematu z subtematami oraz odrębny dla opisu konkretnej metody statystycznej. W celu sprawnego pisania do każdego tematu wyznaczono kraj wiodący, który będzie koordynował prace redakcyjne i recenzyjne — z prawem do ewentualnego znalezienia, w razie potrzeby, innych autorów do określonych podtematów lub konkretnych metod. Kraj wiodący stanie się autorem głównego modułu tematycznego oraz — w drodze dyskusji z krajami współautorskimi — podejmie decyzje o kształcie podtematów i o opisywanych metodach.

Zasadniczymi, wspólnymi elementami wzorców tematu i metody są: nazwy identyfikacyjne, spis treści, streszczenie, główny blok opisowy, pojęcia do słownika metodologicznego oraz syntetyczne hasłowe informacje na temat wykorzystywanych metod teoretycznych, uwarunkowań stosowania czy merytorycznych powiązaniach z innymi modułami. Kwestia, w jaki sposób koordynować owe powiązania, pozostaje jednak ciągle otwarta, podobnie jak ewentualne rozszerzenie roli słownika jako źródła definicji, z powiązaniem lub lokalnymi koncepcjami i synonimami zamieszczonymi w dodatkowej kolumnie, jeśli dany termin ma kilka znaczeń. Każdy podtemat winien zawierać także opis odpowiedniego etapu projektowania badania statystycznego, do którego się odnosi — zgodnie ze schematem GSBPM. Podział prac nad większością projektowanych modułów został już dokonany, autorzy pozostałych zagadnień będą wyłonieni w drodze wzajemnych uzgodnień w drugiej fazie projektu.

W podręczniku zostaną wykorzystane także specyficzne doświadczenia niektórych partnerów projektu, np.:

- dorobek niderlandzkiego projektu *Method Series* (opisanego m.in. w 30 artykułach zamieszczonych na stronie CBS), którego celami są: dokumentacja efektywnych i stosowanych metod statystycznych, standaryzacja stosowanych już sposobów, edukacja oraz użyteczność dla zewnętrznych użytkowników;
- ocena obciążenia respondentów (*Perceive Response Burden*) — rozwiązanie stosowane w Szwecji;
- oprogramowanie *CherryPi* służące do automatyzacji edycji danych gospodarczych (Niderlandy);
- wyniki badania satysfakcji użytkowników informacji statystycznych według ich grup oraz kategorii tychże informacji (Grecja);
- kanadyjski system edycji i imputacji danych GEIS (*Generalized Edit and Imputation System*) wykorzystywany także w Europie.

Podczas opracowywania publikacji wzięta będzie pod uwagę różnorodność rozwiązań i problemów występujących w poszczególnych krajach. Możliwe jest przeprowadzenie w tym celu bezpośrednich konsultacji z potencjalnymi respondentami. Podczas pisania należy zachować obiektywizm, tzn. nie sugerować potencjalnemu czytelnikowi wyboru konkretnych metod, pozostawiając jemu te decyzje, naświetlając za to zalety i wady rozwiązań w odpowiednich sytuacjach praktycznych.

Recenzje winny weryfikować zgodność podręcznika ze standardami GSBPM i SDMX oraz przyczyniać się do bieżącej redakcji słownika metodologicznego. Pierwszy (nieformalny) etap recenzji odbędzie się w kraju autorskim. Formalnymi recenzentami zaś będą kraje uczestniczące w projekcie, ale nie opracowujące danego modułu. Zaproponowano przetestowanie wzorca recenzji na istniejących już modułach, zastanowienie się nad tym, kto może być recenzentem (a w szczególności rozważenie możliwości włączenia do tego procesu osób spoza projektu). Nie da się przy tym wykluczyć możliwości, iż autor nie przyjmie zastrzeżeń recenzenta. Szablon recenzyjny, stosowany dla sformalizowania oceny tekstu, zawiera pytania dotyczące jakości opracowanego modułu od strony formalnej, merytorycznej, stopnia zrównoważenia poruszanych tematów, jakości słownika, zgodności ze standardami SDMX i GSBPM, językowej, przejrzystości prezentacji, czytelności dla niespecjalistów itp. Dużym wyzwaniem dla recenzenta i piszącego dany moduł jest to, w jaki sposób praca winna odzwierciedlać specyfikację wzorców tematów i metod. Truizmem wydaje się więc stwierdzenie, iż dla rzetelności recenzji ważna jest efektywna strukturyzacja zagadnień oraz właściwa hierarchia ich ważności i użyteczności w zakresie tematów oraz odpowiednich podtematów.

