

# Ogólny schemat harmonizacji danych z badań panelowych

Zbyszek Sawiński  
Instytut Filozofii i Socjologii PAN

Anna Kiersztyn  
Uniwersytet Warszawski

Metodologiczne Inspiracje 2023  
Badania Ilościowe w Naukach Społecznych –  
- Wyzwania i Problemy

21-22 września 2023, Jabłonna



NATIONAL SCIENCE CENTRE  
POLAND



# Plan prezentacji

- Wyzwania harmonizacji Narodowych Badań Panelowych
- Schemat harmonizacji: prezentacja podejścia
- Implementacja w projekcie CNB-Young (na podstawie POLPAN i SOEP)
- Co dalej?

# Perspektywy harmonizacji badań panelowych

- Rozwój badań panelowych na świecie
- Inicjatywy harmonizacji krajowych badań panelowych
  - European Community Household Panel (1994-2001; 15 krajów)
  - Cross-National Equivalent File (od 1990, 10 krajów)
  - Comparative Panel File (od 2021, 7 krajów)

# Różnorodność rozwiązań badawczych

- Częstość kontaktów z badanymi
  - Szczegółowość danych
  - Błędy pamięci
- Sposób pytania o zdarzenia i epizody:
  - Metoda sekwencyjna
  - Metoda kalendarza
  - Podejścia mieszane

# Geneza propozycji

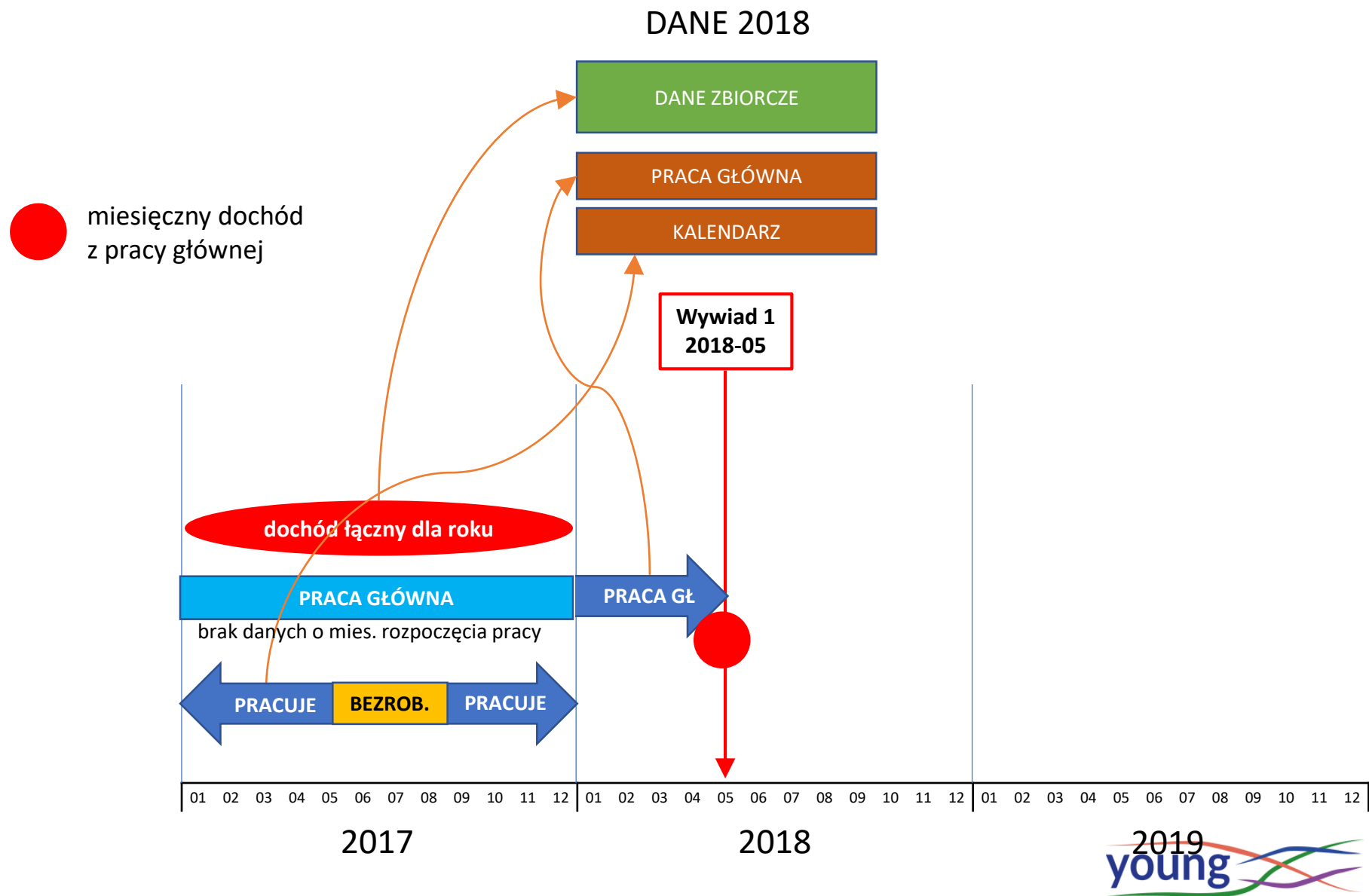
- Projekt CNB-Young (Cross National Biographies – Young), realizowany w latach 2019-2024 przez Uniwersytet Warszawski i IFiS PAN

