

Innowacyjne metody zbierania danych oraz prezentacji wyników badań statystycznych

# GIS W GUS

Odniesienie badanych przez GUS zjawisk demograficznych, społecznych i gospodarczych do przestrzeni wymaga wszechstronnego stosowania technologii GIS. Docelową informatyczną platformą przestrzennej prezentacji danych oraz wyników analiz geostatystycznych będzie Portal Geostatystyczny.

JANUSZ DYGASZEWICZ  
AGNIESZKA NOWAKOWSKA

## ● MODEL REALIZACYJNY SPISÓW

Wszelkie działania – zarówno w zakresie metodyki, jak i organizacji realizacji badań statystycznych – mają na celu zebranie rzetelnych danych stanowiących podstawę opracowania wysokiej jakości wyników informacji w pełni zaspokajających zgłoszone i przewidywane potrzeby użytkowników. Dwa ostatnie lata były dla statystyki publicznej największym wyzwaniem. W 2010 roku, zgodnie z ustawą z 17 lipca 2009 r. o powszechnym spisie rolnym w 2010 r. (DzU nr 126, poz. 1040), w okresie od 1 września do 31 października został przeprowadzony Powszechny Spis Rolny (PSR 2010), a w bieżącym roku od 1 kwietnia do 30 czerwca – Narodowy Spis Powszechny Ludności i Mieszkań (NSP 2011).

Mając na względzie konieczność oszczędnego gospodarowania środkami finansowymi, w realizacji spisów powszechnych wprowadzono nowoczesne i relatywnie tańsze rozwiązania, niż były stosowane dotychczas [patrz GEO-DETA 5/2011 – red.]. Opierały się one na pozyskaniu danych od 16 gestorów z 25 systemów informacyjnych, w tym administracyjnych, oraz wykorzystaniu narzędzi komunikacji elektronicznej. Całkowicie wyeliminowane zostały formularze papierowe. Pozwoliło to na zmniejszenie obciążenia respondentów oraz ograniczenie kosztów druku materiałów spisowych. W spisach powszechnych zastosowane zostały następujące kanały pozyskiwania danych:

- źródła administracyjne,
- samospis internetowy (CAII – *Computer Assisted Internet Interview*),
- wywiad telefoniczny (CATI – *Computer Assisted Telephone Interview*),
- spis za pośrednictwem rachmistrza (CAPI – *Computer Assisted Personal Interview*) wyposażonego w terminal przenośny typu hand held.

Do NSP 2011 pozyskano dane z systemów administracyjnych oraz pozaadministracyjnych na podstawie przepisów ustawy z 4 marca 2010 r. o narodowym spisie powszechnym ludności i mieszkań w 2011 r. (DzU nr 47, poz. 277). Dużym wyzwaniem organizacyjnym i technologicznym dla jednostek statystyki publicznej było zebranie ponad 10 000 zbiorów danych z systemów informacyjnych administracji publicznej prowadzonych przez urzędy gmin i starostwa powiatowe. Przygotowana została platforma informatyczna służąca do pozyskiwania danych z tych systemów za pomocą aplikacji internetowej umożliwiającej transfer zbiorów do serwera zasobowego GUS w bezpiecznym kanale szyfrowanym. Pozyskanie zbiorów danych z systemów rozproszonych wymagało również:

- zorganizowania prac,
- ustalenia koordynatorów przesyłania zbiorów,
- budowy pomocniczej bazy zawierającej dane koordynatorów,
- przygotowania wystąpień do gestorów systemów o przekazanie do GUS zbiorów danych,
- ustalenia aspektów technicznych transferu danych z systemów informacyjnych do serwera zasobowego GUS z wykorzystaniem aplikacji internetowej,

- bezpośredniej współpracy z gestorami systemów.

## ● WYKORZYSTANIE TECHNOLOGII GIS W SPISACH

Innowacją wprowadzoną w obydwu spisach powszechnych było wykorzystanie na każdym etapie prac technologii GIS, w której zostały przygotowane mapy cyfrowe. Były one niezbędnym narzędziem pracy dla rachmistrzów spisowych (w zakresie poruszania się w terenie, weryfikacji operatu itd.), liderów gminnych oraz dyspozytorów wojewódzkich i centralnych, którzy na mapie mogli weryfikować postęp spisu oraz np. marszrutę lub położenie rachmistrza. Mapa cyfrowa zawierała materiały pozyskane z PZGiK (ortofotomapę, granice województw, powiatów i gmin, nazwy miejscowości), granice rejonów statystycznych i obwodów spisowych oraz statystyczne punkty adresowe przygotowane przez służby statystyki publicznej, a także warstwę działek ewidencyjnych (ARiMR) oraz drogi i ulice.

