

# Czego i w jaki sposób można dowiedzieć się z wielkich bibliografii?

prof. dr hab. Adam Pawłowski  
Uniwersytet Wrocławski

dr inż. Tomasz Walkowiak  
Politechnika Wrocławska

28 listopada 2023

# Plan prezentacji

- **POKAZ** [<http://phc.uni.wroc.pl/wbw>]
- **DANE**  
Bibliografia jako źródło danych  
Dane ustrukturyzowane i format MARC  
Bazy Biblioteki Narodowej
- **METODA**  
Automatyczne przetwarzanie języka (NLP)  
Humanistyka cyfrowa i data science
- **WYNIKI**  
Bibliomapa jako przekrój historyczny rozwoju Polski  
Świat w polskich publikacjach  
Płeć a liczba i tematy publikacji  
Analiza dynamiki publikowania w miastach Polski

# Dane – typy bibliografii

Potocznie bibliografia to krótki załącznik do dokumentu.

Dla informatyka to **ustrukturyzowany korpus danych**.

**Podstawowy podział typów bibliografii obejmuje:**

- ogólne (narodowe);
- specjalne (ograniczony zasięg, zakres, typ publikacji)
- dziedzinowe (temat)
- mieszane

Żartując w stylu Orwella, wszystkie bibliografie są ciekawe, ale bibliografie narodowe są ciekawsze.



# Bibliografia jako źródło danych

**Tradycyjnie odbiorcą informacji bibliograficznej jest użytkownik, który poszukuje publikacji na zadany temat.**

**De facto otrzymuje więc nie informację, a metainformację.**

**ale...**

**Wystarczająco obszerny, ustrukturyzowany korpus danych jest jednocześnie źródłem wiedzy.**

**W świecie humanistyki cyfrowej wielkie bibliografie są więc także prymarnym źródłem informacji i wiedzy o świecie.**

# Formaty

**W bazach tego rodzaju dominuje rodzina formatów MARC, czyli Machine-Readable Cataloging, norma ISO 2709:2008.**

Pierwsza wersja powstała w Bibliotece Kongresu w **latach 60.**, od 1971 była standardem w USA.

**Najbardziej rozpowszechnioną wersją jest dziś MARC 21.**

**Większość dorobku piśmienniczego ludzkości jest więc opisana standardem, liczącym dziś ponad 60 lat.**

# Dygresja: jak długo jeszcze?

The MARC standard format is dying!



PAUL MOSS

## Replacing MARC: Where to Start

The MARC format for transmittal of bibliographic records has been an unparalleled success for interlibrary communication. This success, however, has also brought about a world that deals exclusively in MARC and is inherently bound by its limitations. Several efforts over the years have been made to break free of these perceived limitations, but these efforts often miss the crucial mis-step of MARC-like thinking: that a library interchange format should be the only way to ingest, expose, or build systems around bibliographic data. In order to begin a transition away from MARC, each function MARC serves should be examined independently and may be replaced by a different technology.

In June 2016 the Acquisitions & Bibliography  
October 2015 to test the efficacy of BIBFRAME.

# Format MARC 21 – przykład instrukcji

## 041 Język dokumentu (P)

### Wskaźniki

Pierwszy - wskaźnik przekładu

- 0 dokument nie jest przekładem i nie zawiera przekładów
- 1 dokument jest przekładem lub zawiera przekład

Drugi - źródło kodu

- @ *MARC Code List for Languages.*

### Kody podpół

- \$a Język tekstu (P)
- \$b Język streszczeń (P)
- \$h Język tekstu oryginalnego i/lub pośredniego (P)

### Zawartość pola

Pole 041 zawiera informację dotyczącą języków dokumentu, jeżeli nie jest możliwe umieszczenie pełnej informacji o językach dokumentu w polu 008/35-37. Ma zastosowanie w opisie dokumentów wielojęzycznych (także jeśli jednym z języków jest język migowy) lub zawierających tłumaczenie oraz jeśli język streszczeń, abstraktów lub materiałów towarzyszących jest inny niż język tekstu głównego.

Informację o języku podaje się w postaci trzyliterowego kodu według *MARC Code List for Languages*.

Pole 041 używane jest w połączeniu z pozycjami 35–37 pola 008. Kod podany w tych pozycjach powinien być identyczny z kodem wpisanym w pierwszym podpolu \$a pola 041.

Informacje o języku mogą być podawane także w formie słownej w polu 546.

### Znaki umowne

Na końcu pola 041 nie stawia się kropki.

### Wskaźniki

Pierwszy wskaźnik - wskaźnik przekładu

- 0 - dokument nie jest przekładem i nie zawiera przekładów
- 008/35-37 eng
- 041 0@ \$a eng \$a fre

*dokument nie jest przekładem, tekst jest w językach angielskim i francuskim*

**1 - dokument jest przekładem lub zawiera przekład**

008/35-37 eng

041 1@ \$a eng \$h rus

*dokument jest przekładem z języka rosyjskiego na angielski*

**Drugi wskaźnik - źródło kodu**

@ - *MARC Code List for Languages*

Wartość @ (spacja) oznacza, że kody podawane w polu 041 są przejmowane z *MARC Code List for Languages*.

**Podpola**

**\$a - Język tekstu**

Podaje się trzyliterowe kody języków tekstu dokumentu podane w porządku ich dominacji w tekście lub w porządku alfabetycznym, jeżeli żaden z języków nie dominuje. Dla każdego kodu języka tekstu powtarza się podpole \$a. Pierwszy kod powinien być zgodny z kodem wymienionym w polu stałej długości (008/35-37). Jeżeli tekst jest w bardzo wielu językach (np. więcej niż sześciu), a w pozycjach 35-37 pola 008 podano kod „mul”, w pierwszym podpolu \$a podaje się kod języka tytułu dokumentu lub pierwszego tytułu (jeżeli jest on w więcej niż jednym języku), w drugim podpolu \$a podaje się kod „mul” (wiele języków).

Jeżeli dokument zawiera tekst oryginalny i towarzyszący mu przekład, w powtórzonym podpolu \$a podaje się zarówno kod języka przekładu, jak i oryginału. Język oryginału wymienia się także w podpolu \$h.

008/35-37 pol

041 0@ \$a pol \$a eng \$a rus

546 @@ \$a Tekst pol., ang., ros.

*tekst w języku angielskim, polskim, rosyjskim z przewagą języka polskiego*

**\$b - Język streszczeń**

Podaje się w porządku alfabetycznym kody języków streszczeń różnych od języka tekstu głównego. Dla każdego kodu języka streszczeń powtarza się podpole \$b.

041 0@ \$a eng \$a mul \$b eng \$b fre

*tekst w więcej niż sześciu językach z przewagą języka angielskiego, streszczenia w języku angielskim i francuskim*

# Format MARC 21

**Opis wszystkich pól formatu MARC 21 mieści się na 175 stronach podręcznika.**

**Albo w wielu grubych tomach.**





# Dane ustrukturyzowane

**Dane ustrukturyzowane zawierają powtarzalny układ pól. Są łatwe w przetwarzaniu, ale płytkie pod względem informacyjnym.**

008/00-05	Data wprowadzenia rekordu
008/06	Kod statusu publikacji
008/07-14	Daty publikacji
008/15-17	Kod miejsca publikacji
008/18	Kod częstotliwości
008/19	Kod regularności
008/20	Zawartość nieokreślona
008/21	Kod typu dokumentu ciągłego
008/22	Kod formy zapisu oryginału
008/23	Kod formy zapisu
008/24	Kod formy piśmienniczej całości dokumentu ciągłego
008/25-27	Kod formy piśmienniczej poszczególnych części
008/28	Kod publikacji urzędowej
008/29	Kod materiałów z imprezy
008/30-32	Zawartość nieokreślona
008/33	Kod alfabetu użytego w tytule
008/34	Kod zasad opisu
008/35-37	Kod języka tekstu
008/38	Kod zmiany formy zapisu w stosunku do występującej w dokumencie
008/39	Kod źródła opisu

## 600-651 Hasła przedmiotowe

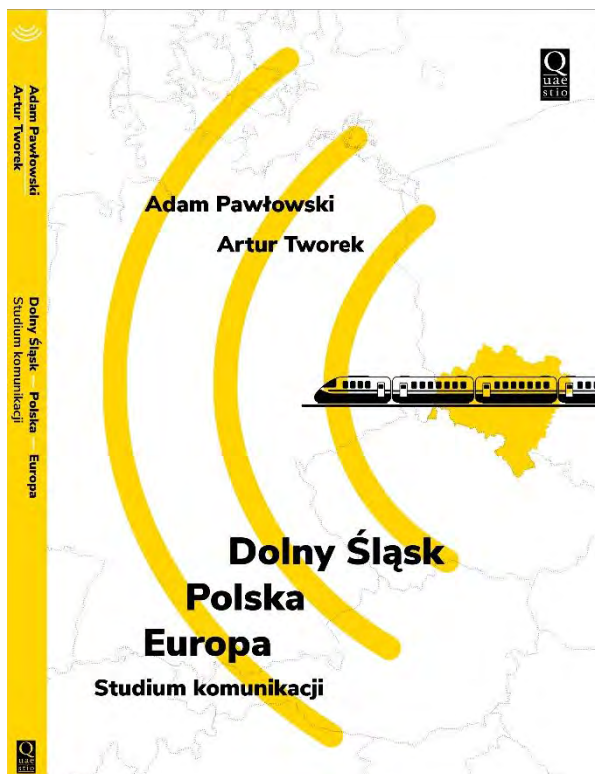
- 600 Hasło przedmiotowe - nazwa osobowa (P)
- 610 Hasło przedmiotowe - nazwa ciała zbiorowego (P)
- 611 Hasło przedmiotowe - nazwa imprezy (P)
- 630 Hasło przedmiotowe - tytuł (P)
- 650 Hasło przedmiotowe - nazwa pospolita (P)
- 651 Hasło przedmiotowe - nazwa geograficzna (P)

## 700-740 Hasła dodatkowe

- 700 Hasło dodatkowe - nazwa osobowa (P)
- 710 Hasło dodatkowe - nazwa ciała zbiorowego (P)
- 711 Hasło dodatkowe - nazwa imprezy (P)
- 740 Hasło dodatkowe - tytuł (P)



# Rekordy a humanistyka cyfrowa



<b>Title</b>	Dolny Śląsk - Polska - Europa: studium komunikacji / Adam Pawłowski, Artur Tworek.
<b>Creator</b>	Pawłowski, Adam (historia) Autor > Tworek, Artur (1965- ) Autor >
<b>Format</b>	Oficina Wydawnicza Atut Wydawca > 334 strony : faksymilia, fotografie, ilustracje, mapy, wykresy ; 29 cm + mapa.
<b>Publisher</b>	Wrocław : Oficyna Wydawnicza Atut - Wrocławskie Wydawnictwo Oświatowe
<b>Creation Date</b>	2021
<b>Note</b>	Bibliografia na stronach 235-246. Indeksy. Współfinansowanie: Uniwersytet Wrocławski
<b>Form and type</b>	Książki Publikacje naukowe
<b>Subject</b>	1801-1900 > 1901-2000 > Komunikacja społeczna > Regionalizm > Transport > Dolny Śląsk > Województwo dolnośląskie (1999- ) >
<b>Genre</b>	Opracowanie
<b>Discipline and Approach</b>	Historia Media i komunikacja społeczna Socjologia i społeczeństwo Transport i logistyka
<b>Identifier</b>	ISBN : 9788379777082
<b>UKD</b>	913(438) 316.77(438) 656
<b>Language</b>	Polish

#### Legenda:

- – toponim (rzutowanie na mapę)
- – antroponim (rzutowanie na bazę bibliograficzną)
- – słowo kluczowe (rzutowanie na ontologię)
- – gatunek piśmienniczy
- – numer klasyfikacji UKD (rzutowanie na słowniki i bazy leksykalne)

