

# Warsztat badacza: Repozytoria danych

22 maja / 16 czerwca 2020 r.

Marta Kołczyńska

Instytut Studiów Politycznych Polskiej Akademii Nauk

[mkolczynska@gmail.com](mailto:mkolczynska@gmail.com)

# Plan dnia

Prezentacja

Projekty wtórne – przykłady

Wikipedia

# Analiza wybranych danych wtórnych

1. Pytanie badawcze
2. Opis danych (dostosowany do typu danych)
  - Rodzaj danych, źródło / sposób pozyskania
  - Adekwatność wybranych danych
  - Populacja, typ próby, poziom realizacji, rozmiar próby
  - Opis wybranych zmiennych; treść pytań i kategorie odpowiedzi
  - Problemy dot. jakości danych i ich wpływ na analizy
3. Analiza
  - Metoda analizy
  - Tabele / wykresy / schematy
4. Wnioski

Prezentacja 15-20 minut, 16 czerwca

Tekst 3-5 stron, do końca czerwca (mkolczynska@gmail.com)

Projekty wtórne

# IPUMS



U.S. Census and American Community Survey microdata from 1850 to the present. [Learn More](#)

VISIT SITE



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present. [Learn More](#)

VISIT SITE



World's largest collection of census microdata covering nearly 100 countries, contemporary and historical. [Learn More](#)

VISIT SITE



Health survey data for Africa and Asia, including harmonized data collections for [DHS](#) and [PMA](#). [Learn More](#)

VISIT SITE



Tabular U.S. Census data and GIS boundary files from 1790 to the present. [Learn More](#)

VISIT SITE



Integrated data on population and the environment from 1960 to the present. [Learn More](#)

VISIT SITE



Historical and contemporary time use data from 1965 to the present. [Learn More](#)



Historical and contemporary U.S. health survey data from [NHIS](#) (1963-present) and [MEPS](#) (1996-present). [Learn More](#)



Survey data on the science and engineering workforce in the U.S. from 1993 to the present. [Learn More](#)

# Trzy pokolenia polskiej transformacji

#3GEN ERACJE POLSKIEJ TRANSFORMACJI  
#3GEN ERATIONS OF THE POLISH TRANSFORMATION

Ireneusz Sadowski, ISP PAN: [3gen.spoleczenstwo.pl/](http://3gen.spoleczenstwo.pl/)

Dane wtórne:

1988: pokolenie ur. 1957-59

2001/2003: pokolenie ur. 1957-59, 1970-71

Dane pierwotne:

2019: pokolenie ur. 1957-59, 1970-71, 1988-89

# Harmonizacja/łączenie danych sondażowych

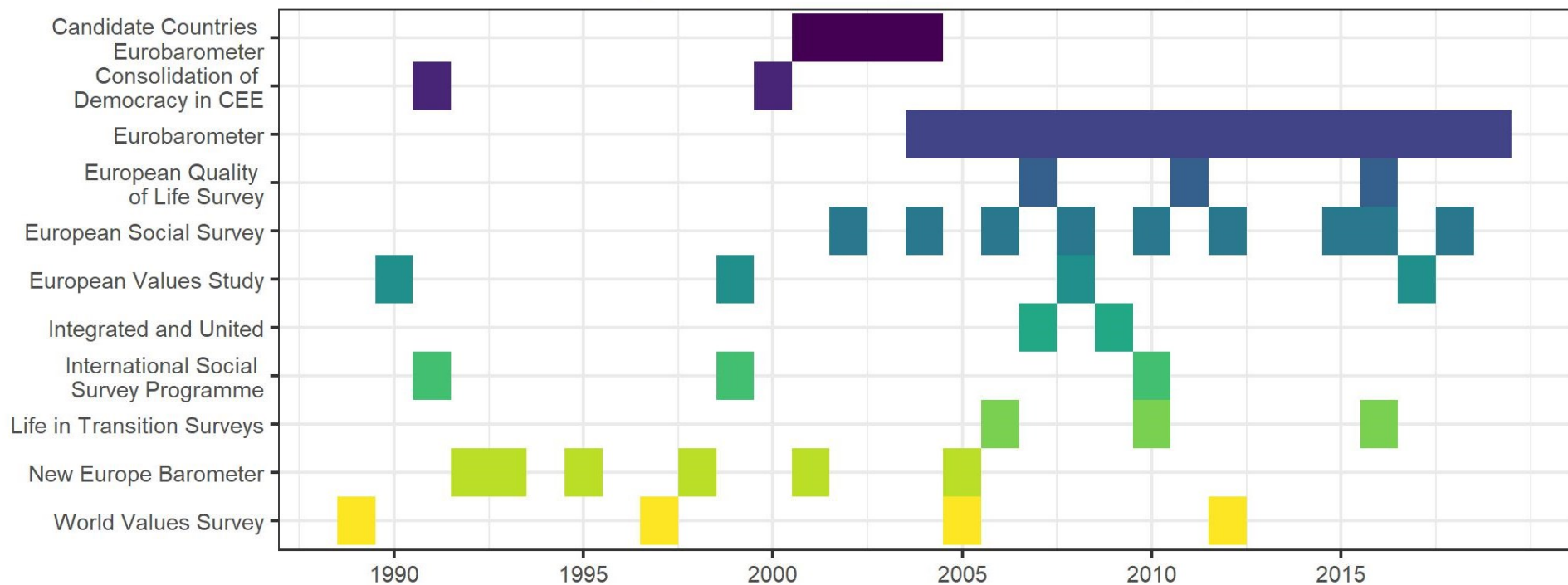
*Survey Data Recycling*, [dataharmonization.org](http://dataharmonization.org)

*Old and new boundaries: National Identities and Religion*,  
[onbound.international](http://onbound.international)

*Uwarunkowania i konsekwencje zaufania politycznego*

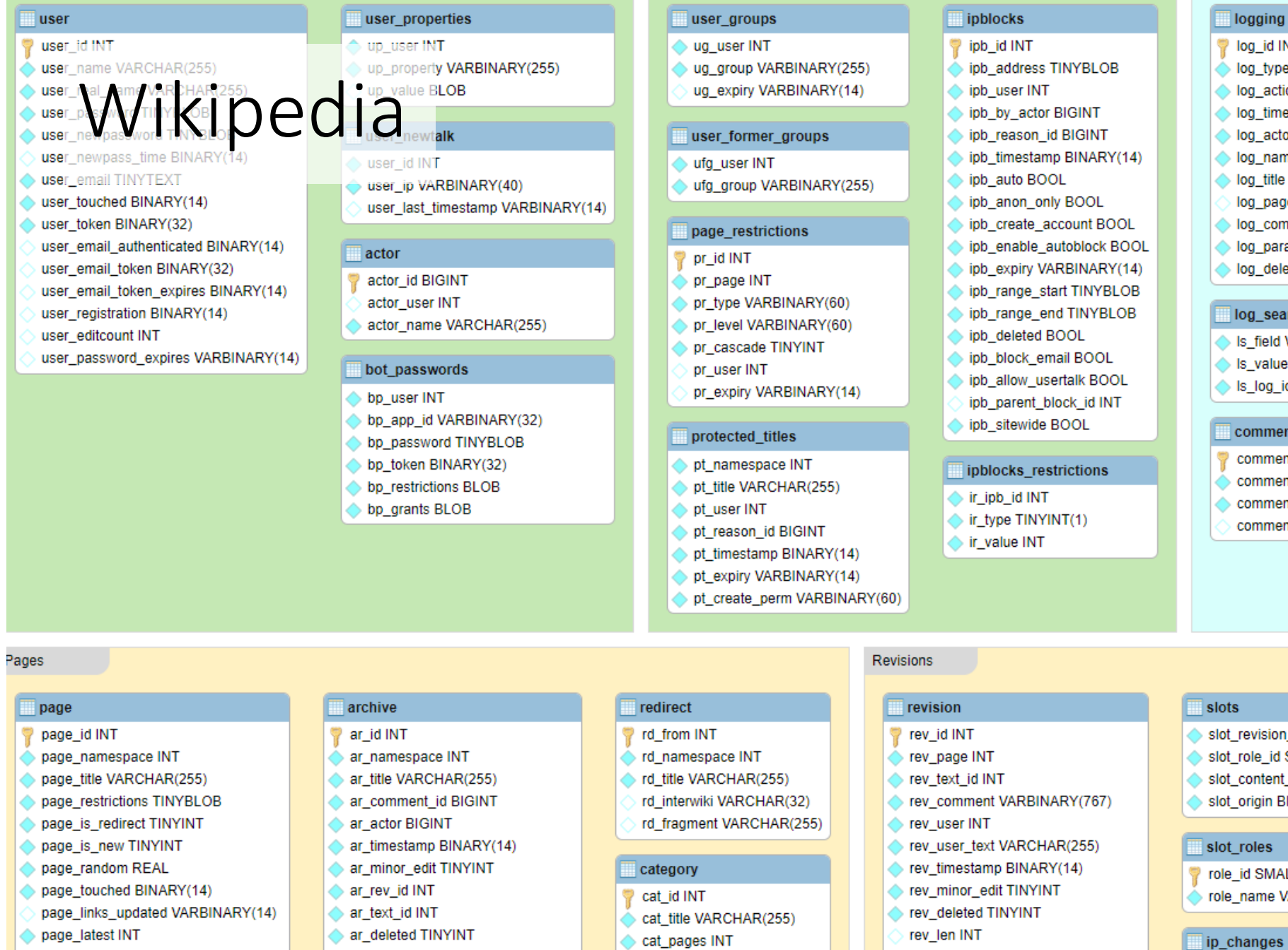
# Harmonizacja/łączenie danych sondażowych

