

Jan Paradysz

Nowe źródła danych w klasycznym
paradygmacie informacji statystycznej

STATYSTYKA – WIEDZA – ROZWÓJ
KONFERENCJA Z OKAZJI MIĘDZYNARODOWEGO ROKU
STATYSTYKI – ŁÓDŹ 17-18 X 2013 roku

Plan prezentacji

1. Wprowadzenie
2. Doświadczenie po PSR 2010 oraz NSP 2011
3. Internetowe źródła danych o gospodarce nieelektronicznej
4. Internetowe źródła danych o gospodarce elektronicznej
5. Sondaż internetowy jako źródło informacji wspomagających w wybranych badaniach reprezentacyjnych
6. Problemy i wyzwania przed statystyką publiczną w zakresie sondażu internetowego
7. Wnioski

Coś na kształt wprowadzenia

Celem referatu jest zwrócenie uwagi na nowe źródła informacji statystycznej, które się pojawiły dzięki ostatnim dwóm wielkim badaniom w Polsce - Powszechny Spis Rolny 2010 (PSR 2010) oraz Narodowy Spis Ludności i Mieszkań (NSP 2011) – oraz na możliwości dalszego wykorzystania źródeł administracyjnych. Obydwa badania sprawiły, że Polska dołączyła do krajów europejskich o najlepiej zorganizowanej statystyce publicznej. Jest to jednak dopiero początek drogi. Doświadczenie zdobyte w tych badaniach oraz przełamanie szeregu barier powinno być utrwalane w codziennej działalności GUS. Niektóre ze źródeł, o wielkim znaczeniu dla polskiej gospodarki, zostały wykorzystane w PSR 2010 i NSP 2011 w niewielkim tylko zakresie. Przykładem tego rodzaju braków w pokryciu informacyjnym naszego kraju są kompletnie dotąd niewykorzystane zasoby Ministerstwa Finansów.

Coś na kształt wprowadzenia

Dużym wyzwaniem dla statystyki publicznej są nowe informacje gospodarcze powszechnie dostępne w cyberprzestrzeni. Wskazują one na coraz większy zasięg gospodarki elektronicznej oraz na coraz większe pokrycie informacyjne w Internecie tradycyjnej gospodarki „w realu”. Jak pokazują liczne opracowania z zakresu statystyki urzędowej – por. artykuły w Journal of Official Statistics – eksploracja internetowych źródeł danych staje się jednym z największych wyzwań dla współczesnej statystyki publicznej.

Czego nie ma w Internecie, to to nie istnieje. Zatem można powiedzieć, że w dobie lawinowo rosnącej informacji *mamy coraz większe braki w pokryciu informacyjnym polskiej gospodarki.*

Doświadczenie po PSR 2010 oraz NSP 2011

Na czym polegają doświadczenia z obu spisów i co to może mieć wspólnego z sondażem internetowym?

1. Ocena dotychczasowych tradycyjnych spisów powszechnych
2. Wykonanie olbrzymiej pracy nad zewidencjonowaniem, oceną jakości oraz dostosowaniem do celów statystyki publicznej
3. Poruszenie olbrzymich pokładów innowacyjności tkwiących pracownikach statystyki publicznej
4. WYROBIE NIE U OSÓB ZAANGAŻOWANYCH W PRACACH PRZYGOTOWAWCZYCH DO SPISÓW NAWYKÓW WYKORZYSTYWANIA WIELU ŹRÓDEŁ W MYŚL ZASADY, ŻE NIE MA ŹRÓDEŁ IDEALNYCH, ALE WIELE Z NICH ZNAJDUJE SIĘ NA CAŁKIEM PRYZWOITYM POZIOMIE
5. Podjęcie konkretnych prac nad wykorzystaniem najnowszych osiągnięć z zakresu metody reprezentacyjnej (kalibracja) i estymacji pośredniej
6. Tradycyjne spisy byłyby bezradne wobec pojawienia się nowych zjawisk wynikających z postępu technologicznego i rozwoju polskiej gospodarki
7. Czy to może mieć coś wspólnego z sondażem internetowym – postaramy się pokazać, że tak.

Doświadczenie po PSR 2010 oraz NSP 2011

Jakie są wyzwania?

1. Dalsze prace międzyresortowe nad doskonaleniem funkcji statystycznych rejestrów administracyjnych
2. Ocena wyników obu spisów powszechnych
 - Według zaproponowanych w 2008 r. kryteriów formalnych i merytorycznych
 - W nawiązaniu do analogicznych prac porównawczych w krajach o większych doświadczeniach w przeprowadzaniu spisów opartych na rejestrach
3. Prace nad problemami związanymi z migracjami międzynarodowymi
4. Rozwój prac metodologicznych w zakresie integracji baz danych
5. Rozwój prac metodologicznych w zakresie estymacji pośredniej
6. Pełniejsze wykorzystanie rejestrów podatkowych
7. Poszukiwania nowych źródeł zasilania informacyjnego, w tym także w cyberprzestrzeni.