# Dane: bazy Biblioteki Narodowej

**Polska nie ma obecnie pełnej Bibliografii Narodowej.**

**Dysponuje jednak katalogami Biblioteki Narodowej, które po 1918 r. są niemal kompletne.**

**Katalog BN = zawartość magazynu**

**Bibliografia Narodowa = wszystko, co spełnia kryteria (ale nie musi być fizycznie w zasobach)**

**Katalog BN < Bibliografia narodowa**

# Dane: bazy Biblioteki Narodowej



## Lista dostępnych API

Rekordy bibliograficzne (Biblioteka Narodowa)

Rekordy bibliograficzne (połączone katalogi)

Rekordy wzorcowe

Bazy do pobrania

## Dokumentacja

Dokumentacja

# Dane: bazy Biblioteki Narodowej

## Bazy bibliograficzne

### Biblioteka narodowa

- **bibs-all.marc** 11,3 GB, 10 paź 08:02
- **bibs-artykul.marc** 4,3 GB, 10 paź 08:03
- **bibs-audiobook.marc** 33,5 MB, 10 paź 08:03
- **bibs-czasopismo.marc** 231 MB, 10 paź 08:03
- **bibs-dokument\_elektroniczny.marc** 10,8 MB, 10 paź 08:03
- **bibs-druk\_ulotny.marc** 688 MB, 10 paź 08:03
- **bibs-e\_book.marc** 203 MB, 10 paź 08:03
- **bibs-film.marc** 93,2 MB, 10 paź 08:03
- **bibs-fotografia.marc** 222 MB, 10 paź 08:03
- **bibs-grafika.marc** 167 MB, 10 paź 08:03
- **bibs-ksiazka.marc** 4,38 GB, 10 paź 08:03
- **bibs-mapa.marc** 257 MB, 10 paź 08:03
- **bibs-nagrania\_dzwiekowe.marc** 230 MB, 10 paź 08:03
- **bibs-nuty.marc** 220 MB, 10 paź 08:03
- **bibs-obiekt\_trojwymiarowy.marc** 1,92 MB, 10 paź 08:03
- **bibs-rekopis.marc** 76,1 MB, 10 paź 08:03
- **bibs-rysunek.marc** 42,9 MB, 10 paź 08:03
- **bibs-starodruk.marc** 155 MB, 10 paź 08:03

# Metody

## 3 filary metodologiczne

- Programowanie w zakresie baz danych, łączenia danych i wizualizacji (szczególnie na mapach).
- Przetwarzanie języka naturalnego (NLP)
- Humanistyka cyfrowa (interpretacja, selekcja tematów, podejście holistyczne i integrujące)



## 3 filary metodologiczne

### Zbiór algorytmów i zasobów pozwalających na automatyczną analizę języka (tekstu).

- rozpoznanie form podstawowych (*domowi* → *dom*)
- rozpoznanie struktur składniowych
- analizy ilościowe
- analizy semantyczne

Zastosowanie: analiza pól tekstowych rekordu (np. tytułów)

# Humanistyka cyfrowa

**System powiązanych ze sobą zasobów, funkcjonalności i praktyk poznawczych, powstały dzięki przeniesieniu do sfery cyfrowej i twórczej rozbudowie dziedzictwa humanistyki wyrastającej z kultury druku.**

# Wyniki cz. 1

- rzutowanie rekordów na mapę
- przekrój historyczny

# Skąd brać współrzędne?



## GeoNames

The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge.

  
 [\[advanced search\]](#)

enter a location name, ex: "Paris", "Mount Everest", "New York"

# Jedna nazwa – wiele miejsc

Cieplice  all countries  [\[advanced search\]](#)

286 records found for "Cieplice"

Name	Country	Feature class	Latitude	Longitude
1 <a href="#">Cieplice Śląskie Zdrój</a> Bad Warmbrunn,Cieplice Slaskie Zdroj,Cieplice Slaskie-Zdroj,Cieplice Śląskie Zdrój,Cieplice Śląskie-...	<a href="#">Poland</a> , Lower Silesia Jelenia Góra > Jelenia Góra	populated place population 15,000	N 50° 51' 55"	E 15° 41' 1"
2 <a href="#">Teplice</a> Cieplice,Teplice,Teplice-Sanov,Teplice-Šanov,Teplicé,Teplitse,Teplitz,Teplitz-Schoenau,Teplitz-Schön...	<a href="#">Czechia</a> , Ústecký kraj Teplice District > Teplice	populated place population 51,223	N 50° 38' 25"	E 13° 49' 28"
3 <a href="#">Cieplice</a> Cieplice,Kolonia Cieplice	<a href="#">Poland</a> , Lesser Poland Powiat miechowski > Gmina Golcza	populated place	N 50° 21' 13"	E 19° 54' 59"
4 <a href="#">Cieplice</a>	<a href="#">Poland</a> , Subcarpathia Powiat przeworski > Adamówka	populated place	N 50° 15' 6"	E 22° 37' 41"
5 <a href="#">Trenčianske Teplice</a> Trencanske Teplice,Trench'janske Teplice,Trench'janske-Teplice,Trenchinske Tjeplice,Trencianske Tpepl...	<a href="#">Slovakia</a> , Trenčiansky kraj Trenčín District > Trenčianske Teplice	populated place population 4,438, elevation 267m	N 48° 54' 38"	E 18° 10' 0"
6 <a href="#">Teplice District</a> Daerah Teplice,Distretto di Teplice,District de Teplice,Districte de Teplice,Distrikto Teplice,Okres...	<a href="#">Czechia</a> , Ústecký kraj Teplice District	second-order administrative division population 128,610	N 50° 38' 0"	E 13° 50' 0"
7 <a href="#">Duchcov</a> Duchcov,Dukhcov,Dux,doufutsuofu,dshkwf,du he zuo fu,duhochopeu,Духцoв,دشكوف,ドゥフツェア,杜赫佐夫,후츠포	<a href="#">Czechia</a> , Ústecký kraj Teplice District > Duchcov	populated place population 8,937	N 50° 36' 13"	E 13° 44' 46"
8 <a href="#">Dubí</a> Dubí,Eichwald,doubi,du bi,dubi,Дуби,ドゥビー,杜比,후비	<a href="#">Czechia</a> , Ústecký kraj Teplice District > Dubí	populated place population 7,591	N 50° 41' 8"	E 13° 47' 8"
9 <a href="#">Krupka</a> Graupen,Krupka,ke lu pu ka,Крупка,克鲁普卡	<a href="#">Czechia</a> , Ústecký kraj Teplice District > Krupka	populated place population 13,687	N 50° 41' 4"	E 13° 51' 29"
10 <a href="#">Bílina</a> Bilin,Bilina,bi li na,Билина,比利納	<a href="#">Czechia</a> , Ústecký kraj Teplice District > Bílina	populated place population 15,738	N 50° 32' 54"	E 13° 46' 31"
11 <a href="#">Cieplice</a>	<a href="#">Poland</a> , Warmia-Masuria Powiat elbląski > Elbląg	populated place	N 54° 15' 20"	E 19° 20' 26"

# Jedna nazwa – wiele miejsc

[\[advanced search\]](#)

6 records found for "Wołowiec"

Name	Country	Feature class	Latitude	Longitude
1 <a href="#">Volovec</a> Wołowiec, Wołowiec	SK, PL	mountain elevation 2063m	N 49° 12' 27"	E 19° 45' 46"
2 <a href="#">Wołowiec</a>	<a href="#">Poland</a> , West Pomerania Powiat goleniowski > Gmina Nowogard	populated place	N 53° 46' 1"	E 15° 9' 14"
3 <a href="#">Wołowiec</a>	<a href="#">Poland</a> , Lesser Poland Powiat gorlicki > Gmina Sękowa	populated place	N 49° 31' 10"	E 21° 21' 9"
4 <a href="#">Wołowiec Lake</a>	<a href="#">Canada</a> , Manitoba	lake	N 56° 7' 24"	W 101° 23' 3"
5 <a href="#">Volovets</a> Voloc, Volovec, Volovec', Volovets, Volóc, Wołowiec, Wołowiec, Воловец, Воловець	<a href="#">Ukraine</a> , Transcarpathia Volovets Raion	seat of a second-order administrative division population 5,128, elevation 492m	N 48° 42' 39"	E 23° 11' 6"
6 <a href="#">Wołowiec</a>	<a href="#">Poland</a> , Lower Silesia Wałbrzych County > Jedlina-Zdrój	mountain elevation 776m	N 50° 44' 2"	E 16° 18' 23"

# Jedna nazwa – wiele miejsc

Stanisławów  all countries

Feature Class:

Continent:

fuzzy search :

62 records found for "Stanisławów"

Name	Country	Feature class	Latitude	Longitude
1 <a href="#">Ivano-Frankivsk</a> IFO,Ivano-Frankivsk,Ivano-Frankivsk,Ivano-Frankivsk,Ivano-Frankivsk,Stanislav,Stanislav,Stanislavov...	<a href="#">Ukraine</a> , Ivano-Frankivsk Ivano-Frankivsk Raion	seat of a first-order administrative division population 238,196, elevation 260m	N 48° 55' 23"	E 24° 42' 44"
2 <a href="#">Stanisławów</a> Stanislavuv,Stanislavow,Stanislawów,Станиславув	<a href="#">Poland</a> , Mazovia Mińsk County > Gmina Stanisławów	seat of a third-order administrative division population 2,200	N 52° 17' 21"	E 21° 32' 54"
3 <a href="#">Stanisławów Skrzański</a> Stanislavow,Stanislavow Skrzanski,Stanislawów,Stanislawów Skrzański	<a href="#">Poland</a> , Mazovia Gostynin County > Gostynin	populated place	N 52° 23' 7"	E 19° 29' 36"
4 <a href="#">Stanisławów Studziński</a> Stanislavow,Stanislavow Studzinski,Stanislawów,Stanislawów Studziński	<a href="#">Poland</a> , Łódź Voivodeship Tomaszów County > Czerniewice	populated place	N 51° 40' 28"	E 20° 5' 45"
5 <a href="#">Stanisławów-Kolonia</a> Kolonie Stanislawow,Kolonie Stanislawów	<a href="#">Poland</a> , Mazovia Mińsk County > Gmina Stanisławów	section of populated place	N 52° 17' 18"	E 21° 34' 43"
6 <a href="#">Stanisławów</a> Kolonia Stanislawow,Kolonia Stanislawów,Stanislavow,Stanislawów	<a href="#">Poland</a> , Lublin Radzyń County > Gmina Ulan-Majorat	populated place population 111	N 51° 48' 48"	E 22° 22' 12"
7 <a href="#">Stanisławów Stary</a> Stanislavow Stary,Stanislawów Stary	<a href="#">Poland</a> , Łódź Voivodeship Pabianice County > Lutomiersk	populated place	N 51° 47' 21"	E 19° 12' 14"
8 <a href="#">Stanisławów Nowy</a> Stanislavow Nowy,Stanislawów Nowy	<a href="#">Poland</a> , Łódź Voivodeship Pabianice County > Lutomiersk	populated place	N 51° 48' 5"	E 19° 11' 23"
9 <a href="#">Stanisławów Lipski</a> Stanislavow Lipski,Stanislawów Lipski	<a href="#">Poland</a> , Łódź Voivodeship Tomaszów County > Czerniewice	populated place	N 51° 41' 57"	E 20° 8' 39"
10 <a href="#">Stanisławów</a> Stanislavow,Stanislawów	<a href="#">Poland</a> , Mazovia Kozienice County > Glowaczów	populated place	N 51° 34' 18"	E 21° 23' 32"