Mapy cyfrowe w technologii GIS miały zastosowanie w obydwu spisach podczas następujących etapów prac:

**1. Aktualizacji gminnej**, dzięki której urzędy gmin miały możliwość edycji punktów adresowych on-line.

### **2. Obchodu przedspisowego:**

- w aplikacji rachmistrza spisowego do wprowadzania nowych punktów, edycji istniejących oraz usuwania nieistniejących w terenie,

- w aplikacji lidera gminnego AGMIS oraz aplikacji dyspozytorskiej ADYS, dzięki którym można było na bieżąco monitorować przebieg spisu i reagować.

**TABELA 1. STATYSTYKA BADANIA REPREZENTACYJNEGO W PODZIALE NA POSZCZEGÓLNE KANAŁY POZYSKIWANIA DANYCH**

Mieszkania w skali Polski	Procent	w tym:	Procent
z wypełnioną ankietą	86,7	Procent jednostek w CAPI	92,7
		Procent jednostek w CAII	6,0
		Procent jednostek w CATI	1,1
		Procent jednostek w CAPI/CAII/CATI	0,2
niezamieszkane	13	-	-
anulowane	0,3	-	-

### 3. Samego spisu:

- w aplikacji rachmistrza spisowego do przeprowadzania wywiadu,

- w aplikacji lidera gminnego AGMIS oraz aplikacji dyspozytorskiej ADYS, pozwalając na monitorowanie i raportowanie przebiegu obchodu oraz spisu, a także planowanie i zarządzanie pracą rachmistrzów na podległych obszarach.

Najważniejszym aspektem całego przedsięwzięcia dotyczącym statystycznych punktów adresowych jest ich jakość. Punkty te zostały przygotowane przez pracowników statystyki publicznej zajmujących się od wielu lat prowadzeniem i aktualizacją Krajowego Rejestru Urzędowego Podziału Terytorialnego Kraju TERYT. Następnie podlegały weryfikacji w urzędach gmin. Ostatnim etapem prac było sprawdzenie istnienia i potwierdzenie ich lokalizacji w terenie przez rachmistrza spisowego. Dzięki etapowemu procesowi przygotowania statystycznych punktów adresowych należy uznać je za wiarygodne.

Mając taką warstwę danych, GUS niezależni się od zmian dotychczasowego systemu identyfikacji przestrzennej poprzez przejście z przyporządkowania obszarowego (obwody spisowe) do przyporządkowania punktowego x, y. Zmiana ta umożliwi bardziej elastyczne grupowanie danych zbieranych w statystyce publicznej dla dowolnie małych obszarów – oczywiście z zachowaniem tajemnicy statystycznej. Pozwoli także na utworzenie bazy mikrodanych o charakterze przestrzennym umożliwiające dokonywanie analiz geostatystycznych różnych zjawisk społecznych.

### ● STATYSTYKI PRZEPROWADZENIA NSP 2011

W dalszej części artykułu zaprezentowano statystyki dotyczące realizacji spisu w poszczególnych kanałach. Ankieta do badania pełnego składała się z 16 pytań, a do badania reprezentacyjnego, do której wylosowano według określonego algorytmu 20 procent mieszkań w skali

kraju, z 60 do 100 pytań. Z możliwości weryfikacji lub uzupełnienia swoich danych w samospisie internetowym (CAII) skorzystało 12% populacji. W badaniu pełnym dla 91% populacji pozyskano kompletne dane ze źródeł administracyjnych. W badaniu pełnym dla:

- 98,9% populacji – pozyskano dane kompletne,

- 0,9% populacji – dane są niekompletne, najczęściej dotyczy to osób określonych w gminnych zbiorach meldunkowych jako osoby „wymeldowane donikąd”, dla których brak możliwości ustalenia adresu przebywania,

- 0,2% populacji – brak wystarczających danych do podjęcia działań spisowych.

W badaniu reprezentacyjnym zostało spisanych 100% wylosowanych mieszkań. W tabeli 1 przedstawiono statystykę badania reprezentacyjnego w podziale na poszczególne kanały pozyskiwania danych. Natomiast statystyki zakończenia spisu reprezentacyjnego (mieszkania wylosowane) ujęto w tabeli 2. Na kolejnych stronach przedstawiono na mapkach statystyki spisu w poszczególnych kanałach – CAII, CAPI i CATI.

