

Wykorzystanie danych statystyki publicznej w badaniach opinii i rynku – możliwości i bariery metody analizy danych zastanych

Edyta Makowska-Belta

Agenda



Metodologia analizy danych zastanych



Zastosowanie i funkcje w projektach badawczych



Potencjalne bariery



Konkluzje

Metodologia (1)

Czym jest analiza danych zastanych?

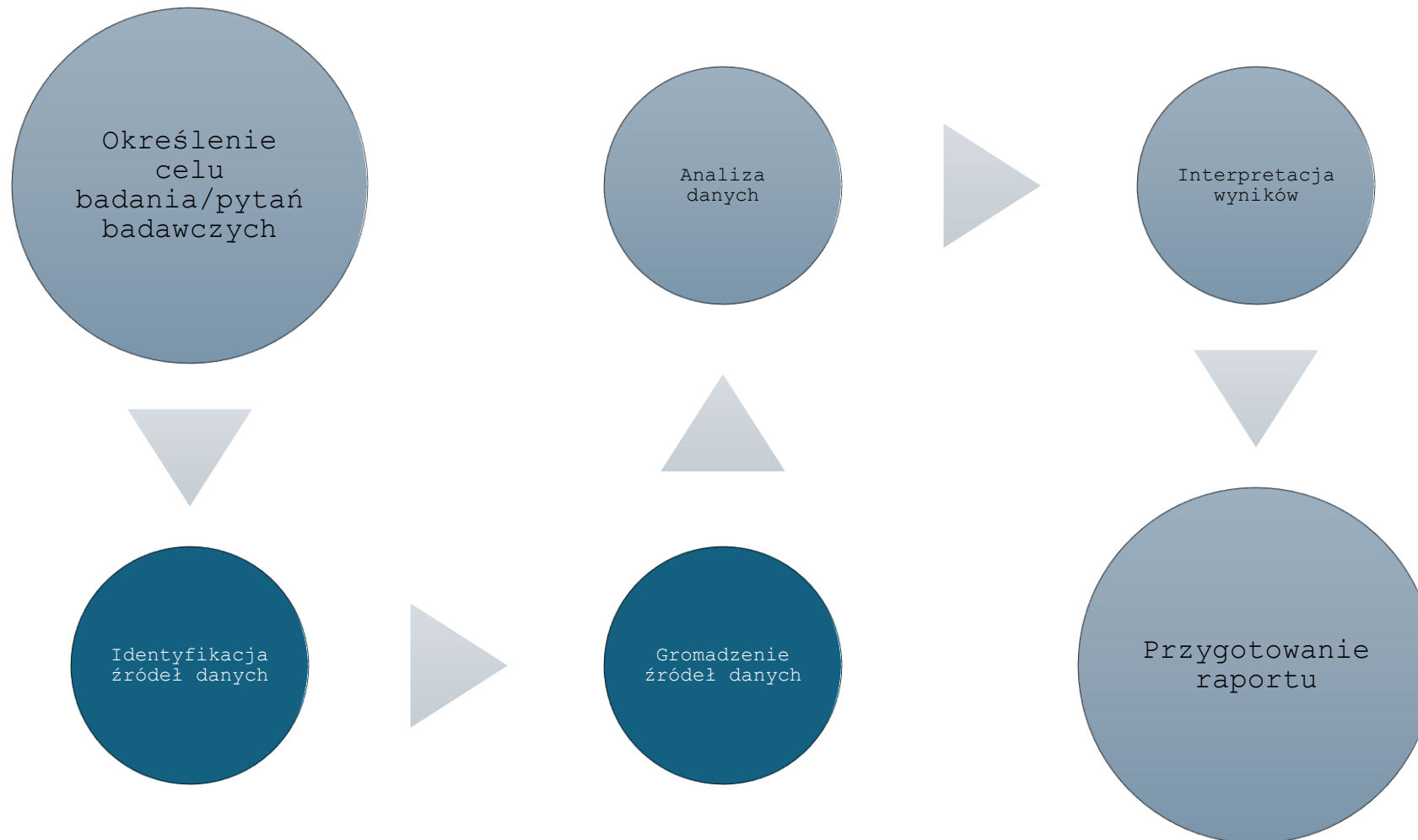
Analiza danych zastanych, znana również jako desk research, polega na zbieraniu, analizowaniu i interpretacji danych, które już istnieją, tj. zostały już zgromadzone przez innych badaczy, organizacje lub instytucje.

W ramach desk research może być przeprowadzona:

- **analiza danych ilościowych** m.in. baz danych, dostępnych statystyk, danych sprawozdawczych, danych z ankiet gromadzonych przez Zamawiającego badanie
- **analiza treści** dokumentów, publikacji, artykułów, wniosków z innych raportów.