Planuje się ponadto stworzenie i rozwijanie treści oraz funkcjonalności specjalnej strony internetowej projektu — będzie ona odgrywać ważną rolę w zapewnieniu ciągłości realizacji strategii rozpowszechniania wyników podjętych działań. Strona ta zostanie także wykorzystana jako narzędzie komunikacji pomiędzy współpracującymi partnerami celem opiniowania i komentowania prowadzonych prac. Wstępna wersja takiej strony wraz z pierwszymi informacjami znajduje się na portalu ESS pod adresem <http://www.essnet-portal.eu/memobust-0>. Nie przewiduje się jednak „wikipedyzacji” podręcznika, ponieważ program koncentruje się na opracowaniu merytorycznym, a przekształcenie takie wymagałoby przeprowadzenia przez doświadczonych programistów odpowiednich wyspecjalizowanych i kosztownych prac. Inicjatywa ta mogłaby jednakże w okresie późniejszym ewentualnie stać się celem jakiegoś innego projektu w ramach MEETS.

POLSKI WKŁAD WNOSZONY DO REALIZACJI PROJEKTU

Tworzenie nowej wersji podręcznika odbywa się z aktywnym udziałem statystyki polskiej. Główną rolę wykonawczą i merytoryczną odgrywa Urząd Statystyczny w Poznaniu wspierany przez Centrum Informatyki Statystycznej. Działania te polegają na autorstwie tekstu oraz recenzowaniu fragmentów powstającego dzieła. Polska była także organizatorem trzeciego spotkania zespołu realizującego projekt, które odbyło się 6 października 2011 r. w Poznaniu. Przy wyborze modułów, które tu mają być opracowywane bądź recenzowane, kierowano się doświadczeniem w statystyce krótkookresowej, informatyzacji sprawozdawczości, estymacji dla małych obszarów, wielowymiarowej analizy danych itp.

Strona Polska jest liderem tworzenia dwóch rozdziałów: „Projektowanie kwestionariuszy” oraz „Odpowiedzi”. W pierwszym przypadku jest to autorstwo niemal wszystkich modułów (jedynie w przypadku testowania formularzy przewidziany jest wkład strony włoskiej).

Główny moduł tematyczny zawiera ogólną charakterystykę i koncepcje, które mogłyby być wykorzystane przy projektowaniu formularzy. Wykorzystane tu zostaną źródła międzynarodowe, takie jak *Handbook of Recommended Practices for Questionnaire Development* (Podręcznik Praktyk Zalecanych dla Doskonalenia Formularzy) czy *Testing in European Statistical System* (Testowanie w Europejskim Systemie Statystycznym). Na tej ostatniej pozycji (oraz poprzedniej wersji podręcznika) oparta jest także część dotycząca testowania formularzy (zawierająca m.in. charakterystykę zasad i testowania, miejsca etapu testowania w strukturze oprogramowania oraz rodzajów procedur testowych).

Moduł zatytułowany „Kwestionariusz zbiorowy/moduły *ad hoc*” poświęcony zostanie pracy w obliczu różnorodnych potrzeb klientów i środków gromadzenia danych (w tym konsolidacji badań), jak również konstrukcji formularzy międzydziedzinowych oraz spersonalizowanych. Projektowanie formularzy elektronicznych obejmuje tworzenie kwestionariuszy samoadministrujących czy administrowanych przez użytkownika oraz opracowywanie instrukcji, jak też dopasowywanie formularzy do właściwości różnych środków gromadzenia danych i oprogramowania, a także ocenę i testowanie.