<https://crossnationalbiographies.edu.pl/>

# Elementy schematu

- Wspólna skala lat i miesięcy
- Uzupełnienie indywidualnych danych o terminy badań
- Odniesienia do pełnych danych

# Obraz historii zatrudnienia i dochodów po pierwszej fali panelu







# Korzyści

- Niskie koszty wejścia
- Minimalna strata informacji
- Elastyczność zastosowań

# Ilustracja

- Rozwiązanie wypracowane na potrzeby CNB-Young dla analiz przebiegu karier
- Cel: maksymalizacja elastyczności przy zachowaniu prostej struktury
- Implementacja schematu dla POLPAN i SOEP
- POLPAN – realizacja raz na 5 lat, dane o karierach gromadzone retrospektywnie dla wszystkich prac od roku ostatniej fali
- SOEP – realizacja raz na rok, szczegółowe dane o pracy głównej w momencie badania (oraz trochę informacji o ew. pracy dodatkowej), osobne pytania o zmianę prac; kalendarz aktywności za poprzedzający rok

# Konstrukcja zbiorów danych

- Jednostka: lata kalendarzowe (*person-years*)
  - Łatwość przejścia na wiek badanych osób
  - Najprostszy zapis w przypadku danych z corocznych badań, np. SOEP
  - Często pyta się tylko o rok, w którym zdarzenie miało miejsce
- Uzupełnienie: zmienne miesięczne
  - Kody informujące o statusie danej aktywności - *Activity Status Codes*
  - Osobne zestawy 12 zmiennych dla każdej aktywności, na temat których gromadzone są informacje i które mogą się na siebie nakładać (praca, nauka, bezrobocie, zajmowanie się domem, przerwy w karierze spowodowane chorobą / urlopem rodzicielskim, ...)

# Rodzaje zmiennych

- Podstawowe charakterystyki respondenta i danych

W tym historia udziału w panelu

Daty poszczególnych fal umożliwiające proste obliczenie głębokości retrospekcji (RDS – retrospective distance scale) dla poszczególnych wydarzeń życiowych / aktywności.

- **Charakterystyki aktywności** (*economic activity spells*)

Praca i pozostałe.

- **Zmienne roczne**

Ułatwienie analiz prostych wskaźników przebiegu karier – ale i większa elastyczność w zapisie informacji z danych źródłowych.