Metodologia (2)

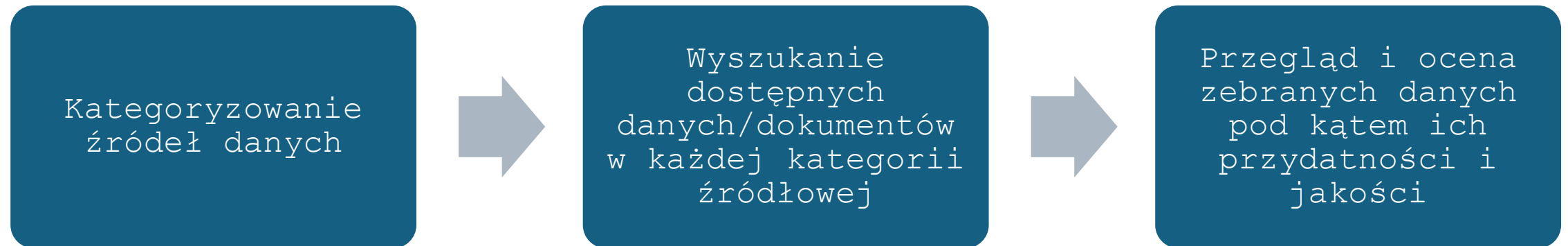
Przebieg procesu analizy danych zastanych w projektach badawczych



Metodologia (3)

Proces gromadzenia źródeł danych

Kluczowym elementem analizy danych zastanych jest zidentyfikowanie potencjalnych źródeł danych odpowiednich dla określonego celu badania oraz funkcji, jaką mają spełnić w projekcie.



Metodologia (4)

Jakie źródła danych można wykorzystać?