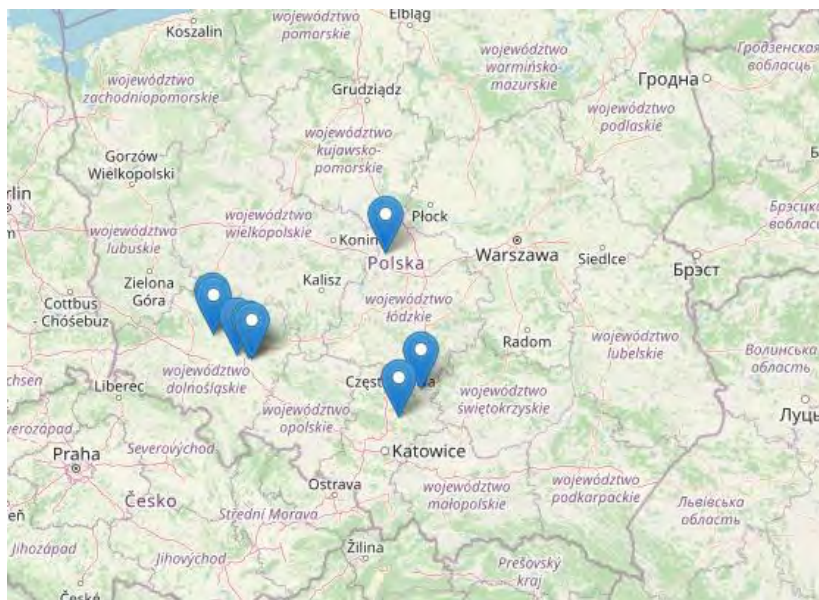
# Jedna nazwa – wiele miejsc

Name	Country	Feature class	Latitude	Longitude
1 <a href="#">Ligota</a> Ligota	<a href="#">Poland</a> , Silesia Bielsko-Biala County > Czechowice-Dziedzice	populated place population 4,454	N 49° 53' 54"	E 18° 57' 3"
2 <a href="#">Ligota Woźnicka</a> Ligota Woźnicka,Ligota Woźnicka	<a href="#">Poland</a> , Silesia Lubliniec County > Woźniki	populated place	N 50° 36' 39"	E 19° 2' 36"
3 <a href="#">Ligota Tułowicka</a> Ligota Tułowicka,Ligota Tułowicka	<a href="#">Poland</a> , Opole Voivodeship Opole County > Tułowice	populated place	N 50° 34' 54"	E 17° 39' 11"
4 <a href="#">Ligota Rybińska</a> Ligota Rybińska,Ligota Rybińska	<a href="#">Poland</a> , Greater Poland	populated place	N 51° 22' 25"	E 17° 42' 33"
5 <a href="#">Ligota Prószkowska</a> Ligota Prószkowska,Ligota Prószkowska,Ligota Prószkowska,Ligota Prószkowska	<a href="#">Poland</a> , Opole Voivodeship Opole County > Gmina Prószków	populated place	N 50° 33' 15"	E 17° 49' 53"
6 <a href="#">Ligota Zamecka</a> Ligota Zamecka	<a href="#">Poland</a> , Opole Voivodeship Kluczbork County > Gmina Kluczbork	populated place	N 50° 58' 28"	E 18° 14' 24"
7 <a href="#">Ligota Turawska</a> Ellguth,Ellguth-Turawa,Ligota Turawska	<a href="#">Poland</a> , Opole Voivodeship Opole County > Gmina Turawa	populated place	N 50° 47' 12"	E 18° 10' 57"
8 <a href="#">Ligota Oleska</a> Ligota Oleska	<a href="#">Poland</a> , Opole Voivodeship Olesno County > Radłów	populated place	N 50° 57' 29"	E 18° 32' 6"
9 <a href="#">Ligota Dolna</a> Ligota Dolna	<a href="#">Poland</a> , Opole Voivodeship Kluczbork County > Gmina Kluczbork	populated place	N 50° 58' 21"	E 18° 10' 30"
10 <a href="#">Ligota Białska</a> Ligota Białska	<a href="#">Poland</a> , Opole Voivodeship Prudnik County > Gmina Biała	populated place	N 50° 24' 39"	E 17° 40' 48"
11 <a href="#">Ligota</a> Ligota	<a href="#">Poland</a> , Opole Voivodeship Krapkowice County > Gmina Krapkowice	populated place	N 50° 26' 47"	E 17° 59' 29"
12 <a href="#">Ligota Mała</a> Ligota Mała,Ligota Mała	<a href="#">Poland</a> , Lower Silesia Dzierżoniów County > Gmina Niemcza	populated place	N 50° 40' 21"	E 16° 48' 47"
13 <a href="#">Ligota Książęca</a>	<a href="#">Poland</a> , Opole Voivodeship Namysłów County > Namysłów	populated place	N 51° 1' 11"	E 17° 34' 11"
14 <a href="#">Ligota Dobrodzieńska</a>	<a href="#">Poland</a> , Opole Voivodeship Olesno County > Gmina Dobrodzień	populated place	N 50° 43' 58"	E 18° 24' 32"
15 <a href="#">Ligota</a>	<a href="#">Poland</a> , Łódź Voivodeship Łask County > Widawa	populated place	N 51° 28' 53"	E 18° 55' 54"
16 <a href="#">Ligota Wielka</a>	<a href="#">Poland</a> , Opole Voivodeship Nysa County > Gmina Otmuchów	populated place	N 50° 29' 11"	E 17° 6' 37"
17 <a href="#">Ligota Górna</a>	<a href="#">Poland</a> , Opole Voivodeship Kluczbork County > Gmina Kluczbork	populated place	N 50° 58' 7"	E 18° 15' 42"
18 <a href="#">Ligota Czamborowa</a>	<a href="#">Poland</a> , Opole Voivodeship Strzelce County > Izbicko	populated place	N 50° 34' 20"	E 18° 10' 52"
19 <a href="#">Ligota Wielka</a> Ligota Wielka	<a href="#">Poland</a> , Lower Silesia Dzierżoniów County > Łagiewniki	populated place	N 50° 45' 31"	E 16° 47' 29"
20 <a href="#">Ligota Dolna</a> Ligota Dolna,Nieder Ellguth	<a href="#">Poland</a> , Opole Voivodeship Strzelce County > Gmina Strzelce Opolskie	populated place	N 50° 29' 27"	E 18° 7' 14"
21 <a href="#">Ligota</a> Ligota	<a href="#">Poland</a> , Greater Poland Ostrzeszów County > Kobyla Góra	populated place	N 51° 22' 17"	E 17° 48' 18"
22 <a href="#">Ligota</a>	<a href="#">Poland</a> , Greater Poland Ostrów County > Gmina Raszków	populated place	N 51° 47' 19"	E 17° 41' 48"
23 <a href="#">Ligota Tworkowska</a> Ligota Tworkowska,Tworkowska	<a href="#">Poland</a> , Silesia Wodzisław County > Gmina Lubomia	populated place	N 50° 1' 26"	E 18° 15' 57"
24 <a href="#">Ligota Polska</a> Alt Ellguth,Ligota Polska	<a href="#">Poland</a> , Lower Silesia Oleśnica County > Oleśnica	populated place	N 51° 14' 48"	E 17° 32' 38"
25 <a href="#">Ligota</a> Ligota	<a href="#">Poland</a> , Lower Silesia Góra County > Góra	populated place	N 51° 42' 1"	E 16° 36' 30"
26 <a href="#">Ligota-Panewniki</a>	<a href="#">Poland</a> , Silesia Katowice > Katowice	section of populated place	N 50° 13' 49"	E 18° 58' 18"
27 <a href="#">Ligota Mała</a>	<a href="#">Poland</a> , Lower Silesia Oleśnica County > Oleśnica	populated place	N 51° 8' 11"	E 17° 21' 45"

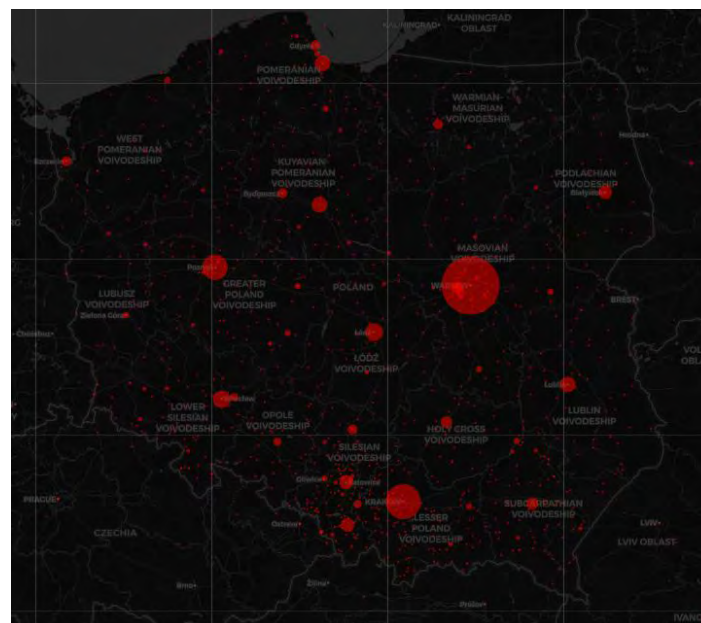


# Wizualizacja

Tak można pokazać 5 punktów.

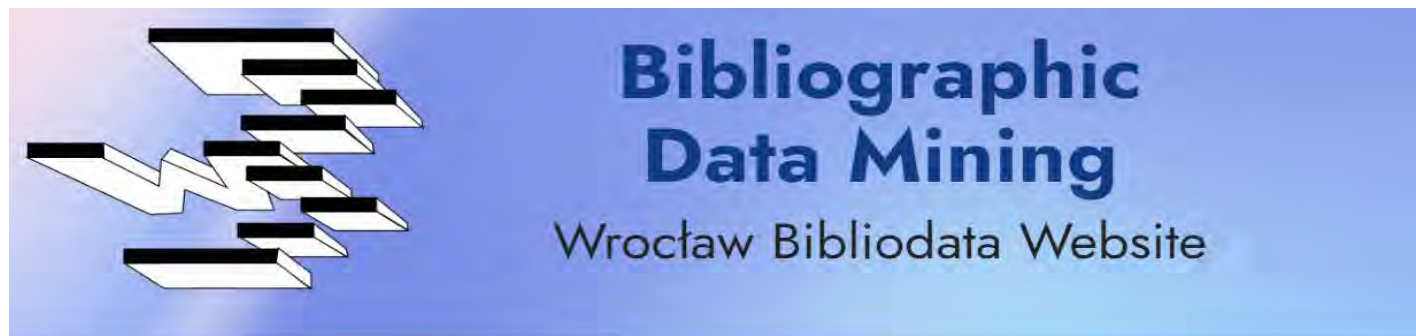


A tak 50.000.



# Gdzie nas znaleźć?

<http://phc.uni.wroc.pl/wbw/>



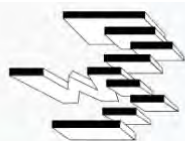
Wielkie bibliografie powstawały przez dziesięciolecia wszędzie tam, gdzie narody lub wspólnoty komunikacji wytworzyły znaczący zasób piśmiennictwa. Nie są one jedynie prostymi rejestrami publikacji, służącymi doraźnym potrzebom informacyjnym społeczeństwa. Efekt skali i zawarte w rekordach (meta)dane sprawiają, że wielkie bibliografie w syntetyczny sposób odzwierciedlają trendy kultury, polityki i cywilizacji. Rozwój technologii cyfrowych pozwolił na wydobywanie i wizualizację tej informacji, a Internet stał się doskonałą platformą jej upowszechniania.

**Wrocław Bibliodata Website** prezentuje w postaci interaktywnej aplikacji efekty rzutowania toponimów (nazw miejscowych) zawartych w danych bibliograficznych na mapy, a także wizualizacje innych informacji, wydobywanych z wielkich bibliografii metodami NLP i text mining. Uwzględniając wymogi metod big data, a także powszechnie stosowane periodyzacje historyczne, jako wzorcowy przyjęto okres tzw. późnej nowożytności i współczesności, którego umownym początkiem jest rok 1800.