Cross-national surveys with trust in parliament items, Poland, 1989-2019





# Wikipedia



# Wikipedia: Zapytania



Quarry Beta

[Home](#)

[New Query](#)

[Recent Queries](#)

[Discuss](#)

[Database tables](#) ▾

Mkolczynska ▾

This query has been published by [The Polish](#).

SQL

[Toggle Highlighting](#)

```
USE plwiki_p;
SELECT CONCAT("[[User:",user_name,"]]") AS user_name, ug_group, user_editcount, COUNT(actor_id) AS recent_user_editcount
FROM user
LEFT OUTER JOIN actor ON actor_user = user_id
LEFT OUTER JOIN revision_actor_temp ON revactor_actor = actor_id
LEFT OUTER JOIN user_groups ON ug_user = actor_user
WHERE actor_user IS NOT NULL
AND revactor_timestamp BETWEEN 20200425000000 AND 20200525000000
AND ug_group LIKE 'autoreview'
/*AND (ug_group NOT IN ('autoreview','sysop','editor','interface-admin','abusefilter','oversight','ipblock-exempt','accountcreator',
GROUP BY user_name
ORDER BY recent_user_editcount DESC;
```

All SQL code is licensed under [CC0 License](#).

Query status: **complete**

Executed in 19.11 seconds as of Wed May 27 2020.

Resultset (175 rows)