Dokumenty
prawne,
strategiczne

Dane
statystyczne

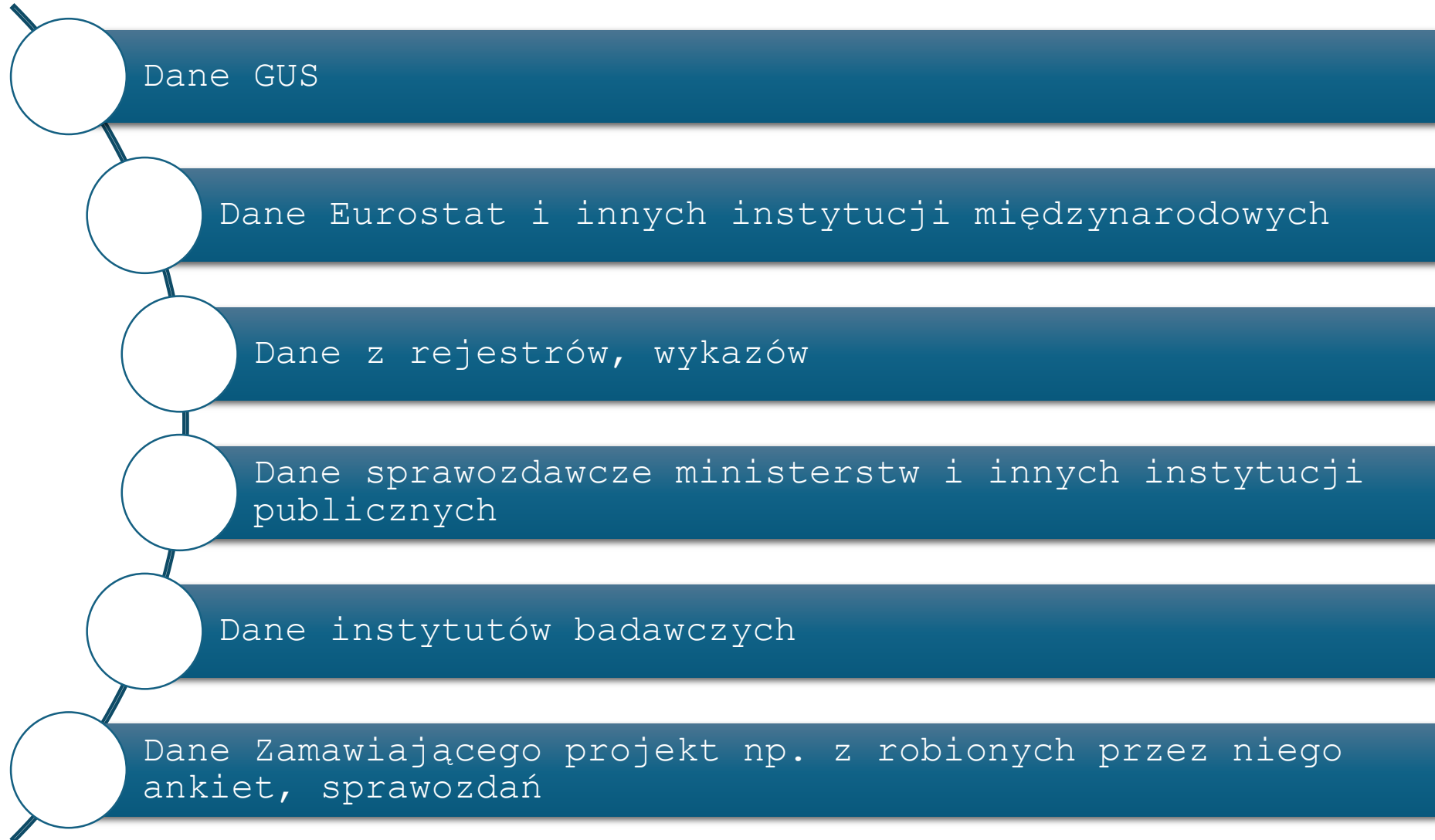
Wyniki
innych badań

Publikacje,
opracowania
naukowe

Artykuły
prasowe,
internetowe

Metodologia (5)

Źródła danych statystycznych



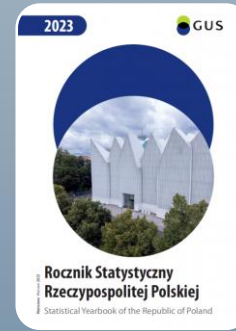
Metodologia (6)

Podstawowe źródła danych GUS



Bazy danych

- Bank Danych Lokalnych
- Działowe Bazy Wiedzy

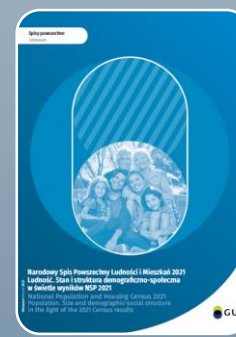


Roczniki statystyczne

- Rocznik Statystyczny RP
- Roczniki wojewódzkie



Publikacje + tablice publikacyjne



Spisy powszechne

Funkcje analizy danych zastanych w projektach badawczych

„Często o wartości przedmiotu decyduje w naszych oczach nawet nie on sam, ale otoczenie, w jakim go ujrzymy, jego kontekst, oprawa.” Ryszard Kapuściński

- **eksploracja** – zebranie wiedzy w celu konceptualizacji badania, doboru odpowiedniej metodologii, czy konstrukcji narzędzi badawczych
- **kontekst (benchmark)** – dostarczenie wiedzy kontekstualnej służącej jako tło dla prezentacji wyników realizowanego badania np. **jak wygląda sytuacja rynkowa w danym obszarze**
- **wskaźniki** – dostarczenie odpowiedzi na pytania/zagadnienia badawcze
- **uzupełnienie** – pogłębienie, wyjaśnienie wniosków płynących z badania

Zastosowanie analizy danych zastanych (1)

Przykłady zastosowania

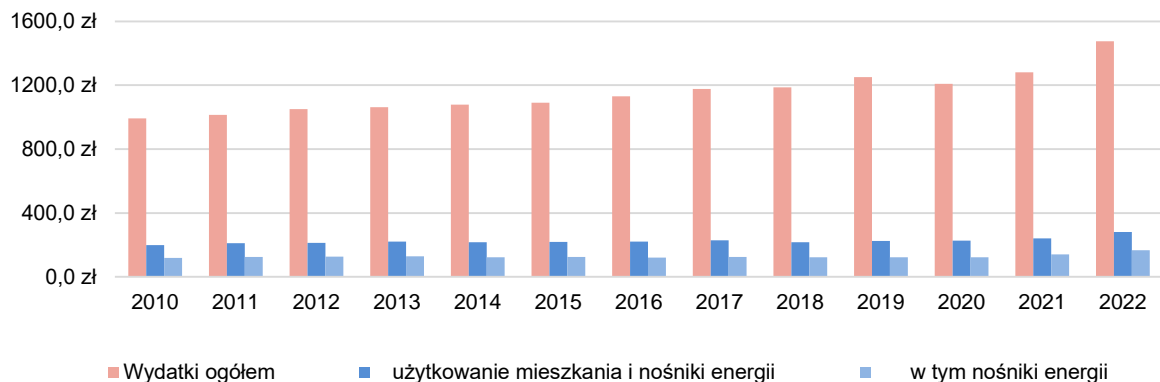
Przykłady zastosowania analizy danych zastanych

- **Badania rynkowe:** Analiza danych zastanych jest często wykorzystywana do oceny rynku, identyfikacji trendów branżowych i analiz konkurencji.
- **Badania społeczne:** Dane zastane mogą być używane do analizy demograficznej, trendów społecznych i zachowań ludzkich.
- **Planowanie strategiczne:** Desk research jest wykorzystywany do analizy danych finansowych, ekonomicznych i demograficznych w celu podejmowania decyzji strategicznych np. szacowanie potencjału inwestycji w populacji (ekstrapolacja wyników przeprowadzonych badań na próbie reprezentatywnej na populację).