W zakresie tematycznym dotyczącym edycji podczas uzyskiwania danych znajdują się następujące zagadnienia: udokumentowane reguły edycji danych, personalizacja, zasady edycji (w tym formularzy elektronicznych) stosowane w obliczu różnych specyficznych sytuacji, jak np. brakujące dane czy obserwacje odstające.

Z kolei w module poświęconym obciążeniom odpowiedzi przedstawiono klasyfikację według rodzaju, czynników wpływających na te obciążenia według miejsca ich powstawania, alternatywne metody klasyfikacji błędów badawczych, praktyczne problemy i przykłady kłopotów w zakresie statystyki działalności gospodarczej oraz działania na rzecz redukcji tych niedogodności, a także standardowy model kosztów badań. Odpowiedni rozdział powinien być uzupełniony o część dotyczącą przetwarzania i analizy jakości odpowiedzi.

Istotne treści merytoryczne — w tym naukowe — wnosi Polska do rozdziału poświęconego imputacji (nad którym prace koordynuje strona niderlandzka). Opracowano moduł poświęcony imputacji opartej na modelach, w którym przedstawiono trzy główne metody tego typu. Pierwsza z nich to imputacja średnią (ang. *mean imputation*), gdzie brakujące dane zastępuje się średnią z dostępnych obserwacji (również w postaci ważonej — C. Arcaro and W. Yung (2001)). Druga to imputacja ilorazowa, w której wykorzystuje się ściśle powiązanie badanej zmiennej z inną zmienną, dla której dane są w pełni dostępne, zakładając, że iloraz średnich dostępnych wartości obu zmiennych jest taki sam jak dla poszczególnych brakujących danych. Opcja ta występuje też w postaci

ważonej. Trzecią możliwość stanowi imputacja regresyjna, gdzie uzupełnianie luk przeprowadza się wykorzystując modele regresji, skonstruowane dla zmiennej z lukami na podstawie jej dostępnych danych oraz dane dla zmiennych pomocniczych silnie z nią powiązanych. Można tu wyróżnić regresję wieloraką z rozszerzeniem iteracyjnym lub predykcyjnym parowaniem średnich (*predictive mean matching*) czy metodę opartą na rezultacie skłonności do braku danych opisaną odpowiednim modelem logistycznym (*propensity score method*) (Yuan, 2010), jak również specjalne modele iteracyjne (Templa i in., 2011) czy algorytmy łączone i sekwencyjne (Chaimongkol, Suwattee, 2004). W zakresie imputacji opartej na dawcach danych (czyli rekordach z pełną informacją) skoncentrowano się na imputacji losowej (*random hot deck*) oraz niektórych metodach deterministycznych, prowadzących do wyboru najlepszego dawcy — w tym rangowej (*rank hot-deck*), jak również opartej na odległościach biorców (rekordów, dla których dokonywana będzie imputacja) od dawców (*distance hot-deck*). Wykorzystano tutaj niektóre narzędzia parowania danych (D’Orazio i in., 2006), syntetyczny przegląd dokonany przez R. R. Andridge i R. J. A. Little (2010) oraz informacje taksonomiczne z książki A. Młodaka (2006).

Oba moduły zaopatrzone w przykłady imputacji dokonywanej zarówno na danych symulowanych, jak i konkretnych informacjach ze statystyki działalności gospodarczej, zaczerpniętych z Internetu. Odrębnego modułu doczekała się specyficzna metoda oparta na mechanizmie tzw. „ruletki statystycznej”, zaproponowana przez B. Stefanowicza (2009) i zastosowana do analiz do celów Narodowego Spisu Powszechnego Ludności i Mieszkań 2011 (Młodak, 2010a, 2010b), a łącząca grupowanie dawców z wykorzystaniem analizy skupień, wybór reprezentantów poszczególnych grup, odległość biorcy od reprezentanta oraz losowy wybór konkretnego dawcy z odpowiedniej grupy. Zaznaczono także sposób korzystania z imputacji na podstawie pomocniczych danych archiwalnych. Swoistym podsumowaniem tych rozważań jest część dotycząca projektowania imputacji, w której zawarto zasady stosowania poszczególnych rozwiązań, doboru zmiennych pomocniczych, oceny dopasowania modelu oraz efektów zakłóceń.

