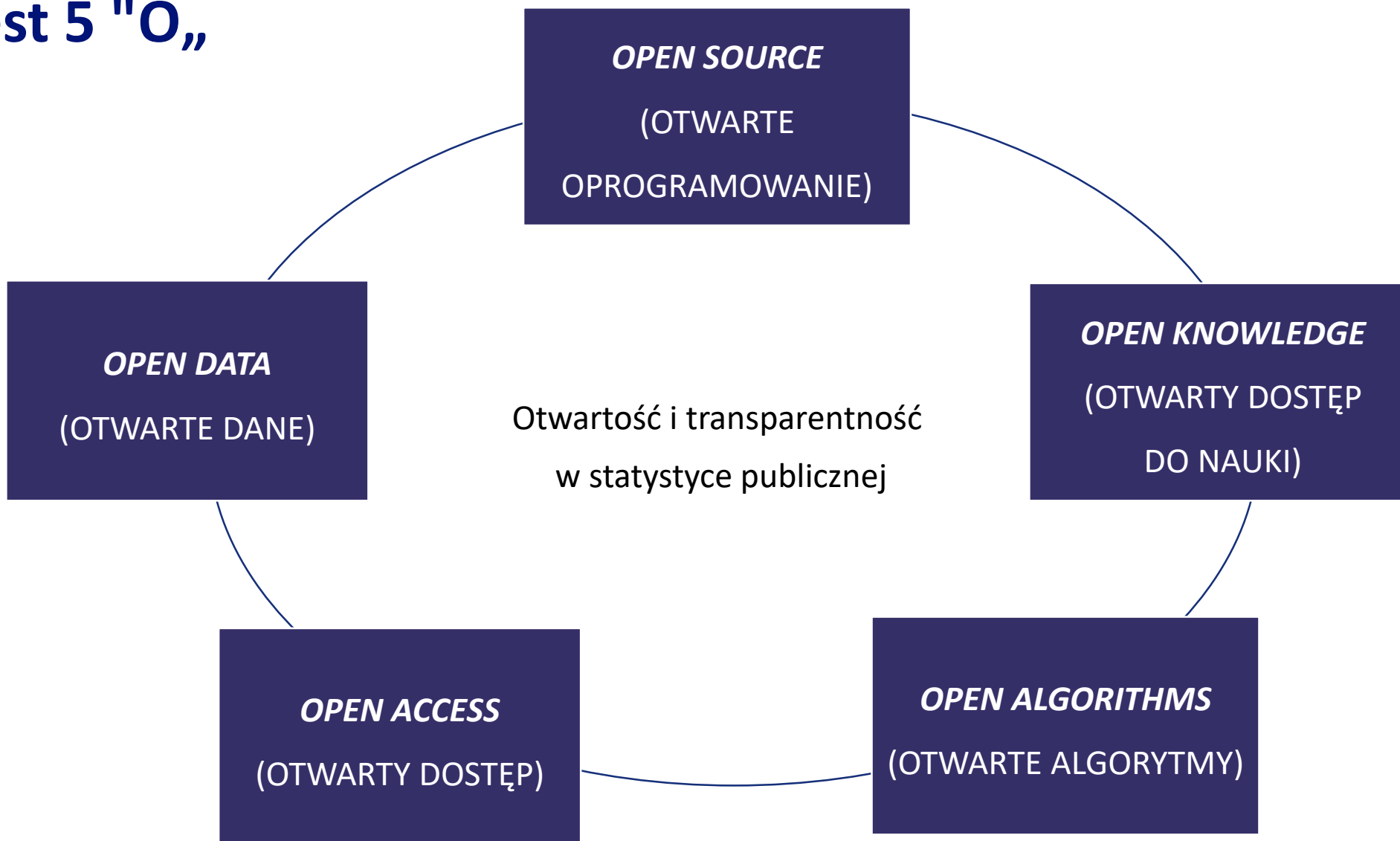


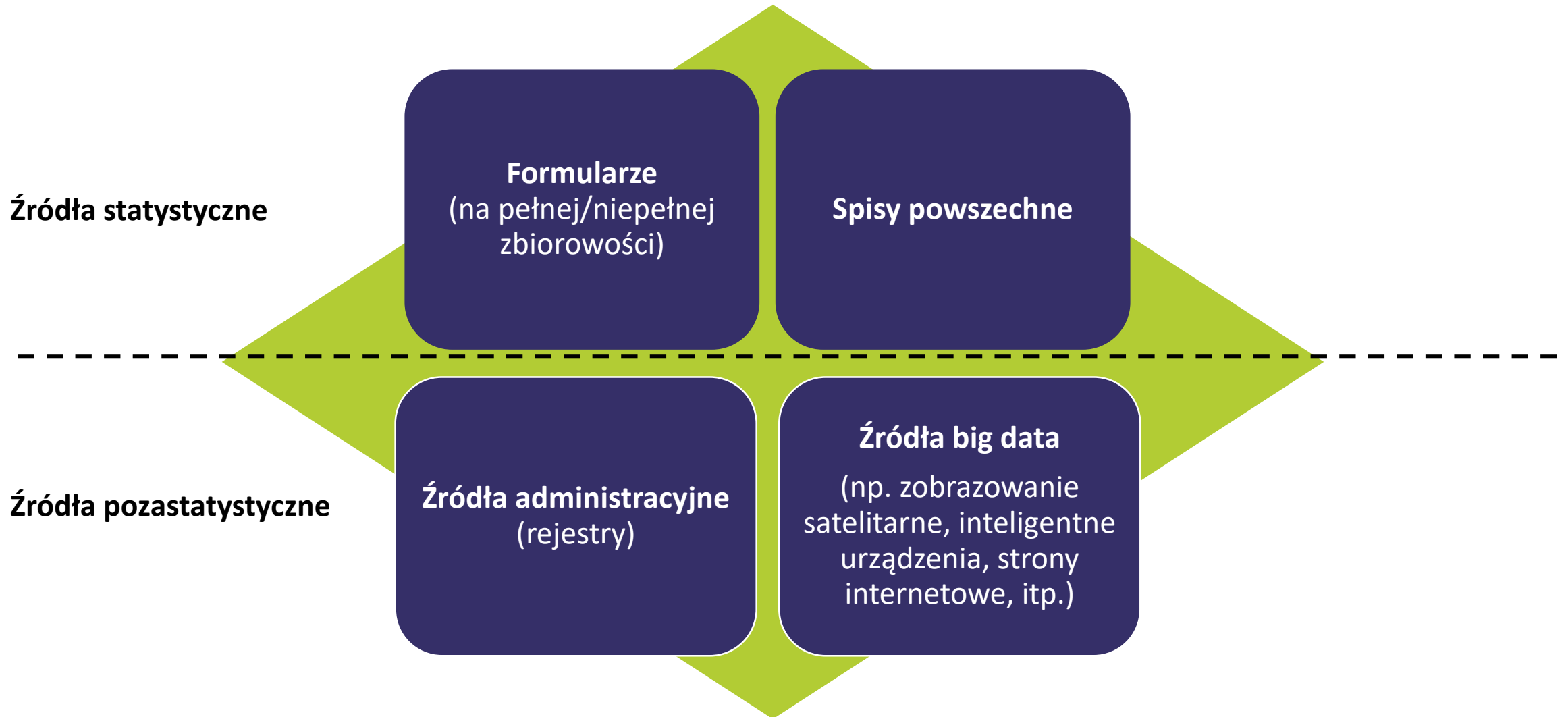
Wtórne wykorzystanie danych administracyjnych w statystyce publicznej – obecnie i w przyszłości

Departament Badań Przestrzennych i Środowiska GUS
Dominika Rogalińska

Manifest 5 "O,,



Źródła danych w statystyce publicznej



Źródło administracyjne – definicje ustawowe

REJESTRY URZĘDOWE

- rejestr publiczny – rejestr, ewidencja, wykaz, lista, spis albo inna forma ewidencji, służące do realizacji zadań publicznych, prowadzone przez podmiot publiczny na podstawie odrębnych przepisów ustawowych;
- inne rejestry i ewidencje prowadzone na podstawie ustaw lub wydanych na ich podstawie aktów wykonawczych, zawierające informacje o podmiotach gospodarki narodowej i ich działalności, informacje o osobach fizycznych, ich życiu i sytuacji oraz dotyczące zjawisk, zdarzeń i obiektów
- systemy zbierania, gromadzenia i przetwarzania informacji przez organy administracji publicznej, Zakład Ubezpieczeń Społecznych, Narodowy Fundusz Zdrowia, Komisję Nadzoru Finansowego, organy rejestrowe, inne państwowe lub samorządowe osoby prawne oraz inne podmioty prowadzące rejestry urzędowe