Zastosowanie analizy danych zastanych (2)

ENERGETYKA – WYDATKI GOSPODARSTW DOMOWYCH

Przeciętne miesięczne wydatki na 1 osobę w gospodarstwach domowych

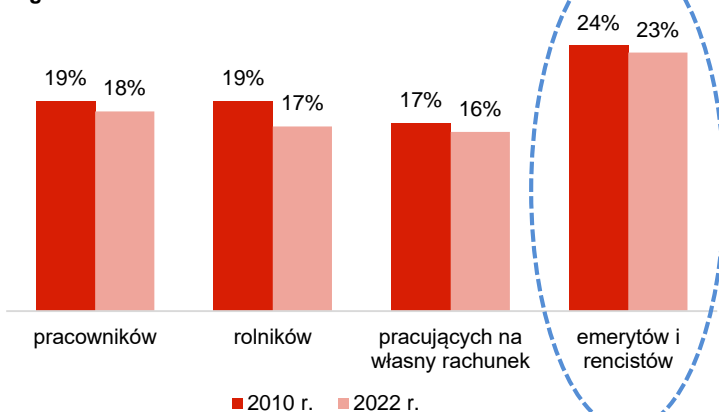


- W 2022 r. wydatki na użytkowanie mieszkania i nośniki energii wyniosły miesięcznie ok. 280 zł na 1 osobę w gospodarstwie domowym, co stanowiło 19% ogółu przeciętnych miesięcznych wydatków. Użytkowanie mieszkania i nośniki energii to druga pod względem wielkości kategoria w strukturze wydatków gospodarstw domowych zaraz po wydatkach na żywność i napoje bezalkoholowe.
- Na same nośniki energii w 2022 r. wydatkowano ok. 168 zł miesięcznie na osobę. Między 2021 r. a 2022 r. przeciętne miesięczne wydatki na nośniki energii na osobę w gospodarstwach domowych wzrosły o 19,4%.
- Wydatki na użytkowanie mieszkania i nośniki energii stanowiły w 2022 r. największy udział w ogóle wydatków wśród gospodarstw 1-osobowych (25%) lub emerytów i rencistów (23%).

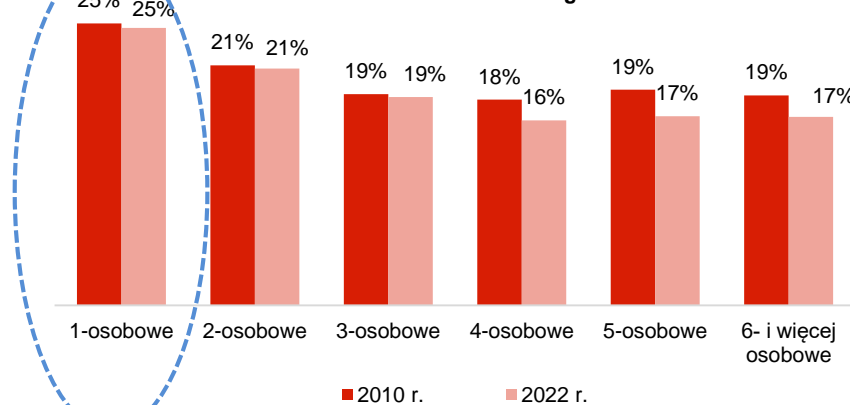
Źródło – Opracowanie własne na podstawie danych GUS, Budżety gospodarstw domowych w 2022 roku – tablice przeglądowe

Udział wydatków na użytkowanie mieszkania i nośniki energii w ogóle wydatków (przeciętnie miesięcznie na 1 osobę w gospodarstwie domowym)

wg GRUP SPOŁECZNO-EKONOMICZNYCH



wg LICZBY OSÓB



Źródło – Opracowanie własne na podstawie danych GUS, Budżety gospodarstw domowych w 2022 roku oraz w 2010 r.