W ramach części wstępnej („Obserwacje ogólne” — liderem są Niderlandy) z Polski wywodzi się moduł poświęcony różnym rodzajom badań statystycznych. Oprócz tego Polska staje się także autorem rozdziałów poświęconych waznieniu (podstawowemu, opartemu na schemacie oraz dopasowaniu i kalibracji wag) i estymacji (przeprowadzanie szacunków, estymacja punktowa, estymacja oparta na modelu, estymacja statystyk krótkookresowych, dla małych obszarów oraz problem obserwacji odstających w kontekście jakości estymacji). W obu przypadkach prace prowadzone są pod kierunkiem Włoch. Przygotowano także części poświęcone estymacji wariancji, błędom badań w przypadku losowania nieprobabilistycznego, efektom błędów nielosowych oraz spójności i porównywalności. Omówiono tam m.in. czynniki wpływające na estymację wariancji, metody estymacji wariancji dla różnych typów badań, dziedzin i klas oraz bada-

nia z uwzględnieniem czynnika czasowego (powtarzalności), jak również wpływ błędów operatu, pomiaru przetwarzania, wynikających z braku odpowiedzi, a także oddziaływania modelu w badaniach reprezentacyjnych na jakość danych wynikowych. Należą one do rozdziału „Ocena jakości” — tu funkcję lidera sprawuje Grecja. Podczas redagowania tej części wykorzystanie znalazły też doświadczenia polskie z zakresu wielowymiarowej analizy danych, w tym metod taksonomicznych (dotyczy to zapewnienia porównywalności różnorodnych cech diagnostycznych i konstrukcji miernika kompleksowego).

Polska będzie także recenzentem (lub współrecenzentem) następujących części: model GSBPM (w ramach „Podstawowych informacji”), potrzeby użytkowników (lider — Grecja), części modułów w koordynowanym przez Norwegię rozdziale poświęconym doborowi prób (chodzi o koordynację losowania prób, losowania probabilistyczne i nieprobabilistyczne oraz koordynację badań z wykorzystaniem różnego rodzaju jednostek statystycznych, a także losowanie zbilansowane i warstwowanie wielomodelowe) oraz rozdziałów: „Gromadzenie danych wtórnych” (lider — Niderlandy). Strona polska gotowa jest również do włączenia się w działania promocyjne dotyczące podręcznika (stworzenie stron zwierciadlanych na portalach urzędów statystycznych, prowadzenie szkoleń *on-line* czy publikacja dzieła w zwartej formie). Wszystkie te prace mogą istotnie przyczynić się do doskonalenia narzędzi statystycznych i ich praktycznego zastosowania również w naszym kraju.

dr hab. Andrzej Młodak — *Urząd Statystyczny w Poznaniu, Ośrodek Statystyki Miast, Państwowa Wyższa Szkoła Zawodowa im. Prezydenta Stanisława Wojciechowskiego w Kaliszu*

LITERATURA

- Andridge R. R., Little R. J. A. (2010), *A Review of Hot Deck Imputation of Survey Non-response*, „International Statistical Review”, vol. 70
- Arcaro C., Yung W. (2001), *Variance estimation in the presence of imputation*, SSC Annual Meeting, Proceedings of the Survey Method Section
- Chaimongkol W., Suwattee, P. (2004), *Weighted Nearest Neighbor and Regression Imputation*, 9th Asia-Pacific Decision Sciences Institute Conference, APDSI-KOPOMS Seoul, Korea, 1—4 July 2004, Seoul, Korea
- D’Orazio M., Di Zio M., Scanu M. (2006), *Statistical Matching. Theory and Practice*, John Wiley & Sons, New York
- Generic Statistical Business Process Model* (2009), Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS), Version 4.0, United Nations Economic Commission for Europe (UNECE), UNECE Secretariat, <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.61/2009/mtg1/zip.32.e.pdf>
- Młodak A. (2010a), *Metodologiczne problemy imputacji danych w spisach powszechnych*, „Kwartalnik Statystyczny”, R. XII, nr 4
- Młodak A. (2010b), *Imputacja danych w spisach powszechnych*, „Wiadomości Statystyczne”, R. LV, nr 8

- Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej*, Centrum Doradztwa i Informacji DIFIN, Warszawa
- SDMX *Content-Oriented Guidelines. Annex 4: Metadata Common Vocabulary 2009* (2009), http://sdmx.org/wp-content/uploads/2009/01/04_sdmx_cog_annex_4_mcv_2009.pdf
- Stefanowicz B. (2009), *Imputacja danych statystycznych* (maszynopis)
- Templa M., Kowarik A., Filzmoser P. (2011) *Iterative stepwise regression imputation using standard and robust methods*, „Computational Statistics & Data Analysis”, vol. 55
- Willeboordse A. (red.) (1997), *Handbook on the Design and Implementation of Business Surveys*, Eurostat, Luxembourg
- Yuan Y. C. (2010), *Multiple Imputation for Missing Data: Concepts and New Development (Version 9.0)*, SAS Institute Inc Rockville, MD, U.S.A

SUMMARY

The paper discusses basic purposes and executive assumptions of the project leading to the modernization of methodology of statistical surveys related to business statistics and an assessment of quality of their results realized within the framework of the European Statistical System „Methodology for Modern Business Statistics” (ESSnet MeMoBuSt). This is venture of eight countries: the Netherlands, Hungary, Poland, Sweden, Switzerland, Norway, Italy and Greece. The group operates from 2011 and intends to release a new version of the methodological handbook concerning these issues in 2013. Specialists from the eight countries will take into account the newest scientific achievements in this field and work out directions of their effective application. The current methodological handbook has been in 1997 and the group focuses on disadvantages and gaps existing in the information contained there in relation to modern realities and challenges. Next component of this presentation is a characteristics of assumptions of the current version of the handbook. It includes also a degree of its consistency with commonly used in European and world statistics scheme of organization of statistical surveys GSBPM (Generic Statistical Business Process Model), adopted by UN, OECD and Eurostat. A description of current statements and divergences concerning detailed vision of the content of particular chapters and review of the already written fragments are discussed. The paper is supplemented with an analysis of the role of the Polish part in this venture and the issues which our country has contributed to this project and which it can contribute in the future.

РЕЗЮМЕ

В статье характеризуются основные цели и реализационные основы проекта модернизации методологии проведения обследований в области экономической статистики и оценки качества их результатов реализованного в рамках программы-гранта Европейской статистической системы «Методология современной разработки статистики пред-

приятый» (ESSnet MeMoBuSt). Задачей этого, запланированного на 2011—2013 гг, мероприятия восьми стран: Голландии, Венгрии, Польши, Швеции, Норвегии, Швейцарии, Италии и Греции является разработка новой версии методологического пособия, учитывающего последние научные достижения, а также направления их эффективного использования. В статье было представлено содержание, разработанного еще в 1997 г., пособия — с указанием его недостатков по отношению к современным реалиям и вызовам. Следующую часть разработки составляет характеристика проекта нового пособия. Она охватывает также степень соответствия его содержания с обычно используемой в европейской и мировой статистике схемой организации статистических обследований GSBPM (Основная модель проведения статистики экономической деятельности), принятой ООН, ОЭСР и Евростатом. Дальше представляются актуальные определения и расхождения по отношению к детальному просмотру содержания отдельных разделов (глав), а также обсуждаются части пособия, которых разработка уже закончена. Статья подчеркивает роль польской стороны в реализации этого мероприятия.