### ● PREZENTACJA WYNIKÓW SPISÓW

Informacje zebrane podczas spisów zgromadzone zostały w specjalnie chronionej Operacyjnej Bazie Mikrodanych (OBM). Dane te, zgodnie z zapisami obydwu wspomnianych ustaw spisowych, zostaną trwale usunięte nie później niż po upływie 2 lat od dnia zakończenia każdego ze spisów. Do tego czasu odpersonalizowane dane trafią do Analitycznej Bazy Mikrodanych (ABM). Upowszechnienie danych spisowych w postaci mikroagregatów, agregatów i tabel wynikowych będzie realizowane w spo-

sób nowatorski – poprzez dostęp do ABM. Uzyskane informacje pozwolą na prowadzenie różnotematycznych analiz (w tym analiz przestrzennych) i tworzenie indywidualnych raportów. Ocenę zawartości zgromadzonych danych umożliwią publicznie dostępne metadane. Dane będą prezentowane w formie odpersonalizowanej, z zachowaniem tajemnicy statystycznej.

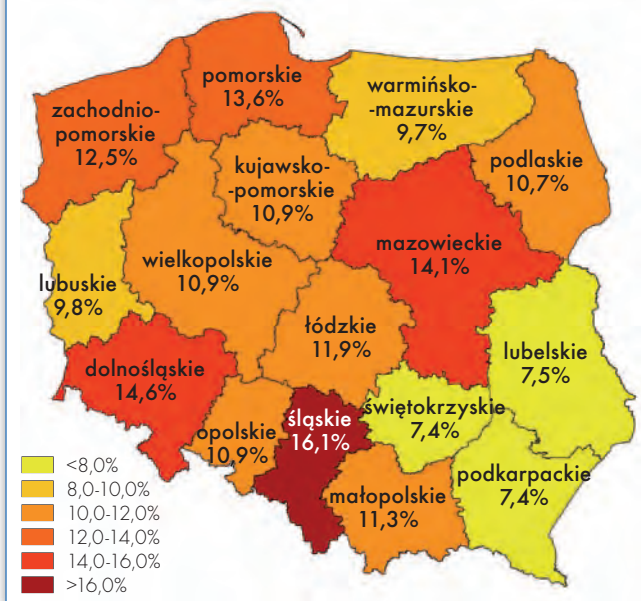
W celu upowszechniania wyników spisów oraz prezentacji ich w ujęciu przestrzennym GUS postanowił stworzyć Portal Geostatystyczny. We wrześniu br. zostało uruchomione postępowanie o udzielenie zamówienia publicznego w trybie przetargu nieograniczonego na wykonanie usługi rozbudowy platformy aplikacyjnej GIS Głównego Urzędu Statystycznego przez opracowanie i wdrożenie Portalu Geostatystycznego (PGS) na potrzeby prezentacji wyników Powszechnego Spisu Rolnego 2010 oraz Narodowego Spisu Powszechnego 2011. Portal stanowić będzie docelową informatyczną platformę przestrzennej prezentacji danych oraz wyników analiz geostatystycznych GUS. W szczególności dotyczyć to będzie danych PSR 2010 oraz NSP 2011 gromadzonych w Analitycznej Bazie Mikrodanych (ABM), jak również innych mikrodanych zagregowanych w bazach analitycznych Systemu Informacyjnego Statystyki Publicznej (SISP).

Zadaniem Portalu Geostatystycznego będzie także wypełnienie zadań wiążących się z implementacją wytycznych dyrektywy INSPIRE dla tematów 1 i 10 III grupy tematycznej, włącznie z obsługą „geoportalu branżowego” dla zdefiniowanego profilu metadanych branżowych GUS. Ponadto portal będzie umożliwiał urzędom gmin bieżącą aktualizację statystycznych punktów adresowych. Serwis pozwoli na dodawanie nowo powstałych punktów zgodnie z prowadzoną w gminach numeracją po-