SYSTEMY INFORMACYJNE ADMINISTRACJI PUBLICZNEJ

Źródła administracyjne w statystyce publicznej

SZANSE

- Zmniejszenie kosztów badań i obciążenia respondentów, poprawa efektywności tworzenia statystyk
- Skrócenie czasu dostarczenia danych do odbiorców, analizy nawet w czasie (prawie) rzeczywistym
- Zwiększenie dokładności i aktualności danych wynikowych i oszacowań

WYZWANIA

- Integracja systemu statystyki publicznej z systemami administracyjnymi
- Odejście od tradycyjnych spisów ludności i mieszkań na rzecz spisów opartych na rejestrach i spisów łączonych
- Dostosowanie rejestrów do potrzeb związanych z prowadzeniem badań statystycznych

ZAGROŻENIA

- Występujący (sporadycznie) brak spójności z analogicznymi źródłami na poziomie regionalnym/lokalnym
- Systemy regionalne/lokalne rozproszone i utrzymywane według różniących się między sobą standardów metodycznych i informatycznych
- Utrzymanie źródeł administracyjnych i dbałość o ich regularne zasilanie to wyzwanie przede wszystkim dla gestorów,
- Nawiązanie współpracy – określenie wspólnych korzyści

Co zyskujemy z przejścia na źródła administracyjne

PRZYKŁAD WDROŻENIA

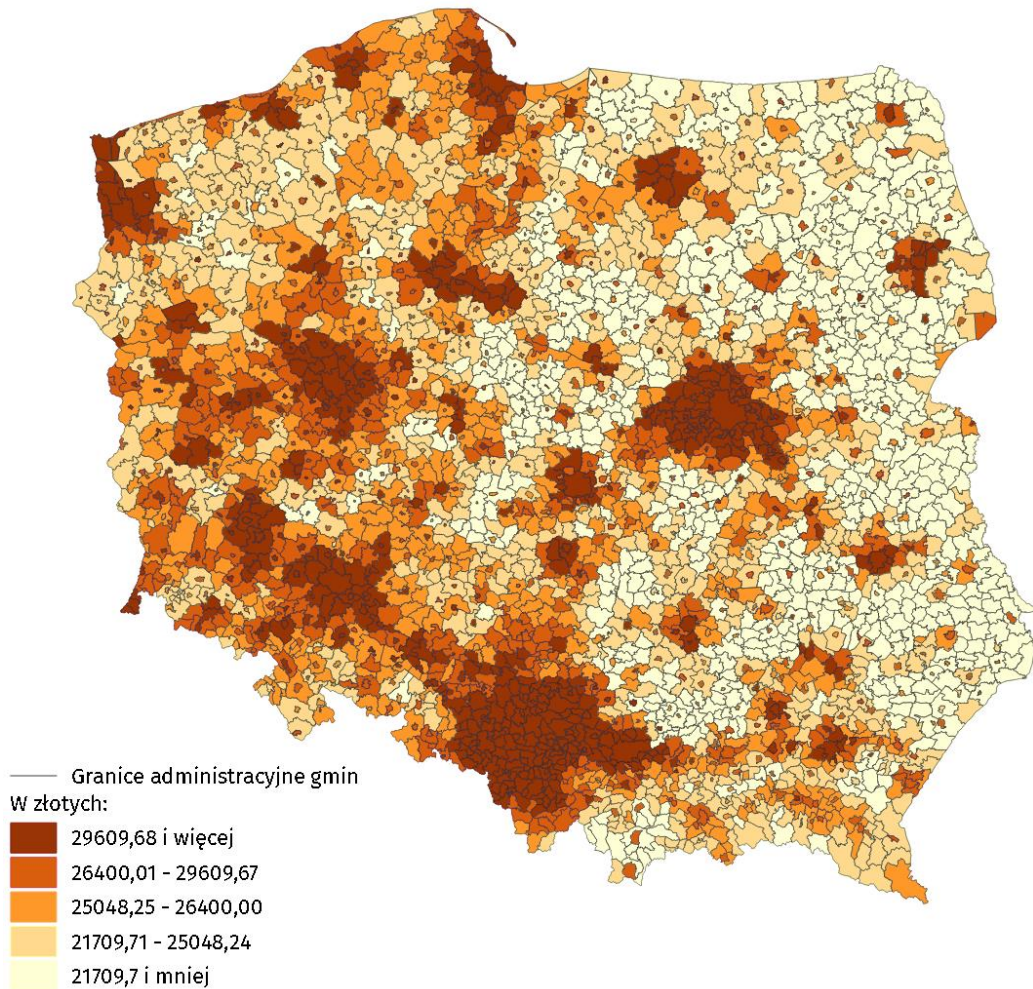
Program badań statystycznych statystyki publicznych 2018 a 2019:

zwiększenie liczby systemów informacyjnych, z których pozyskiwane są dane, z 330 do 347 (wzrost liczby gestorów – ze 148 do 152) na potrzeby 128 badań – 28 zupełnie nowych źródeł danych, niewykorzystywanych wcześniej;

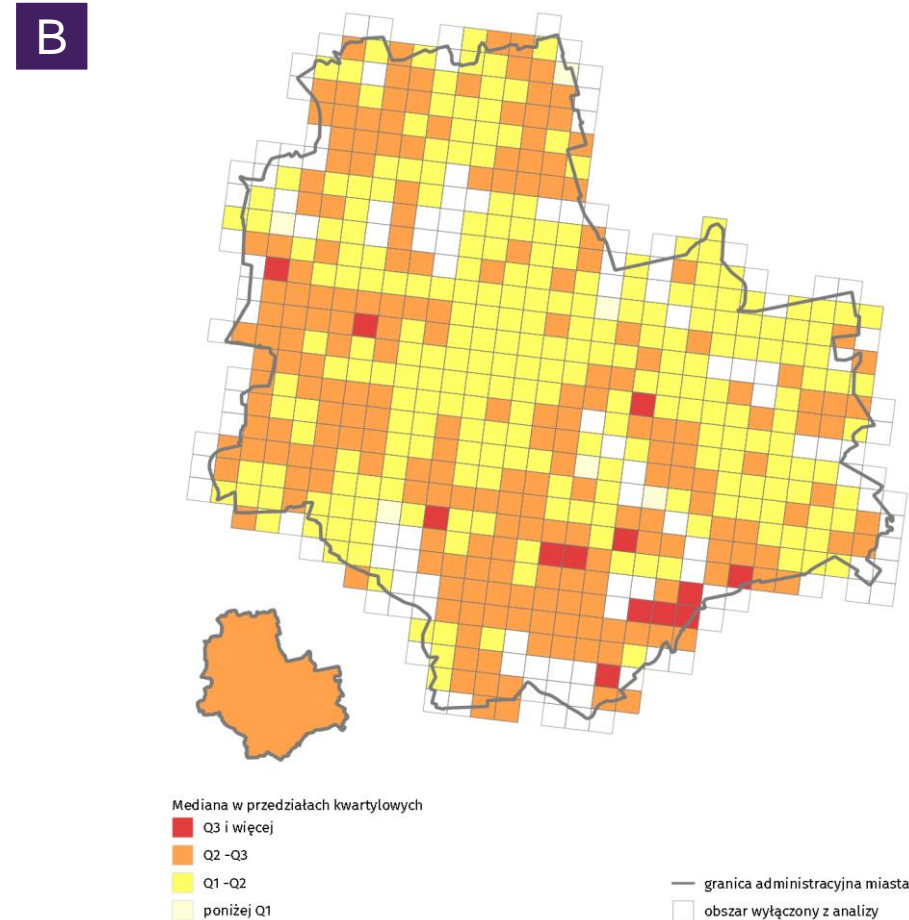
SKUTEK

- Ograniczenie zakresu przedmiotowego 4 sprawozdań;
- Ograniczenie zakresu podmiotowego 6 sprawozdań;
- W jednym przypadku – odejście od przekazywania danych przez wojewodów na rzecz jednego podmiotu;
- Rezygnacja z 2 zestawów danych.