Zapleczem naukowym **Wrocław Bibliodata Website** jest Instytut Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Wrocławskiego, natomiast wsparcie technologiczne zapewnia zespół konsorcjum CLARIN-PL i Politechnika Wrocławska. Pomysłodawcami i wykonawcami projektu są prof. dr hab. Adam Pawłowski (UWr) oraz dr inż. Tomasz Walkowiak (PWr).

# Gdzie nas znaleźć?

<https://bibgeos.clarin-pl.eu/>



## Bibliographies on maps



BN

Topnoms extracted from publication places, titles and publication description of books in Polish (1801-2019) from [National Library of Poland](#)



PBL

Topnoms extracted from publication places of books from [Polish Literary Bibliography](#)



Directions

Historical directions of interest of Polish culture based on the occurrences of toponyms in book titles in [National Library of Poland](#)



Tomasz Walkowiak



Adam Pawłowski



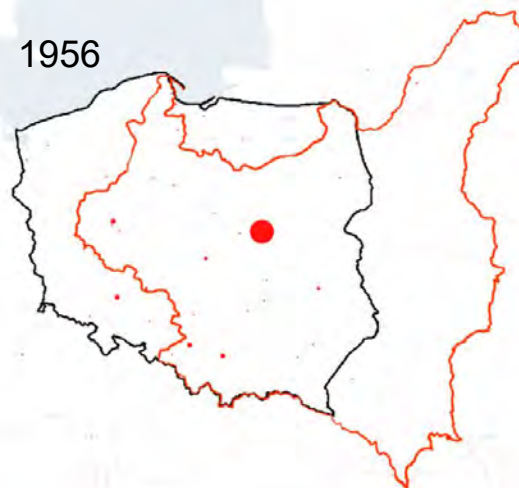
Publications

# Bibliograficzny kontur granic

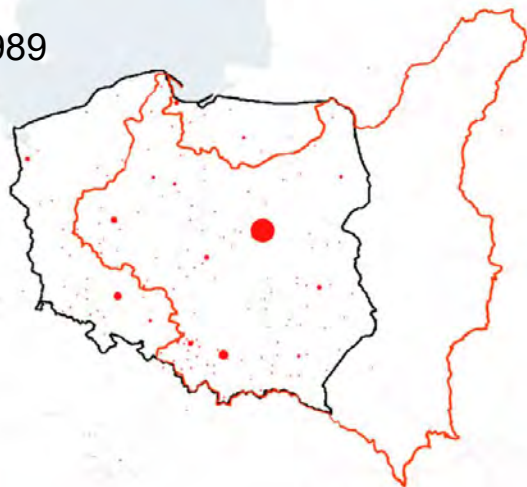
1938



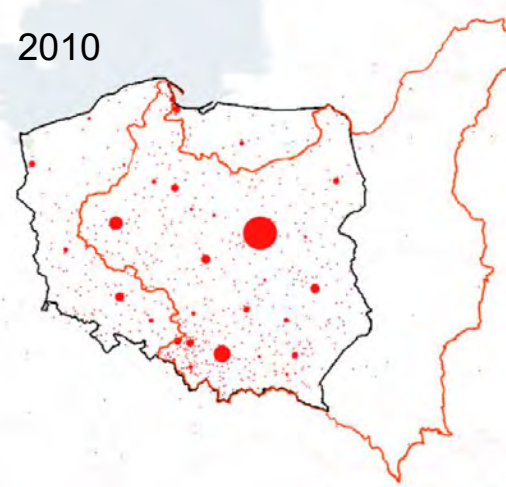
1956



1989

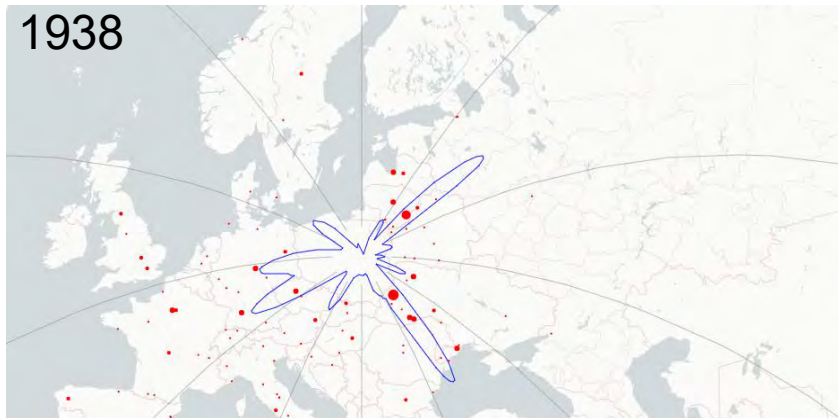


2010

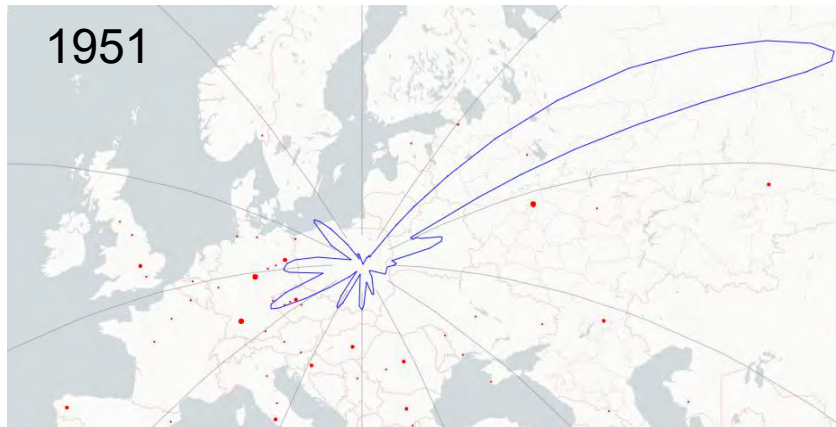


# Kierunki zainteresowań i oddziaływania

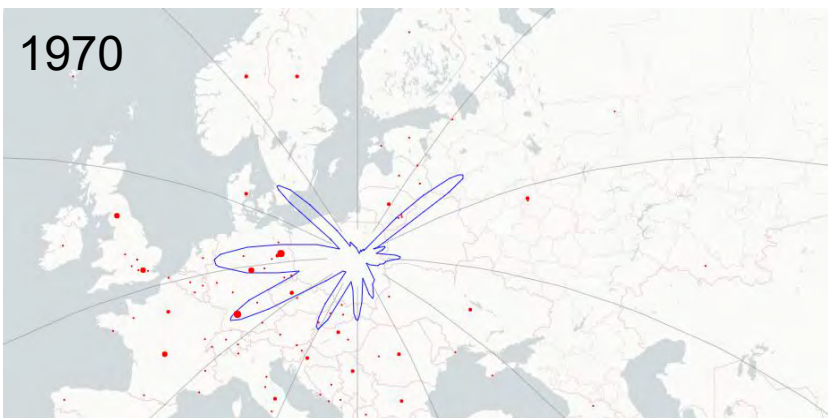
1938



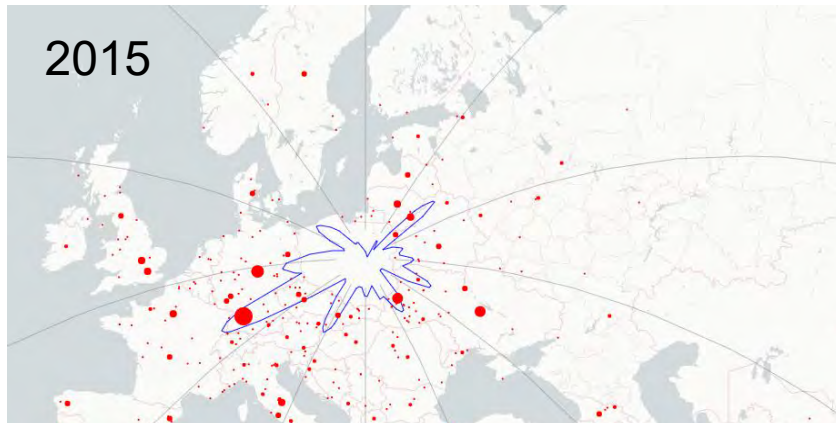
1951



1970



2015



# Dlaczego linie się zaginają?



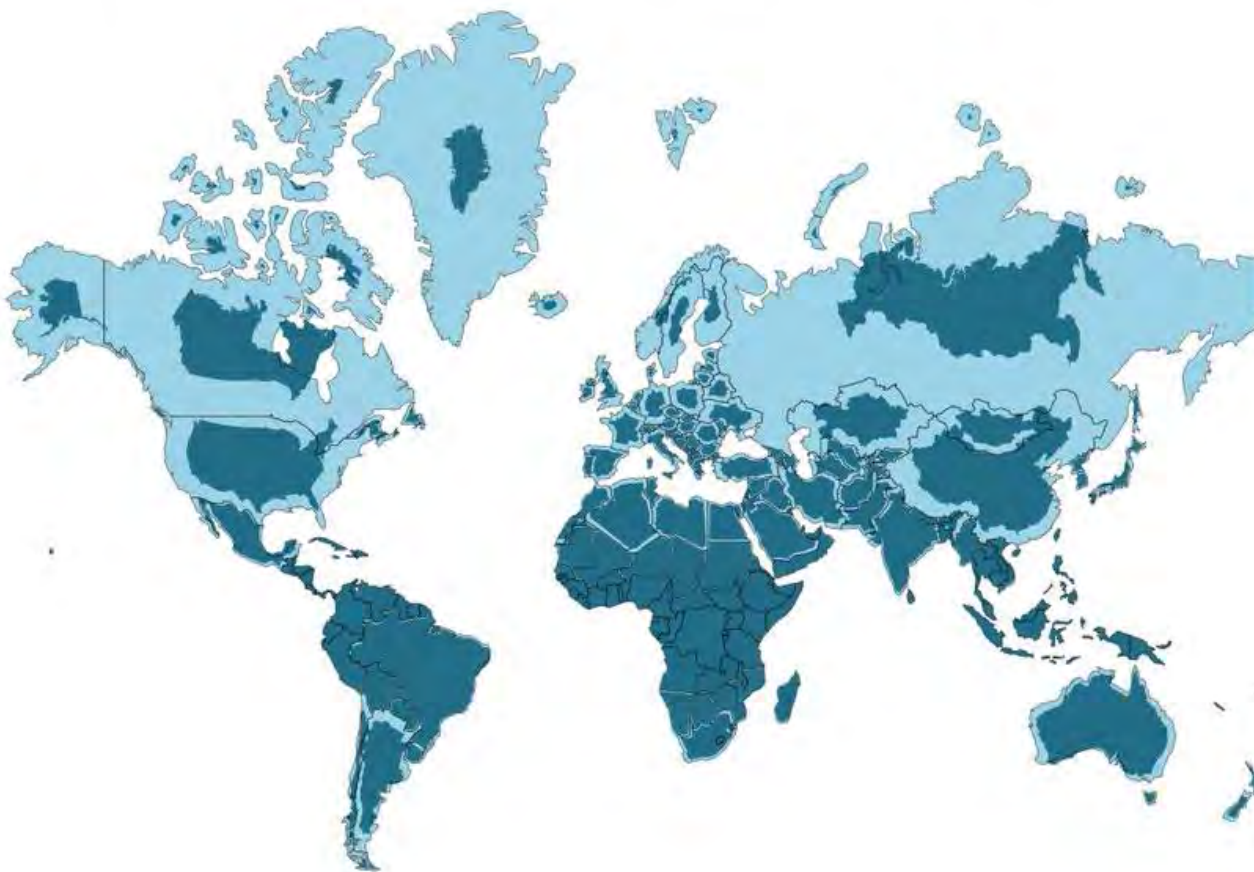
## Gerard Mercator (1512 – 1594)

XVI-wieczny geograf, kosmograf i kartograf pochodzącym z Flandrii. Największą sławę przyniosło mu stworzenie w 1569 r. mapy świata, opartej na nowym rzutowaniu, które przedstawiało trasy żeglarskie o stałym kierunku (tzw. loksodromy) jako linie proste. innowacja ta jest nadal stosowana w mapach morskich.