Zastosowanie analizy danych zastanych (3)

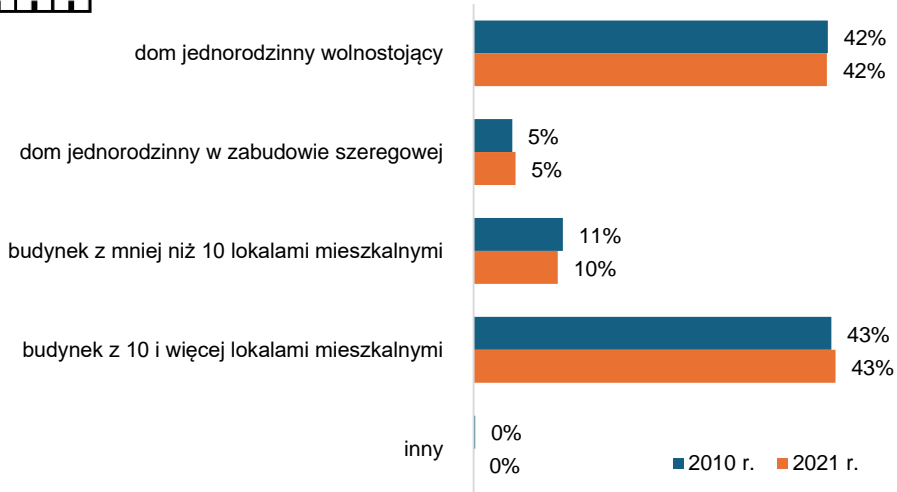
ENERGETYKA – SYTUACJA GOSPODARSTW DOMOWYCH W POLSCE

Gospodarstwa domowe najczęściej mieszkają w budynkach wielorodzinnych z 10 i więcej lokalami mieszkalnymi (43%) lub w domach jednorodzinnych wolnostojących (42%).

Między 2010 r. a 2022 r. zwiększył się dostęp gospodarstw domowych do wyposażenia w ciepłą wodę bieżącą z sieci, gaz z sieci oraz centralne ogrzewanie - lokalne. Spadł natomiast udział gospodarstw domowych, których mieszkanie było wyposażone w gaz (z butli), centralne ogrzewanie z sieci oraz piece na opał.

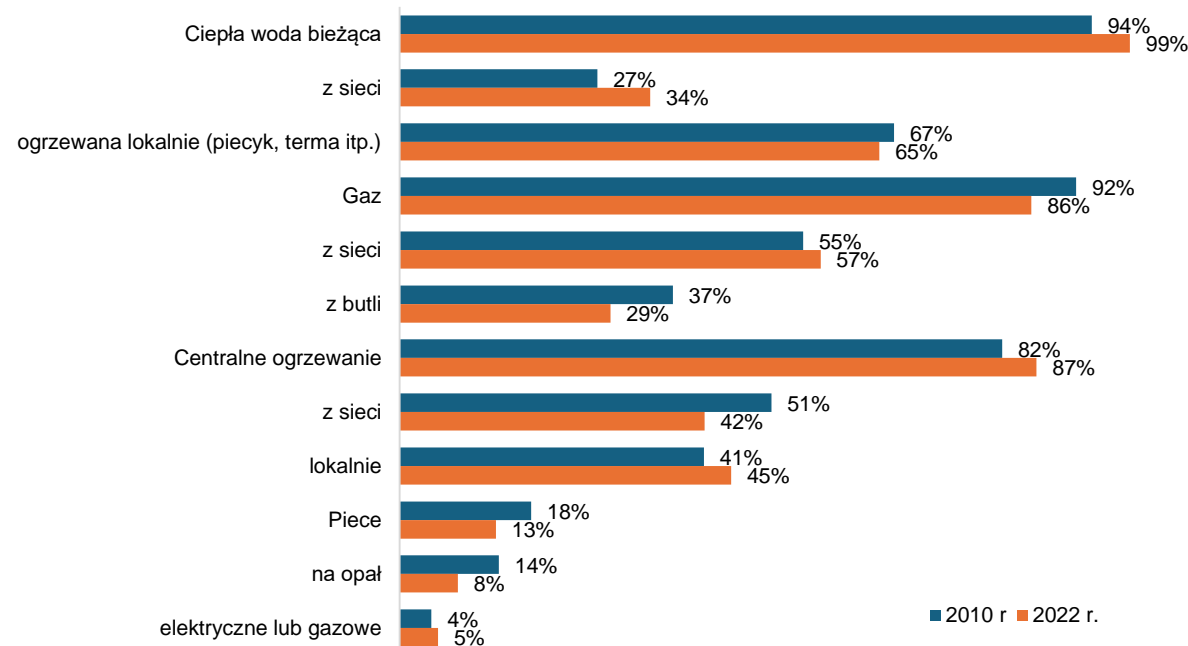


Rodzaj budynku, w którym znajduje się mieszkanie



Źródło – Opracowanie własne na podstawie danych GUS, Dochody i warunki życia ludności Polski - raport z badania EU-SILC 2021 oraz EU-SILC 2010