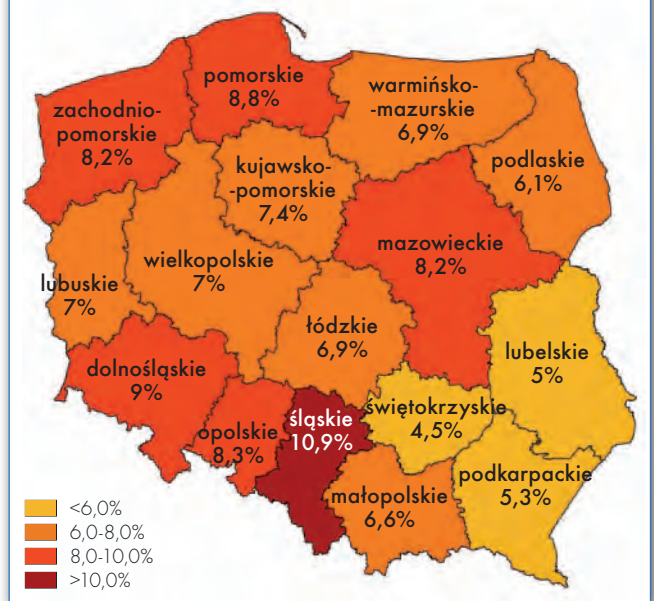
**TABELA 2. STATYSTYKA ZAKOŃCZENIA SPISU REPREZENTACYJNEGO (MIESZKANIA WYLOSOWANE)**

Stan na dzień 30.06.2011 r.	Liczba mieszkań	Udział procentowy
Mieszkanie spisane	2 377 951	86,7
Mieszkanie niezamieszkane, brak respondenta, brak mieszkania	357 252	13,0
Odmowa spisu	3 162	0,1
Inne przypadki	5 671	0,2
Suma	2 744 036	100,00

## UDZIAŁ PROCENTOWY OSÓB, KTÓRE DOKONAŁY SAMOSPISU (12% W SKALI KRAJU)



## UDZIAŁ PROCENTOWY MIESZKAŃ WYŁOSOWANYCH DO BADANIA REPREZENTACYJNEGO, KTÓRE ZOSTAŁY SPISANE W SAMOSPISIE (7,6% W SKALI KRAJU)



rządkową nieruchomości. Dane wprowadzane na serwerze mapowym będą zasilały krajowy rejestr urzędowy podziału terytorialnego kraju – TERYT.

Oprócz korzystania z gotowych analiz przestrzennych, indywidualny użytkownik będzie mógł w ramach Portalu Geostatystycznego redagować własne mapy tematyczne w formie kartogramu na podstawie dowolnej cechy modelu danych tematycznych oraz drukować opracowane przez siebie mapy.

Planuje się, iż portal będzie pozwalał generować obszary analizy na podstawie: buforowania wstawionego punktu, na-

rysowania grafiki poligonowej oraz wygenerowania raportu sumarycznego (raport tabelaryczny może być uzupełniony mapą obszaru i wykresami wybranych cech).

Założono, iż system będą tworzyły następujące części:

1. Podsystem Utrzymania Bazy Danych Geostatystycznych.
2. Podsystem Administrowania Systemem.
3. Podsystem Zarządzania Treścią (CMS).
4. Podsystem Publikowania i Aktualizacji Informacji Geostatystycznych:

- Moduł Przeglądarki Danych Geostatystycznych,

- Moduł Zaawansowanej Przeglądarki Danych Geostatystycznych.

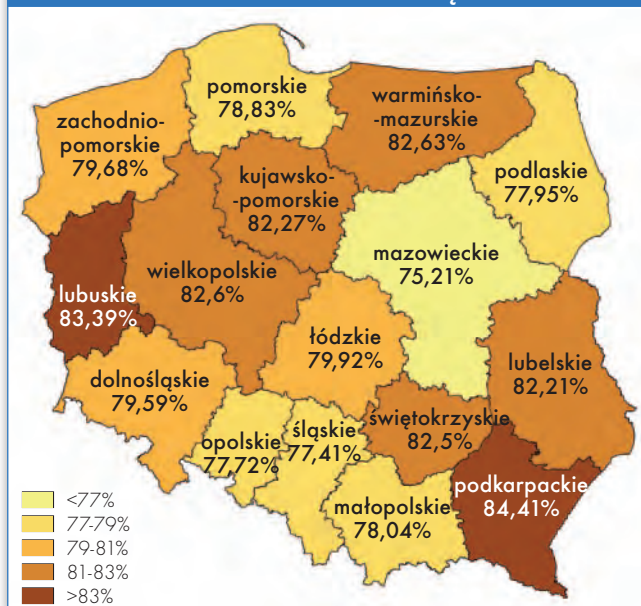
5. Komponenty – serwisy mapowe:

- Moduł Aktualizacji Gminnej (rozbudowa Aplikacji Aktualizacji Gminnej),

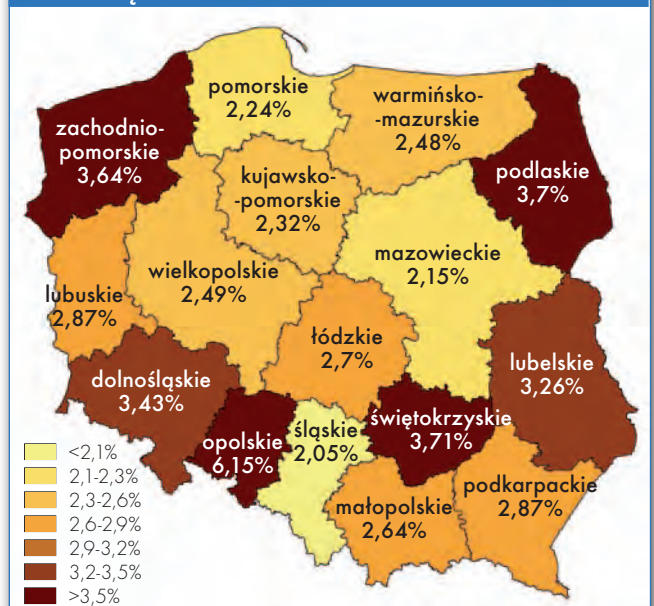
- Podsystem Usług INSPIRE zawierający Portal Metadanych oraz komponenty – usługi sieciowe OGC/ISO: WMS, WFS, WCS, CSW.

W celu optymalnej realizacji założonych w zamówieniu zadań zostało ono podzielone na następujące etapy:

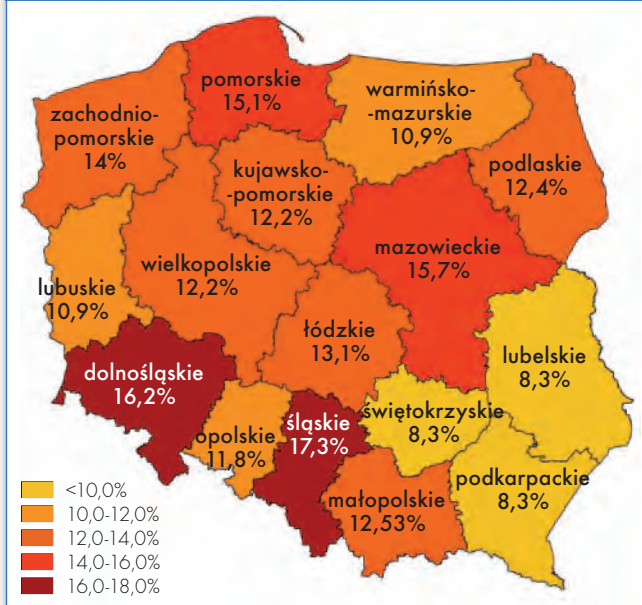
## UDZIAŁ PROCENTOWY MIESZKAŃ WYŁOSOWANYCH DO BADANIA REPREZENTACYJNEGO, KTÓRE ZOSTAŁY SPISANE METODĄ CAPI



## UDZIAŁ PROCENTOWY OSÓB PODLEGAJĄCYCH BADANIU PEŁNEMU, KTÓRE ZOSTAŁY SPISANE METODĄ CAPI



## UDZIAŁ PROCENTOWY OSÓB PODLEGAJĄCYCH BADANIU PEŁNEMU, KTÓRE DOKONAŁY SAMOSPISU (13,3% W SKALI KRAJU)



●I – opracowanie projektu wdrożenia systemu (PWS) w terminie do 12 grudnia 2011 r.;

●II – opracowanie, wdrożenie oraz konfiguracja systemu potwierdzonego przez PWS – do 31 stycznia 2012 r.;

●III – szkolenie użytkowników zamawiającego w zakresie obsługi i funkcji systemu – do 29 lutego 2012 r.;