Internetowe źródła danych o gospodarce nieelektronicznej (tradycyjnej)

Czy możemy być usatysfakcjonowani dotychczasową informacją o cenach?

1. Ceny produktów i usług są oknem, przez które statystyka publiczna może mieć wgląd na kondycję ekonomiczną przedsiębiorstwa
2. Krytyka tradycyjnych źródeł informacji o cenach produktów i usług
 - Zupełnie nie są wykorzystane zasoby kas fiskalnych
 - Brak informacji o cenach transakcyjnych
 - Trudności powiązania cen ofertowych z wielkością produkcji i sprzedaży
3. Możliwości, jakie daje sondaż internetowy dla niektórych dziedzin życia
 - Ceny transakcyjne pokoi hotelowych, połączeń lotniczych, imprez artystycznych itd.
 - Wolumin sprzedaży dóbr i usług oferowanych w gospodarce nieelektronicznej
 - Ceny ofertowe nieruchomości i trudności pozyskiwania informacji w tym zakresie
 - Chlubny wyjątek na rynku nieruchomości – holenderska <http://fonda.nl/>

Internetowe źródła danych o gospodarce elektronicznej

Coraz większe rozmiary aktywności ekonomicznej w Internecie

1. Przedsiębiorstwa w cyberprzestrzeni

- Allegro
- Amazon
- e-Bay
- Fevad
- Booking.com

2. Rodzaje aktywności internetowej

- B2B – między przedsiębiorstwami.
- B2C – między przedsiębiorstwem a konsumentem
- C2C – między konsumentami

3. Obecność w Internecie „realnych” podmiotów gospodarczych i instytucji

- Banki, sklepy, przedsiębiorstwa produkcyjne, związki wyznaniowe, partie polityczne

Sondaż internetowy jako źródło informacji wspomagających w wybranych badaniach reprezentacyjnych

Zasoby internetowe mogłyby być źródłem wspomagającym badania ankietowe

1. W zakresie gospodarki hotelowej

- Obłożenie łóżek hotelowych
- Ceny transakcyjne
- Położenie w przestrzeni miejskiej
- Trendy i cykliczność
- Konkurencja między hotelami pozycja rynkowa
- Oceny klientów

2. W zakresie turystyki

- Ruch turystyczny i mobilność przestrzenna
- Analiza dynamiki i cykliczności ruchu turystycznego

Sondaż internetowy jako źródło informacji wspomagających w wybranych badaniach reprezentacyjnych

Zasoby internetowe mogłyby być źródłem wspomagającym badania ankietowe

3. W zakresie rynku nieruchomości na rynku wtórnym
 - Struktura nieruchomości,
 - Dynamika i koniunktura gospodarcza

Problemy i wyzwania przed statystyką publiczną w zakresie sondażu internetowego

1. Potencjalne rodzaje operatu losowania w zakresie gospodarki hotelowej i turystycznej
 - klienci
 - hotele
 - pokoje bądź łóżka hotelowe
 - operaty przestrzenne
2. Schematy doboru próby
3. Rodzaje estymacji,
4. Częstość pobierania informacji
5. Reakcja przedmiotu badań
 - Pozytywna w gospodarce hotelarskiej, sprzedaż biletów na środki komunikacji i imprezy sportowo-kulturalne
 - Negatywna na rynku nieruchomości

Wnioski

1. Wysokie standardy, które wyznaczyły polskiej statystyce publicznej PSR 2010 i NSP 2011 należy nie tylko utrzymać, ale zwiększać poprzez poszerzanie zakresu informacji pozyskiwanych z rejestrów administracyjnych, w szczególności podatkowych
2. Sondaż internetowy rozumiany jako pozyskiwanie informacji istniejących w Internecie miałby już teraz duże znaczenie dla statystyki publicznej, zatem ...
 - Powinno się rozpocząć prace studialne nad pozyskiwaniem i wykorzystaniem pełnowartościowych i wiarygodnych informacji w Internecie o życiu gospodarczym Polski.
 - W niektórych przypadkach należy założyć systematyczną obserwację internetowej aktywności podmiotów życia gospodarczego zgodnie z regułami jakości danych.
 - Reprezentatywność danych pozyskiwanych w Internecie
 - Doskonalenie estymacji pośredniej
 - Łączenie źródeł internetowych z pozostałymi informacjami