Zastosowanie: Stratyfikacja dochodowa mieszkańców miast



A. Mediana przychodów (w złotych) ludności w 2018 r. według gmin w kraju



B. Przeciętny poziom dochodów mieszkańców Warszawy w 2018 r. (mediana w przedziałach kwartylowych)

Dane administracyjne vs. big data – cechy wspólne

pozyskanie wiedzy na potrzeby
realizacji polityk publicznych

ponadprzeciętny
wolumen danych

możliwość wykorzystania tych
samych narzędzi i techniki

sprostanie współczesnym
wyzwaniom informacyjnym
społeczeństwa

Big data – definicja przez warunki

objętość

(ang. volume)

szybkość

(ang. velocity)

różnorodność

(ang. variety)

wartość

(ang. value)

wiarygodność

(ang. veracity)

wizualizacja

(ang. visualization)

zmienność danych

(ang. variability)

trafność

(ang. validity)

rentowność

(ang. viability)

**trudność
przetwarzania**
(ang. viscosity)



Źródła big data w statystyce publicznej

SZANSE

- Zmniejszenie kosztów badań i obciążenia respondentów
 - Skrócenie czasu dostarczenia danych do odbiorców, analizy nawet w czasie (prawie) rzeczywistym
 - Nowe informacje, nowe kompetencje w statystyce publicznej
-

WYZWANIA

- Współpraca z dostawcami danych z sektora prywatnego
 - Regulacje prawne nienadążające za postępem technologicznym
 - Integracja z istniejącymi systemami statystyki publicznej, porównywalność danych
 - Dostosowanie infrastruktury informatycznej
-

ZAGROŻENIA

- Opłaty za dostęp do danych
- Tajemnica statystyczna, ochrona prywatności
- Jakość i stabilność danych, niewystarczające pokrycie informacyjne
- Odpływ specjalistów do sektora prywatnego

Big data w statystyce publicznej

ŹRÓDŁA DANYCH

- dane internetowe
- zobrazowanie satelitarne
- inteligentne urządzenia
- czujniki drogowe
- AIS (pol. System Automatycznej Identyfikacji)

TEMATY

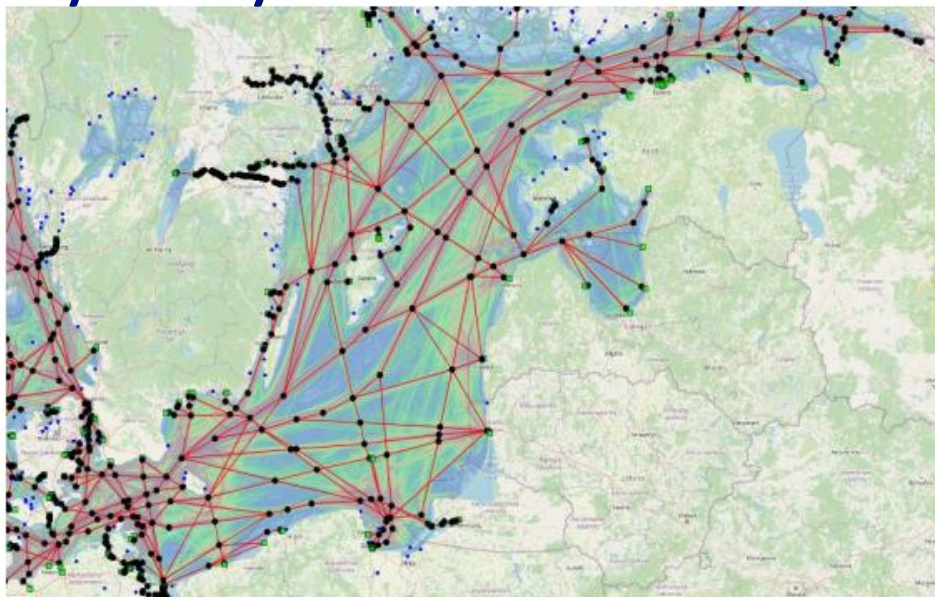
- webscraping
- machine learning
- kodowanie i klasyfikacja
- imputacja braków danych
- optymalizacja szacunków
- jakość danych
- covid-19