●IV – wdrożenie serwisów mapowych dla 100% danych PSR2010 w zakresie informacyjnym określonym w Rozdz. 7 Dodatek nr 2 Planowane serwisy tematyczne PSR 2010 potwierdzonym przez PWS – do 31 marca 2012 r.;

w innych badaniach. W bieżącym roku rozpoczęły się prace nad przygotowaniem operatów przestrzennych do badań społecznych, a w kolejnych latach planuje się rozbudowę rejestru statystycznego, jakim jest Baza Jednostek Statystycznych. Jest to baza danych, której podstawą jest rejestr REGON, jednak w odróżnieniu od niego jest aktualizowana na podstawie wyników badań statystycznych oraz informacji pochodzących ze źródeł administracyjnych (np. Zakładu Ubezpieczeń Społecznych, Krajowego Rejestru Sądowego). Stanowi ona podstawowe (wspólne dla autorów wszystkich badań) źródło da-

nych wykorzystywanych nie tylko przy określaniu operatu i próby, ale też zbior informacji niezbędnych na etapie redagowania wyników badań. Rozbudowa tego rejestru będzie polegała na dodaniu do adresów przedsiębiorstw również ich współrzędnych x, y (uzyskanie warstwy przestrzennej), aby można było przeprowadzać analizy przestrzenne, np. dojazdy do pracy czy szkoły.

Wykorzystanie nowoczesnych technologii, nowe podejście do spisów powszechnych, rezygnacja z papieru, szerokie zastosowanie danych z rejestrów administracyjnych wymagało ogromnego wysiłku na wszystkich etapach, począwszy od prac przygotowawczych, a skończywszy na opracowaniu wyników. Potrzebne było także wsparcie zarówno organów administracji rządowej, samorządowej, jak i ogółu społeczeństwa. Ostatecznie jednak wspólny wysiłek się opłacił i dla wszystkich stało się jasne, że od nowych rozwiązań nie ma odwrotu. Coraz bardziej oczywiste jest także to, że odniesienie badanych przez Główny Urząd Statystyczny zjawisk demograficznych, społecznych i gospodarczych do przestrzeni wymaga wszechstronnego stosowania technologii GIS.

●V – Wdrożenie serwisów mapowych dla 100% danych NSP2011 w zakresie informacyjnym określonym w Rozdz. 8 Dodatek nr 3 Zakres informacyjny NSP 2011 potwierdzonym przez PWS – do 31 maja 2012 r.;

●VI – Świadczenie usługi asysty technicznej systemu – do 30 listopada 2012 r.

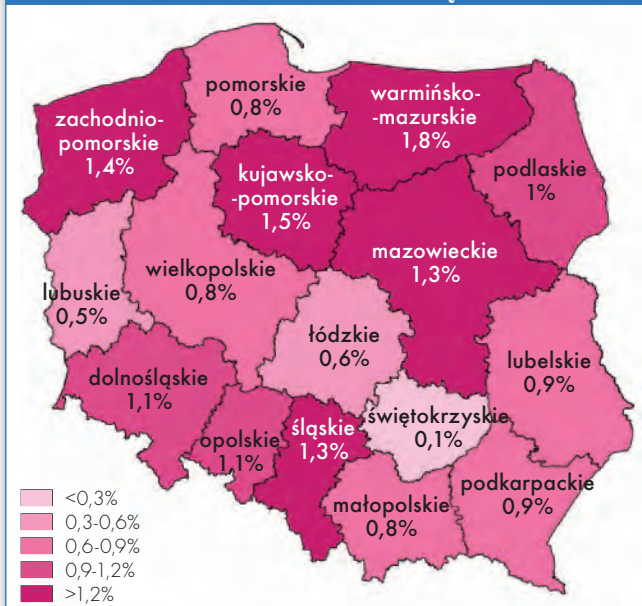
W przyszłości statystyka publiczna zamierza wykorzystywać narzędzia GIS również

W przyszłości statystyka publiczna zamierza wykorzystywać narzędzia GIS również

W przyszłości statystyka publiczna zamierza wykorzystywać narzędzia GIS również

JANUSZ DYGAŚZEWICZ  
(dyrektor Departamentu Programowania i Koordynacji Badań GUS)  
AGNIESZKA NOWAKOWSKA  
(naczelnik Wydziału Informacji Geoprzestrzennej GUS)

## UDZIAŁ PROCENTOWY MIESZKAŃ WYŁOSOWANYCH DO BADANIA REPREZENTACYJNEGO, KTÓRE ZOSTAŁY SPISANE METODĄ CATI



## UDZIAŁ PROCENTOWY OSÓB PODLEGAJĄCYCH BADANIU PEŁNEMU, KTÓRE ZOSTAŁY SPISANE METODĄ CATI