# Epizod pracy (job spell)

- Praca aktualna i wg kalendarza (SOEP) kodowane jako osobne „epizody”.
- **Osobne epizody pracy dla różnych fal.** Zalety rozwiązania:
  - Brak konieczności arbitralnych wyborów przy łączeniu prac „na szwach” (np. POLPAN)
  - porównania z surveyami nie stosującymi dependent interviewing (NLSY vs POLPAN)
  - zachowanie informacji o danych źródłowych (jedna fala – jeden epizod)
  - uwzględnienie zmian charakterystyk pracy z fali na falę
- Zbiór zawiera informacje umożliwiające łączenie epizodów będących tą samą pracą

# Charakterystyki „epizodów prac”

- Kod typu epizodu pracy (praca główna; praca dodatkowa; praca zakończona; inne – w SOEP prace wg kalendarza, w POLPAN prace za granicą)
- Zmienne metodologiczne np. czy zastosowano dependent interviewing / ID tej pracy w poprzedniej fali
- Data rozpoczęcia / zakończenia pracy
- Wybrane charakterystyki pracy (inne można dołączyć korzystając z odwołania do danych źródłowych)
- Miesięczne kody statusu (*Activity Status Codes*; ASC)

# Kody statusu dla każdego „epizodu pracy”

- 000 no activity this month
  - 101 intermediate month of activity
  - 111 first month of activity reported
  - 112 first month of activity censored  
**(miesiąc po poprzedniej fali)**
  - 121 last month of activity reported
  - 122 last month of activity censored  
**(praca trwa w momencie badania)**
  - 131 first and last month of activity reported (praca trwa 1 m-c)
  - 132 first or last month of activity censored
- 
- 001 brak danych
  - 101 osobny kod braków danych jeśli znamy rok rozpoczęcia / zakończenia pracy ale nie znamy miesiąca**

Druga cyfra pozwala łatwo określić, czy praca wykonywana w styczniu jest kontynuacją pracy z poprzedniego roku / czy grudzień jest ostatnim miesiącem pracy.

Trzecia cyfra odróżnia początek / koniec pracy wg daty podanej w wywiadzie od danych odciętych.

Jeśli praca zaczęła się przed poprzednią falą, dane traktujemy jako odcięte.

# Roczne podsumowania

- Celem jest nie tylko ułatwienie pracy na danych – ale przede wszystkim lepsze dostosowanie do specyfiki danych. Np. dla SOEP:
  - pytanie o łączny dochód z pracy w poprzednim roku.
  - nie da się z kalendarza określić liczby prac, ale jest osobne pytanie o to, ile razy R. zmieniał pracę w ciągu ostatniego roku.
- A ogólniej:
  - Czasem wiemy, ile tygodni lub miesięcy trwała dana aktywność (np. przerwa w zatrudnieniu), ale brak danych o miesiącu jej rozpoczęcia, które pozwoliłyby na zapis tej informacji w ASC



# Formuły implementacji propozycji

- Zależnie od możliwości trwałej współpracy z dysponentem danych i zasad udostępniania danych
  - Zmienne dołączone do dystrybuowanego zbioru danych
  - Osobny zbiór danych
  - Jedynie w formie skryptu
- Pytanie o strategię dystrybucji i zapewnienia widoczności wśród potencjalnie zainteresowanych badaczy.
- Problem sustainability – co po zakończeniu projektu?

# Podsumowanie

- W przypadku danych biograficznych z badań panelowych dotychczasowe podejścia do harmonizacji wymagają uzupełnienia o etap wstępny.
- Tym etapem jest uproszczenie i ujednolicenie formatu danych (layout, zestaw zmiennych, jednostka, system kodowania) bez straty informacji i z możliwością „dotarcia do źródła”.
- Ujednolicenie formatu pozwala „zobaczyć różnice” i w bardziej świadomy sposób podejść do analiz porównawczych i harmonizacji – jednak konkretne decyzje co do dalszej obróbki danych pozostawione są użytkownikom.
- Można sprawdzić, jak te decyzje wpływają na wynik analiz.

Dziękujemy za uwagę.

**Dynamics of youth employment precarity:  
Drivers, trajectories, and outcomes in a cross-national perspective (CNB-Young)**

Grant OPUS NCN nr. 2018/31/B/HS6/02043

[crossnationalbiographies.edu.pl](http://crossnationalbiographies.edu.pl)

**[Working paper: [chaber@is.uw.edu.pl](mailto:chaber@is.uw.edu.pl)]**