WYKORZYSTANIE

- internetowe oferty pracy (rynek pracy od strony popytu)
- charakterystyki przedsiębiorstw
- statystyka rolnictwa
- monitorowanie obszarów zurbanizowanych
- statystyka turystyki
- statystyka morska / portowa
- inteligentne statystyki (wykorzystanie aplikacji internetowych i komórkowych do badań ankietowych)

Przykłady

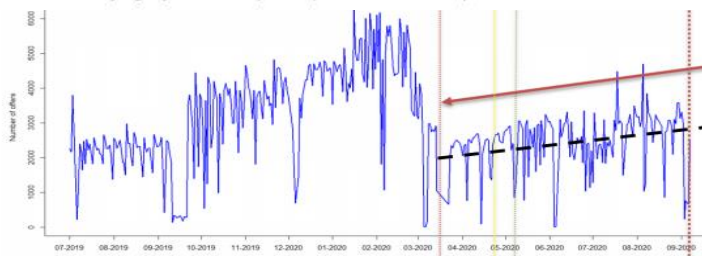
A



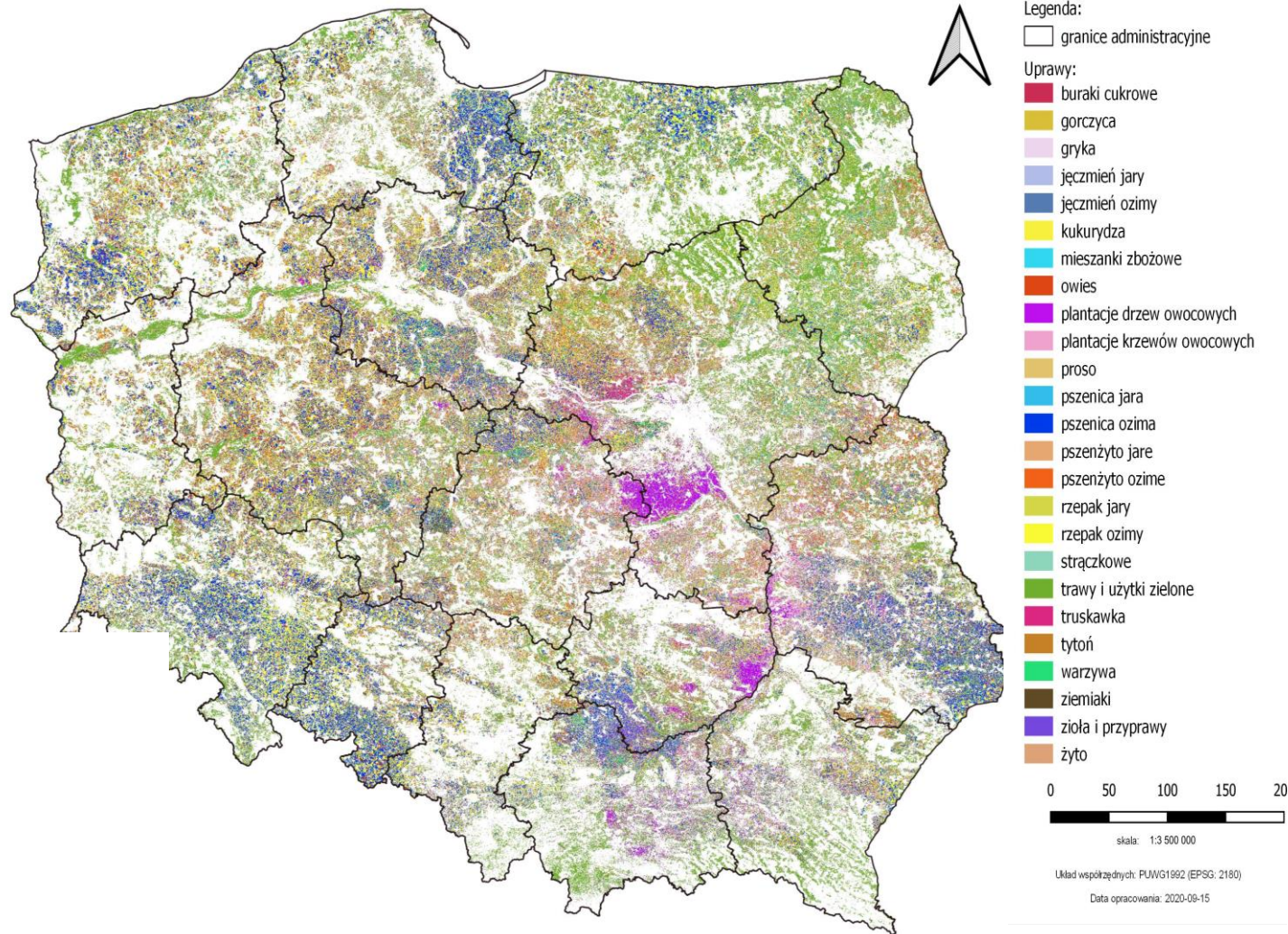
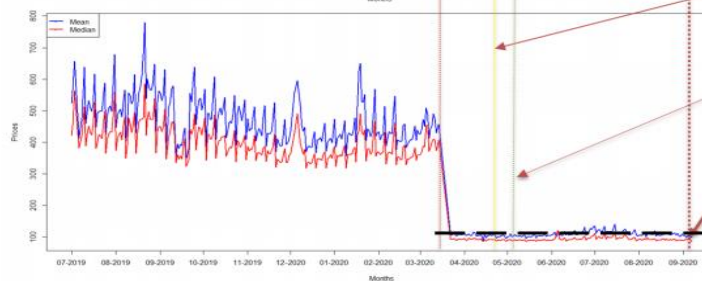
Wizualizacja grafu dla Bałtyku. Opracowanie - zespół AM.