Mercator był pionierem kartografii. Obok Gemmy Frisiusa i Abrahama Orteliusa jest powszechnie uważany za jednego z założycieli niderlandzkiej szkoły kartografii i geografii. Jest też powszechnie uważany za najbardziej znaczącą postać tej szkoły. W swoich czasach był znanym twórcą globusów i instrumentów naukowych.

# Czy ziemia na pewno jest okrągła?

World Mercator projection with true country size and shape added



# Wyniki cz. 2

Płeć autorów a liczba i tematy publikacji (Polska vs Niemcy)



# Przesłanki ogólne

Napisanie i wydanie książki jest w sensie statystyki wskaźnikiem agregującym inne wskaźniki szczegółowe.

- wykształcenie → średnia wykształcenia, poziom edukacji
- poziom zamożności → średni dochód
- prestiż społeczny
- akceptacja społeczna

Założenie, że liczba publikowanych książek przez daną grupę społeczną jest przybliżeniem miary emancypacji, jest więc słuszne.

Dodatkowo: pewne miary ekonomiczne są niemożliwe do obliczenia w Europie Środkowej w czasie objętym badaniem (od 1800).

# Płeć autorów PL/DE: cel i hipotezy

## Cele szczegółowe:

- ukazać proporcje kobiet i mężczyzn wśród autorów książek wydanych w latach 1801-2021 w dwóch europejskich krajach (wspólnotach komunikacji): Polsce i Niemczech
- wydobyć z tytułów książek tematy (topiki) charakterystyczne dla autorów
- zidentyfikować treści, które odróżniają piśmiennictwo obu płci

## Podejście i metodyka:

Zastosowano techniki NLP, statystykę matematyczną i programowanie baz danych do wielkich baz bibliograficznych.

Umożliwiło to przeprowadzenie badania w perspektywie tzw. *długiego trwania* (w rozumieniu metodyki badań szkoły Annales).

# Płeć autorów PL/DE: dane

Dane: katalog Biblioteki Narodowej i Deutsche Nationalbibliothek

Format: MARC 21

Dodatkowo: Polska wiek XIX – Bibliografia Estreicherów (do litery k)

Filtrowanie danych z obu baz:

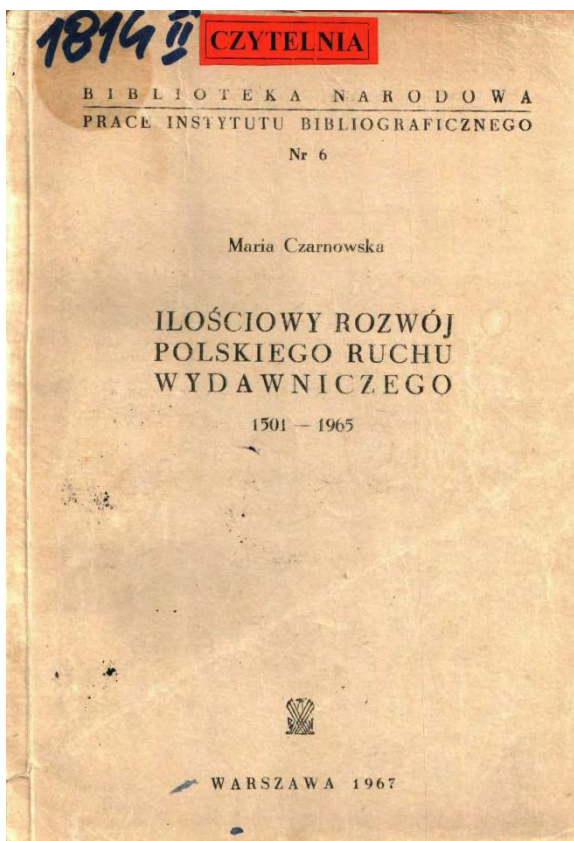
- uwzględniono język PL i DE
- uwzględniono niepuste pola 'tytuł', 'autor', 'data wydania'
- pominięto rekordy nietekstowe (mapy, nuty, nagrania itd.) i periodyki

Czyszczenie danych:

- automatyczne oddzielenie imienia od nazwiska
- ujednoczenie kodowania umlautów (mogą być dwuznakami)
- usunięcie dubletów z Bibliografii Estreicherów

**Objętość końcowa: PL = ok. 1,4 mln rek., DE = ok. 7,5 mln rek.**

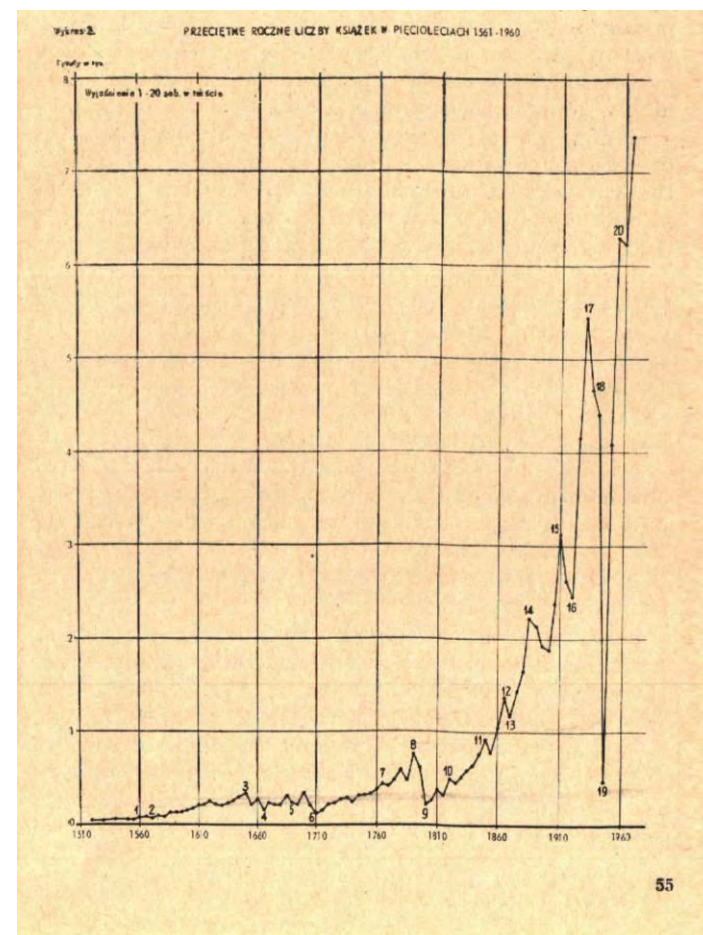
# Chwała pionierom: Maria Czarnowska



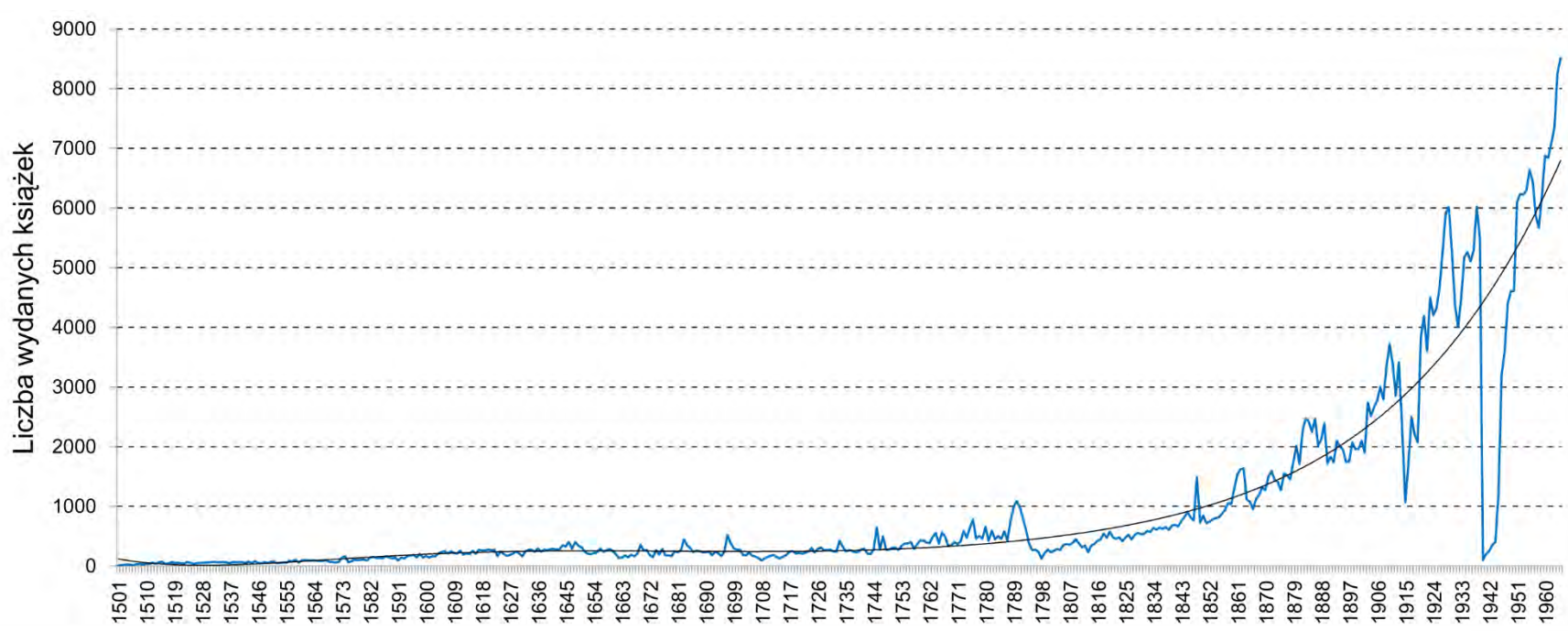
**ANEKSY**  
I. KSIĄŻKI W LATACH 1501-1965

Rok	Tytły	Rok	Tytły	Rok	Tytły	Rok	Tytły
1501	9	1531	64	1561	96	1591	95
2	14	2	69	2	92	2	142
3	23	3	61	3	90	3	127
4	30	4	64	4	99	4	166
5	18	5	71	5	87	5	174
6	13	6	59	6	87	6	184
7	35	7	44	7	85	7	147
8	34	8	69	8	97	8	205
9	24	9	60	9	65	9	141
1510	37	1540	64	1570	71	1600	167
1511	35	1541	65	1571	58	1601	142
2	56	2	46	2	70	2	162
3	38	3	78	3	143	3	162
4	60	4	50	4	162	4	214
5	74	5	80	5	66	5	235
6	37	6	50	6	76	6	249
7	41	7	44	7	104	7	210
8	64	8	66	8	99	8	247
9	55	9	49	9	104	9	217
1520	55	1550	70	1580	111	1610	208
1521	51	1551	76	1581	94	1611	259
2	36	2	47	2	144	2	191
3	64	3	53	3	148	3	210
4	59	4	65	4	134	4	203
5	33	5	53	5	138	5	253
6	47	6	84	6	128	6	204
7	52	7	72	7	162	7	269
8	50	8	105	8	147	8	260
9	60	9	58	9	125	9	261
1530	58	1560	92	1590	142	1620	273