[Download data](#) ▾

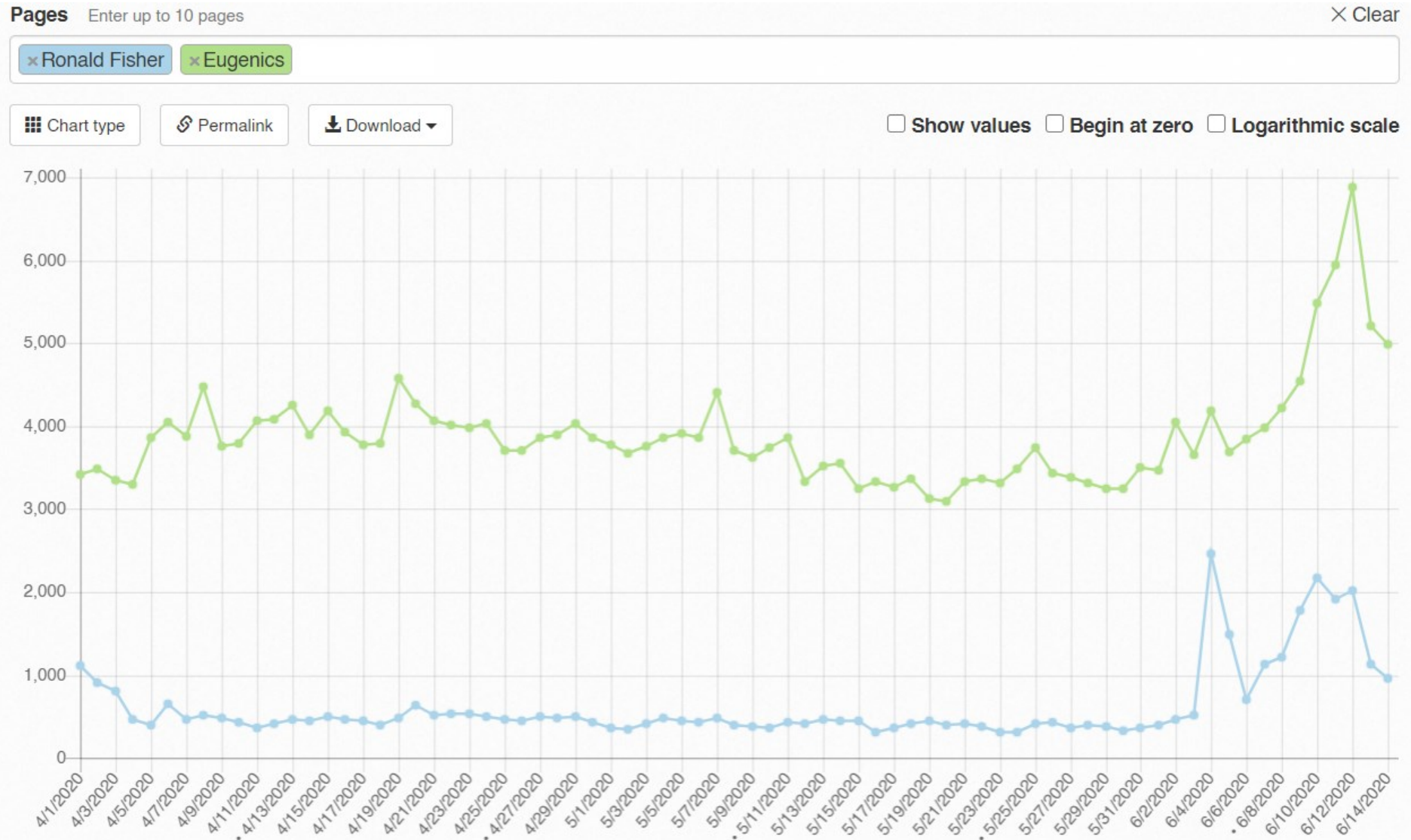
Show  entries

Search:

user_name	ug_group	user_editcount	recent_user_editcount
[[User:Macuk]]	autoreview	270068	5547
[[User:Lukasz2]]	autoreview	23322	2834

<https://quarry.wmflabs.org/> ; <https://quarry.wmflabs.org/query/23322>

# Wikipedia: Liczba odsłon wg stron



<https://pageviews.toolforge.org/> ; [https://en.wikipedia.org/wiki/Wikipedia:Web\\_statistics\\_tool](https://en.wikipedia.org/wiki/Wikipedia:Web_statistics_tool)

# Linki

Otwarte dane jakościowe

<https://journals.sagepub.com/doi/pdf/10.1177/1609406918823863>

<https://ukdataservice.ac.uk/use-data/guides/methods-software/qualitative-reuse>

# Podziękowania



N A R O D O W E   C E N T R U M   N A U K I

Prezentowane treści powstały w ramach projektu „Uwarunkowania i konsekwencje zaufania politycznego: Polaryzacja polityczna i demokratyczna użyteczność zaufania w perspektywie porównawczej” (2019/32/C/HS6/00421).