Wyposażenie mieszkań w instalacje techniczno-sanitarne

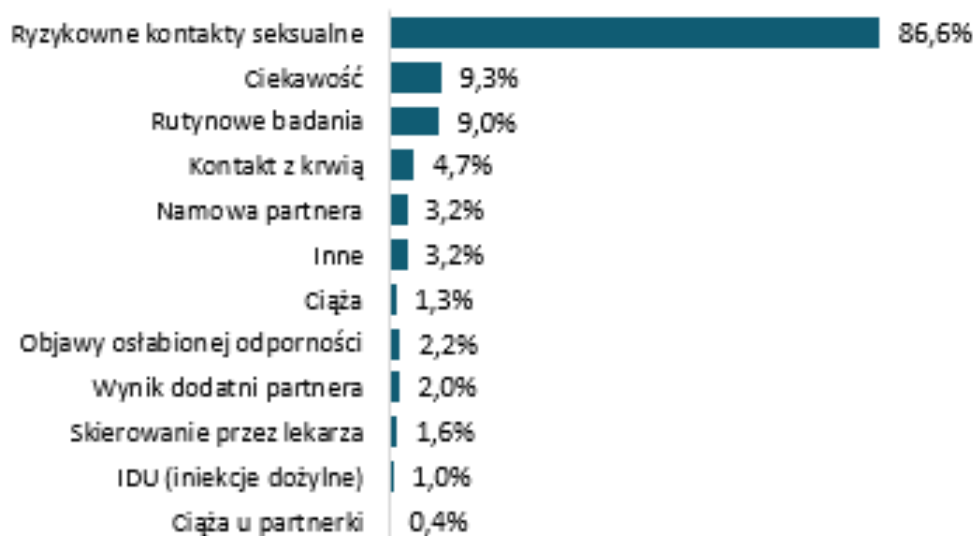


Źródło – Opracowanie własne na podstawie danych GUS, Budżety gospodarstw domowych w 2022 roku oraz w 2010 r.

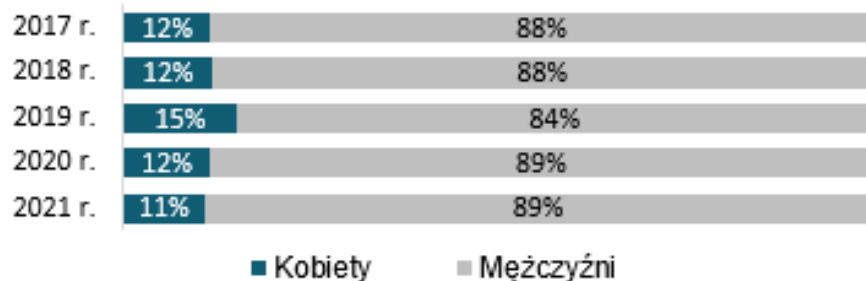
Zastosowanie analizy danych zastanych (4)

PROFILAKTYKA HIV/AIDS

POWODY WYKONANIA TESTÓW NA HIV W PUNKTACH KONSULTACYJNO-DORADCZYCH W 2021 R.