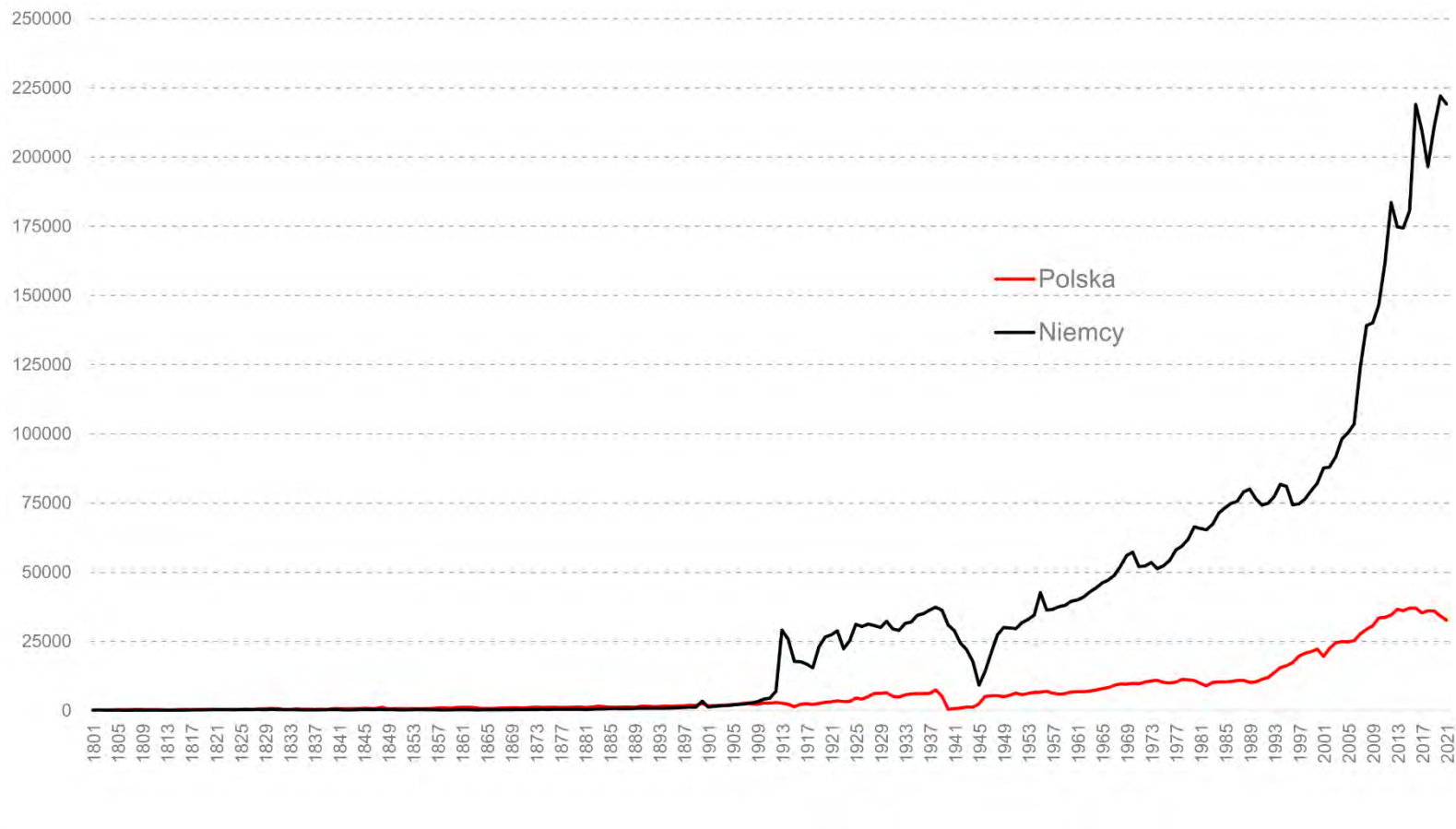
170 → = 3188  
= 3124  
= 500



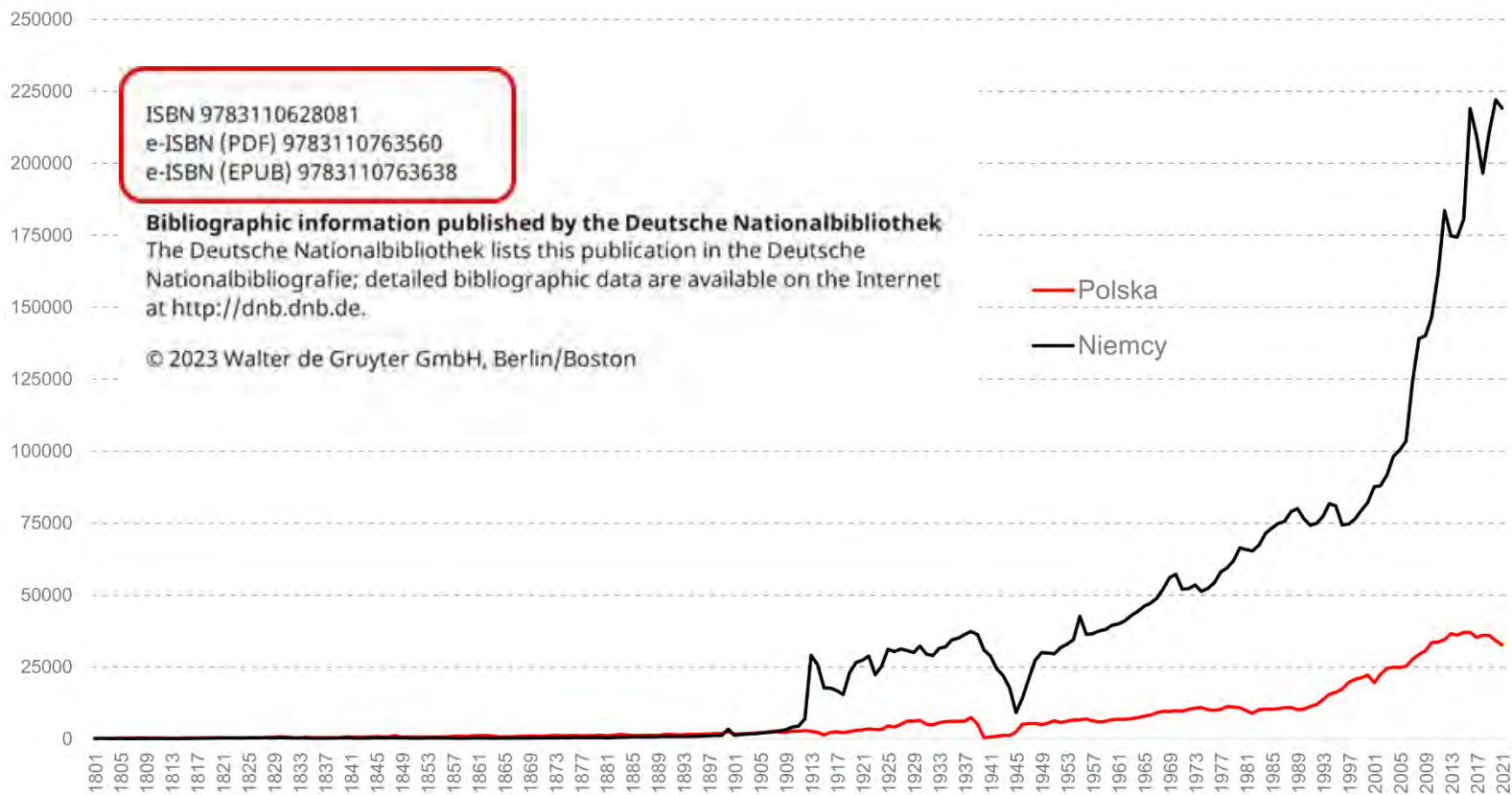
# Chwała pionierom: Maria Czarnowska



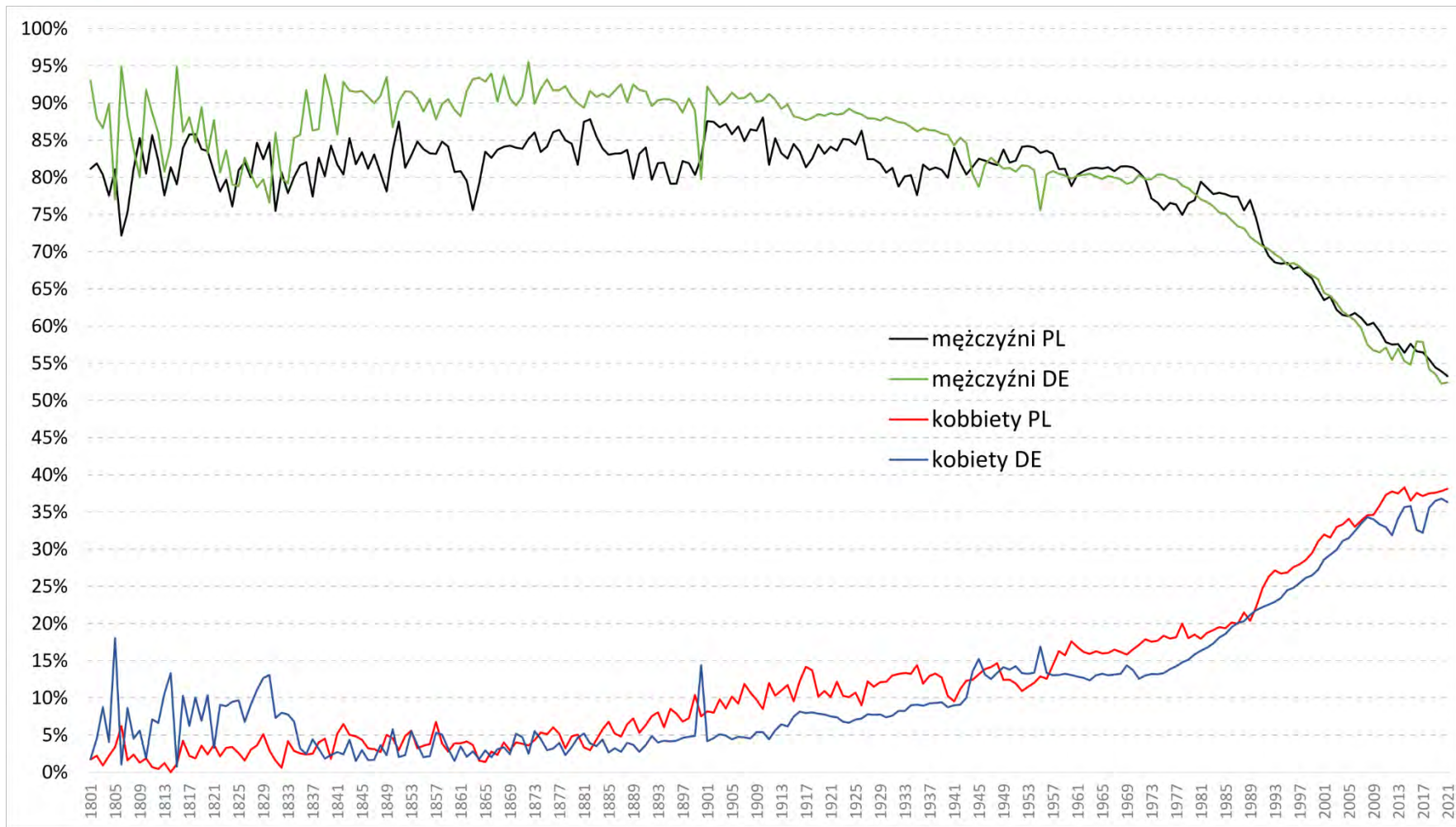
# Liczby publikacji w bazach PL i DE



# Liczby publikacji w bazach PL i DE

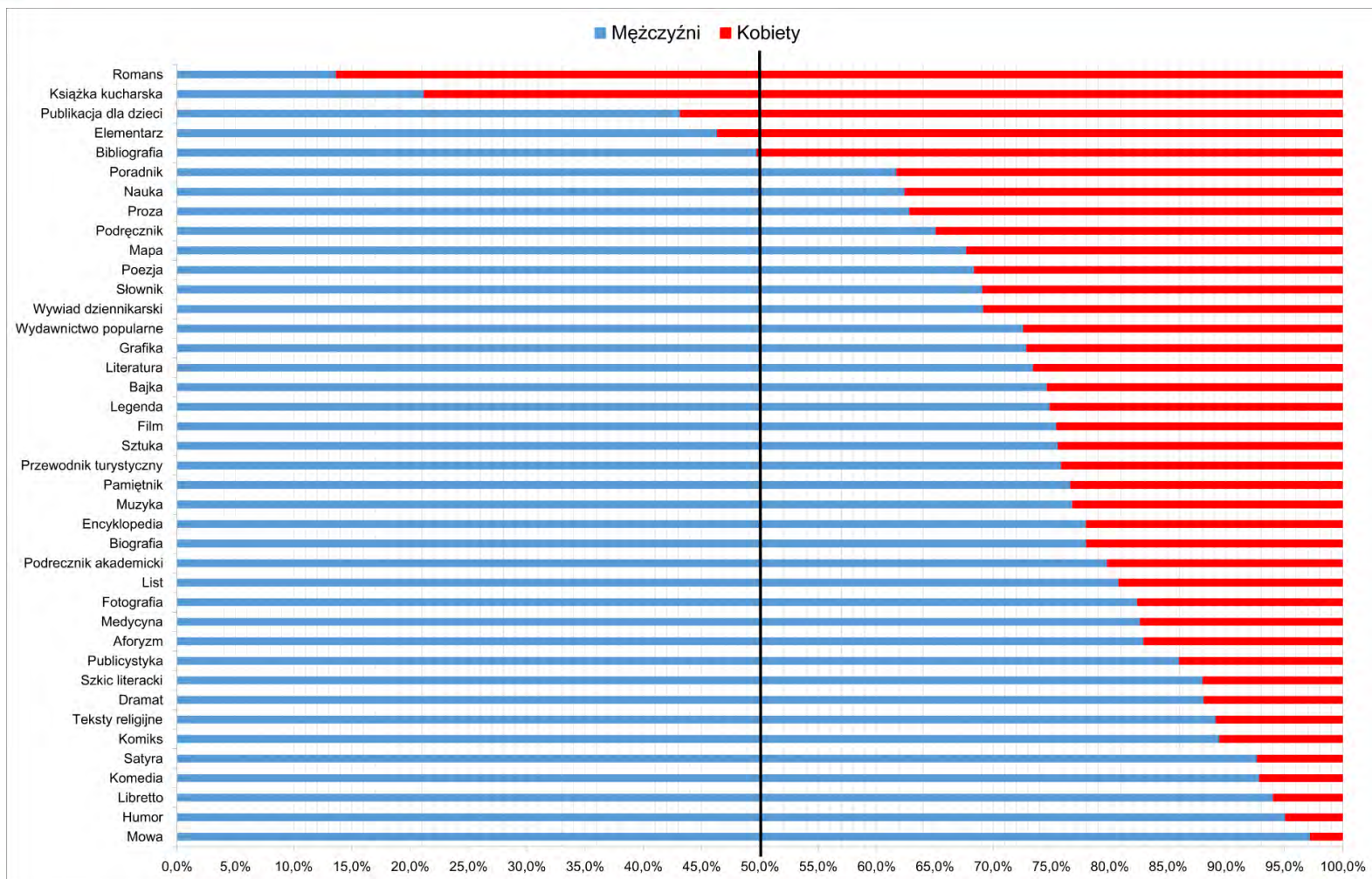


# Różne państwa i języki – jeden proces





# Gatunki piśmiennicze (Polska)



# Tematy różnicujące (Polska)

Okres	Topiki kobiet (C)	Topiki mężczyzn (C)
1801-1910	KOBIETA (5,7) MŁODZIEŻ (4,8) POWIEŚĆ (2,8)	OBRONA DOKTORATU (8,8) SPRAWOZDANIE (5,6) TEKST AKADEMICKI (4,3)
1911-1945	MŁODZIEŻ (4,9) ROMANS/POWIEŚĆ (3,1) POWIEŚĆ (2,1)	PRAWO (3,9) WYKŁAD/PRZEMÓWIENIE (3,1) TEKST AKADEMICKI (2,2)
1946-1980	DZIECKO (4,7) JĘZYK (2,6) SZKOŁA (2,6)	ELEKTRYCZNOŚĆ (4,7) KONSTRUKCJA/BUDOWA (4,1) TRANSPORT (4,0)
1981-1999	SZKOŁA (4,2) DZIECI (3,8) ROMANS/POWIEŚĆ (3,3)	TRANSPORT (4,8) ELEKTRYCZNOŚĆ (4,4) KONSTRUKCJA/BUDOWA (3,8)
2000-2021	SZKOŁA (6,3) DZIECI (3,6) KOBIETA (3,6)	MASZYNA (4,8) WOJNA (4,7) INFORMATYKA (3,9)

# Tematy różnicujące (Niemcy)

Okres	Topiki kobiet (C)	Topiki mężczyzn (C)
1801-1910	POWIEŚĆ (4,6) ROMANS (2,5) LISTY (1,9)	NIEMCY (3,3) PRZEMÓWIENIE/SPRAWOZDANIE (3,1) RELIGIA (3,0)