C

Liczba ofert



Ceny



- A. Model szacowania pracy przewozowej
- B. Wynikowy szacunek głównych ziemiopłodów w 2020 r.
- C. Wpływ Covid-19 na podaż i cenę obiektów noclegowych

Działania rozwojowe statystyki publicznej związane z danymi zastanymi (1/2)

CEL – ROLA STATYSTYKI

- Szybsze pozyskiwanie wysokiej jakości i bardziej adekwatnych danych;
- Rozbudowa krajowego systemu informacyjnego w celu dostosowania go do wymagań związanych z programowaniem polityki spójności;
- Kompleksowe monitorowanie i ewaluacja rozwoju dla potrzeb prowadzonych polityk publicznych.



LEPSZE ZARZĄDZANIE ROZWOJEM

Działania rozwojowe statystyki publicznej związane z danymi zastanymi (2/2)

BADANIA

- Prace eksperymentalne jako element procesu badawczego w statystyce, służące pozyskiwaniu nowych informacji wynikowych oraz doskonaleniu obecnych metod gromadzenia i przetwarzania danych;
- Udział w zewnętrznych (międzynarodowych, międzyinstytucjonalnych, wielostronnych) projektach badawczych, których celem jest wdrożenie do produkcji statystycznej pozastatystycznych źródeł danych oraz nowoczesnych metod i technik ich przetwarzania;

WSPÓŁPRACA

- Środowisko naukowe jako wsparcie w rozwoju metodologii badań oraz metod przetwarzania danych;
- Gestorzy danych jako niezbędne ogniwo w pozyskaniu i implementacji źródeł pozastatystycznych (w tym w zakresie wypracowania spójności metodologicznej i definicyjnej, formatu i schematów przekazywania danych oraz metadanych);
- Dywersyfikacja źródeł finansowania i partycypacja partnerska;

EDUKACJA

- Podnoszenie świadomości na temat potrzeby utrzymywania i doskonalenia systemów administracyjnych (zasilanie, rewizja danych, nowoczesne rozwiązania informatyczne, ułatwiające przetwarzanie danych);
- Szkolenia i wymiana wiedzy w zakresie pozastatystycznych źródeł danych oraz nowoczesnych metod i technik ich pozyskiwania, przetwarzania, gromadzenia i udostępniania;
- Udział w opracowywaniu wytycznych, w szczególności w obszarze nowych kierunków działania (tj. budowanie kompetencji big data - poziom organizacji i jednostki) oraz współpraca międzynarodowa w tym zakresie.

<https://stat.gov.pl/statystyki-eksperymentalne/>

Dziękuję za uwagę

Dominika Rogalińska

Dyrektor Departamentu Badań Przestrzennych i Środowiska GUS

d.rogalinska@stat.gov.pl