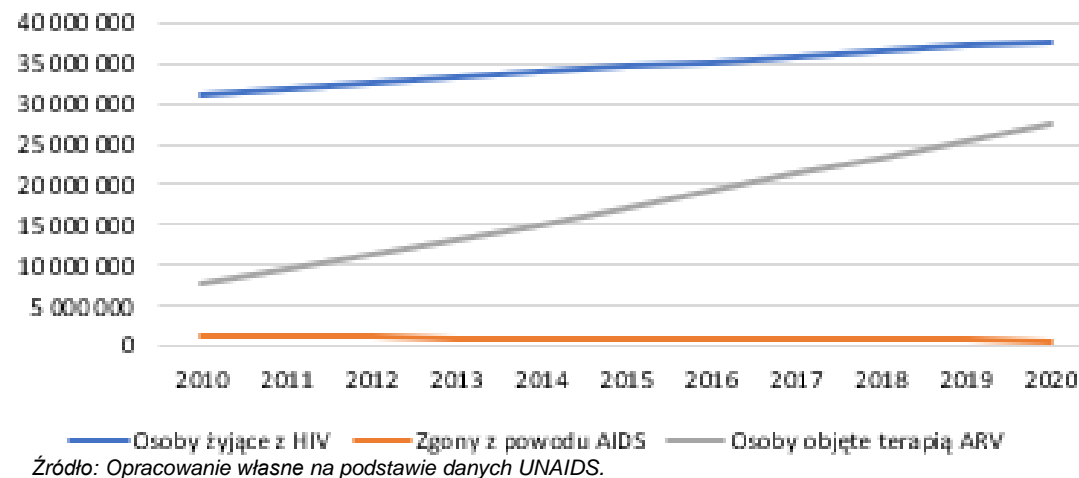


OSOBY Z ZAKAŻENIEM HIV WG PŁCI



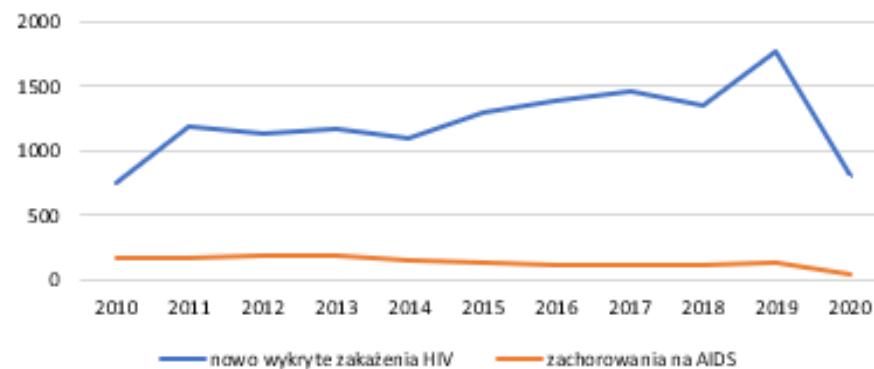
Źródło: Opracowanie własne na podstawie danych Zamawiającego.

SYTUACJA EPIDEMIOLOGICZNA HIV/AIDS NA ŚWIECIE W LATACH 2010–2020



Źródło: Opracowanie własne na podstawie danych UNAIDS.

ZMIANA LICZBY ZAKAŻEŃ HIV I ZACHOROWAŃ NA AIDS W POLSCE W LATACH 2010–2020



Źródło: Opracowanie własne na podstawie danych Narodowego Instytutu Zdrowia Publicznego PZH – Państwowy Instytut Badawczy, biuletyny roczne „Choroby zakaźne i zatrucia w Polsce”. Dla 2020 r. dane wstępne.

Możliwości analizy danych zastanych



Efektywność kosztowa:

Gromadzenie danych zastanych jest zazwyczaj tańsze niż prowadzenie badań pierwotnych.



Szybkość: Dostępne dane można zwykle zgromadzić i przeanalizować w krótszym czasie niż zbieranie danych pierwotnych.



Dostęp do dużej ilości informacji:

Desk research pozwala na korzystanie z szerokiego zakresu danych pochodzących z różnych źródeł, co może dostarczyć bogatych informacji.



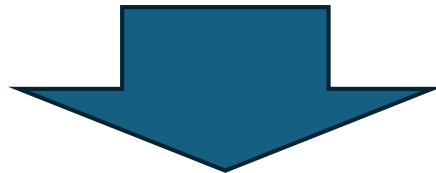
Zasoby historyczne: Analiza danych zastanych pozwala na badanie trendów i zmian w czasie, dzięki dostępności historycznych danych.

Potencjalne bariery (1)

Analiza danych zastanych jest efektywnym sposobem zbierania informacji, jednak napotyka na szereg barier, które mogą wpływać na jakość i użyteczność wyników.



Ograniczona kontrola nad danymi: Badacz nie ma wpływu na to, jak dane zostały zebrane, przetworzone i zaprezentowane.

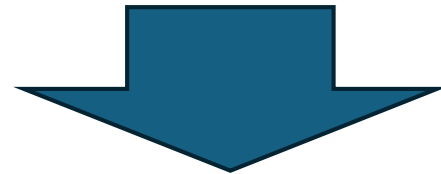


Badaczowi trudno ocenić wiarygodność, dokładność i jakość danych.

Potencjalne bariery (2)

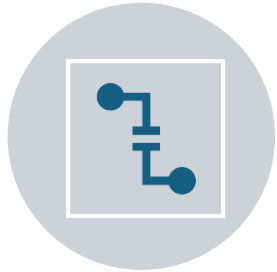


Ograniczona wiedza: Badacz może nie posiadać szczegółowej wiedzy na temat metodologii zbierania (np. techniki badawcze, wielkość i dobór próby) lub przetwarzania danych (np. jak czyszczono zbiory, jak agregowano zmienne, jakie procedury ważenia zastosowano)

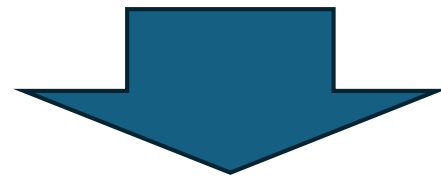


- Badaczowi trudno ocenić wiarygodność, dokładność i jakość danych.
- Mogą występować błędy metodologiczne, interpretacyjne, pojęciowe.

Potencjalne bariery (3)



Brak aktualności danych: Dane zastane mogą być przestarzałe i nieodzwierciedlające obecnych realiów rynkowych, społecznych czy technologicznych.