1911-1945	POWIEŚĆ (3,5) OPOWIEŚĆ (3,3) DZIECI (3,2)	PODATKI (5,7) EKONOMIA (3,0) PRAWO (2,9)
1946-1980	ROMANS/POWIEŚĆ (3,3) KOBIETA (3,1) DZIECI (2,8)	NAUKA (3,6) PRAWO (3,2) ELEKTRYCZNOŚĆ (2,8)
1981-1999	DZIECI (2,8) ROMANS/POWIEŚĆ (2,3) MEDYCYNĄ (2,0)	INFORMATYKA (3,2) MATEMATYKA (2,8) NAUKI STOSOWANE (2,8)
2000-2021	ROMANS/POWIEŚĆ (3,1) DZIECI (2,8) GOTOWANIE (2,1)	BADANIA NAUKOWE (2,3) FINANSE (2,1) RELIGIA (1,8)

# Rozkład płci autorów w artykułach naukowych (PL)

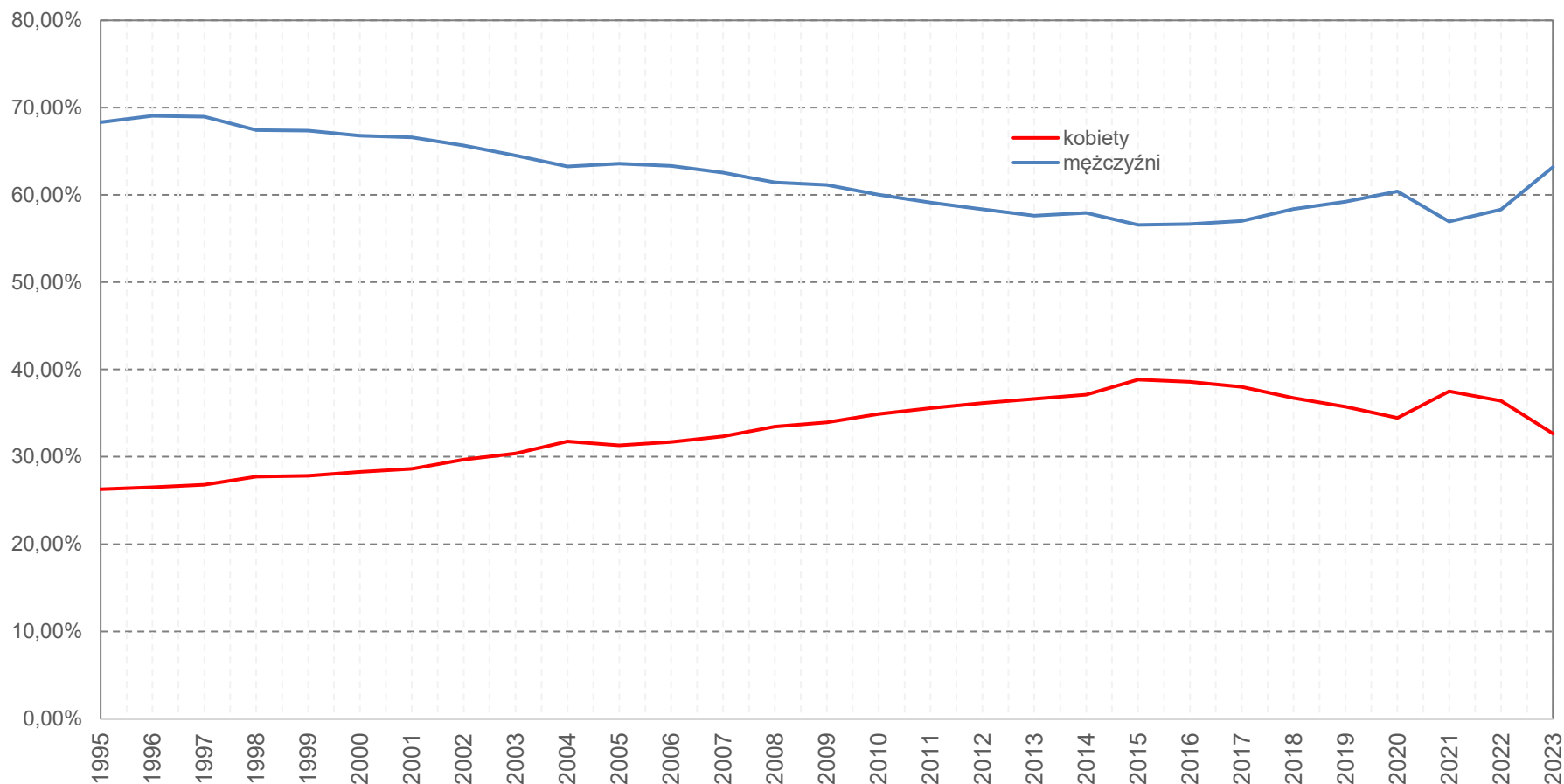
Źródło: dane Biblioteki Narodowej

Objętość:

<b>kobiety</b>	<b>944.478</b>	<b>33,98%</b>
<b>mężczyźni</b>	<b>1.695.587</b>	<b>61,01%</b>
<b>nierozpoznane</b>	<b>139.241</b>	<b>5,01%</b>
<b>ogółem</b>	<b>2.779.306</b>	

# Rozkład płci autorów w art. naukowych (PL)

Procentowy rozkład płci autorów artykułów (Baza BN od 1995)



# Wyniki cz. 3

## Dynamika publikowania w miastach Polski

# Miasta polskie a publikacje

- 1) Miasta duże i średnie z Warszawą (liczby tyt. / proporcje)
- 2) Miasta duże i średnie bez W-wy (liczby tyt. / proporcje)
- 3) Wrocław i Poznań (liczby tyt. / proporcje)
- 4) Warszawa i Kraków (liczby tyt. / proporcje)
- 5) Warszawa, Kraków, Lwów 1801-1939 (proporcje)
- 6) Wschód i zachód Polski (proporcje)