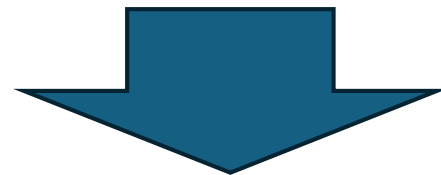


- Może to prowadzić do błędnych analiz, rekomendacji i strategii, nieadekwatnych do obecnych realiów.
- Np. dane demograficzne mogą pochodzić z ostatniego spisu ludności, który był przeprowadzony kilka lat temu.

Potencjalne bariery (4)



Brak specyficzności: Dane zastane często są zbierane w szerokim kontekście i mogą nie odpowiadać specyficznym potrzebom badawczym danego projektu.

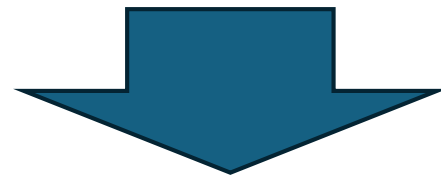


Może utrudniać precyzyjne odpowiedzi na pytania badawcze i prowadzić do konieczności uzupełniania badania innymi metodami pozyskania danych np. badania pierwotne.

Potencjalne bariery (5)



Problemy z dostępnością danych: Nie wszystkie dane są łatwo dostępne. Niektóre mogą być chronione prawami autorskimi, objęte tajemnicą statystyczną lub handlową, czy dostępne tylko za opłatą.

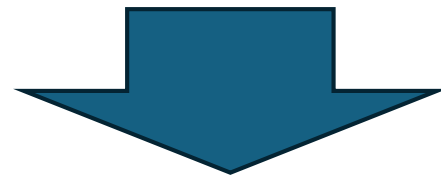


Może utrudniać kompleksową analizę i wymagać dodatkowych środków na zakup dostępu do potrzebnych informacji.

Potencjalne bariery (6)



Kompatybilność danych: Dane pochodzące z różnych źródeł mogą być zbierane przy użyciu różnych metodologii, formatów, struktur.

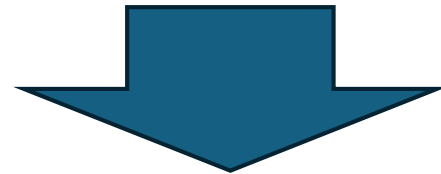


- Może to utrudniać łączenie i analizę danych, powodować trudności metodologiczne i interpretacyjne.
- Może wymagać dodatkowego czasu i zasobów na przetworzenie, ujednoczenie przed rozpoczęciem analizy.
- Różne definicje i kategorie używane w różnych badaniach mogą prowadzić do niespójności definicyjnej, interpretacyjnej.

Potencjalne bariery (7)




Błędy w interpretacji: Różni badacze mogą wyciągać różne wnioski na podstawie tych samych danych, jeżeli nie znają w pełni lub nie rozumieją kontekstu procesu badawczego oraz przetwarzania danych



Subiektywność i błędy interpretacyjne, które mogą prowadzić do różnych wniosków i rekomendacji, wpływając na jakość i spójność wyników badania.

Ocena danych statystyki publicznej



- ⑩ Rzetelność - definicje, duże lub pełne próby


- ⑩ Wiarygodność - działanie zgodnie ze standardami i regulacjami

- ⑩ Spójność i dokładność - zbierane systematycznie, przy użyciu standardowych metodologii

- ⑩ Szeroki zakres tematyczny
Szeregi czasowe

- ⑩ Dane regionalne

- ⑩ Publiczny dostęp



- ⑩ Aktualność - brak większości danych w czasie rzeczywistym, niektóre dane (np. spisy powszechne) są aktualizowane co kilka/kilkanaście lat

- ⑩ Dostępność - trudność w użytkowaniu (poruszaniu się po bazach danych, format danych)

- ⑩ Brak specyficzności - brak potrzebnych przekrojów, zagadnień, standardowe kategorie i definicje

- ⑩ Tajemnica statystyczna

Konkluzje

Istniejące dane statystyczne mogą stanowić ważne uzupełnienie procesu badawczego w badaniach opinii i rynku.

Dane statystyki publicznej dają gwarancję jakości prezentowanych danych, również w długich szeregach czasowych.

Możliwość zastosowania w projektach badawczych analizy danych zastanych zależy od tego, jakie dane mamy do dyspozycji, czego dotyczą, co opisują, czy odpowiadają na nasze potrzeby.

Kontakt

Edyta Makowska-Belta

Starszy Konsultant ds. Badań

Fundacja Instytut Badań Rynkowych i
Społecznych

Tel.: 507-082-085

E-mail: e.makowskabelta@ibris.pl