# Skąd te dane?

Bibliography National Library of Poland Toponyms from publication places

List of books with Poznań in 2000

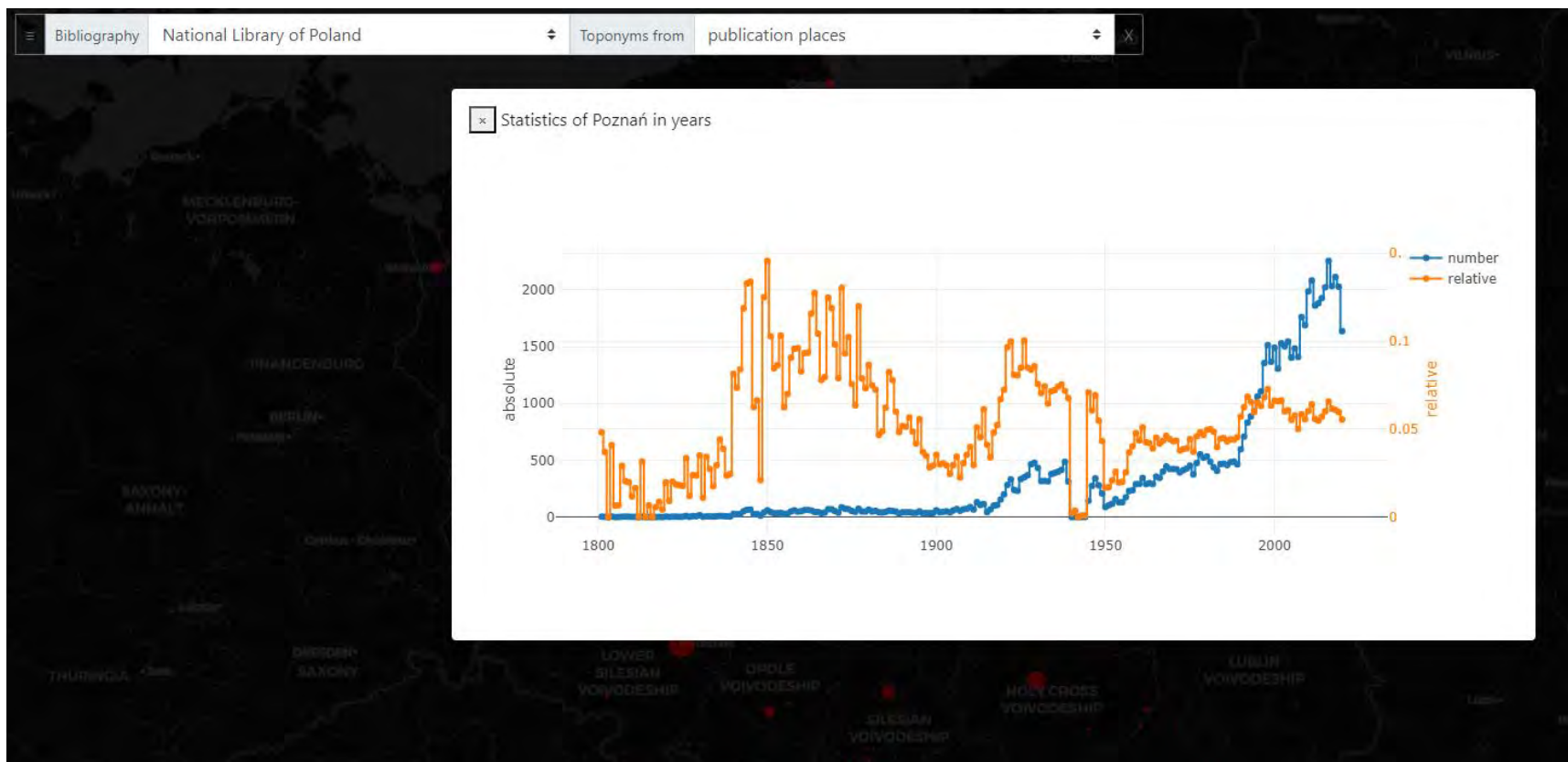
Authors	Title	Publisher
Bogdan Hojdis	O współistnieniu słów i obrazów w kulturze polskiego średniowiecza	"Tum"
	Immunologia porównawcza i rozwojowa zwierząt	Wydaw. Naukowe PWN
Grzegorz Dziamski	Lata dziewięćdziesiąte	Galeria Miejska "Arsenał"
	Postępy psychoterapii	PW [Przedsiębiorstwo Wielobranżowe] "Interfund"
Gerard Górnicki	Bitwa szalała do wieczora...	PW [Przedsiębiorstwo Wielobranżowe] "Interfund"
Kazimierz Rogoziński	Usługi rynkowe	Wydaw. AE
Danuta Krzemińska	Finanse przedsiębiorstwa	Wydaw. Wyższej Szkoły Bankowej
Marek Hendrykowski	Film jako źródło historyczne	"Ars Nova"
Czesław Osękowski	Wybory do sejmu z 19 stycznia 1947 roku w Polsce	Wydaw. Poznańskie
	Nazizm, Trzecia Rzesza a procesy modernizacji	Wydaw. Poznańskie

Showing 1 to 10 of 1491 rows 10 rows per page

< 1 2 3 4 5 ... 150 >

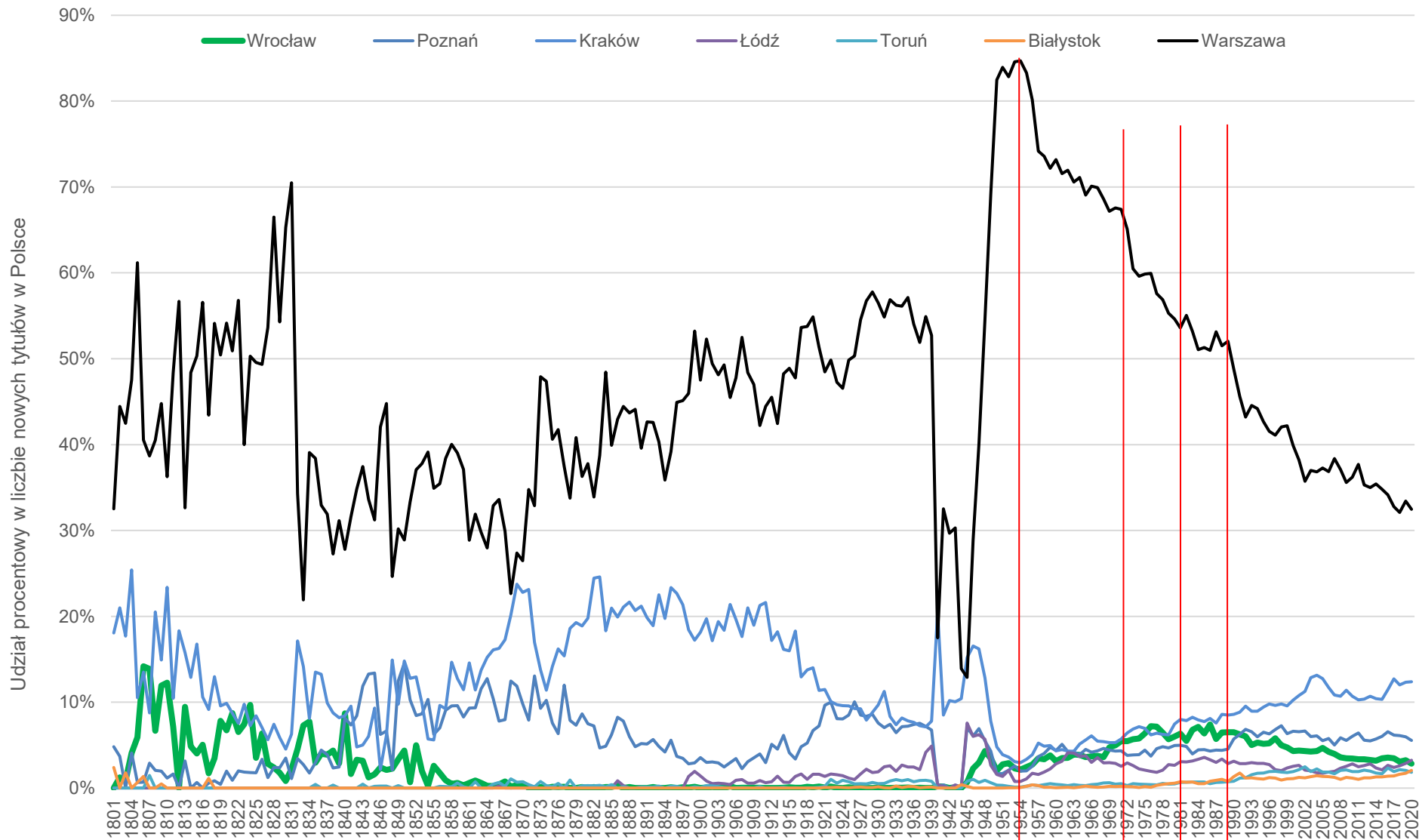


# Skąd te dane?



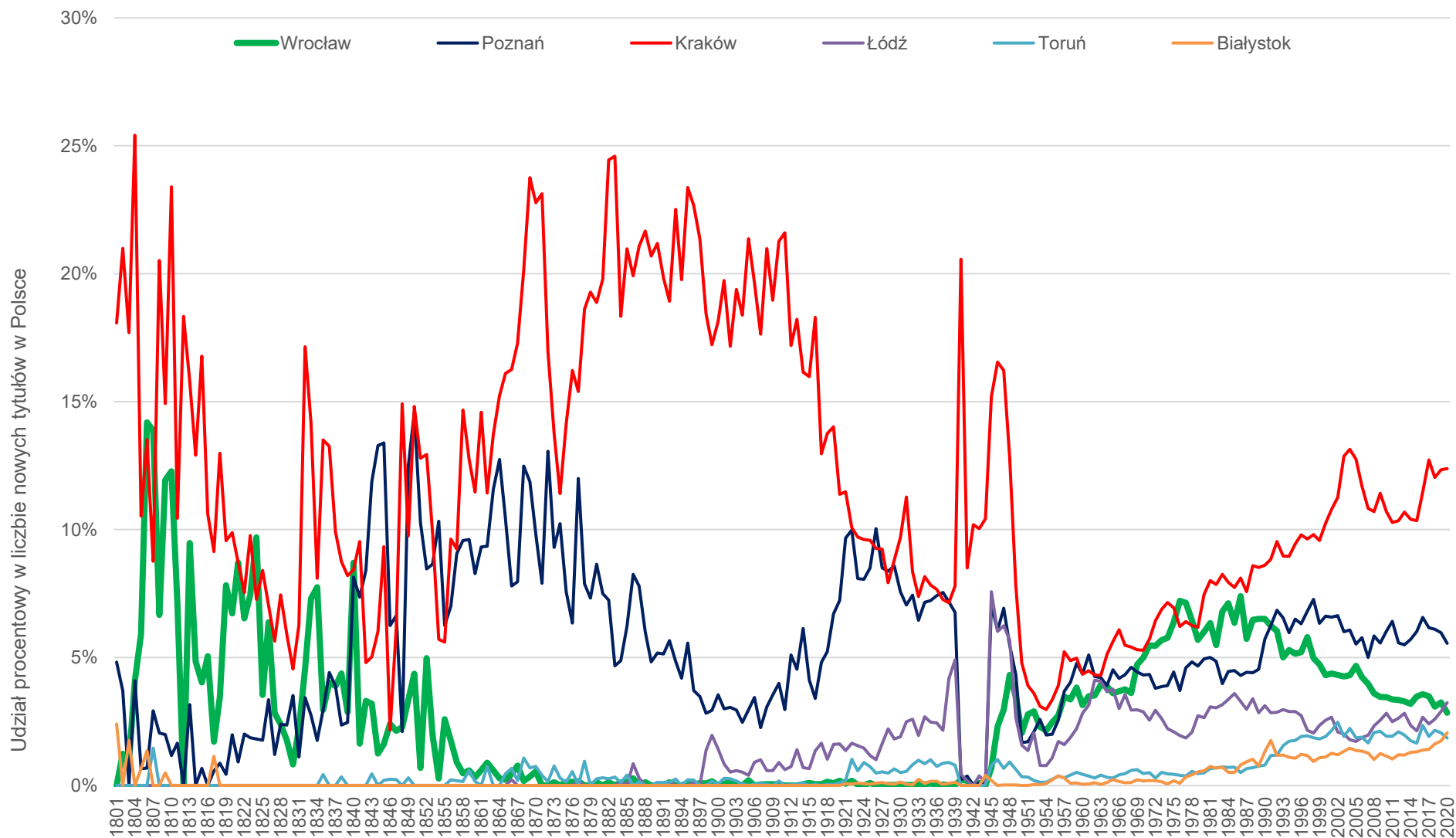


# Miasta duże i średnie z Warszawą (proporcje)

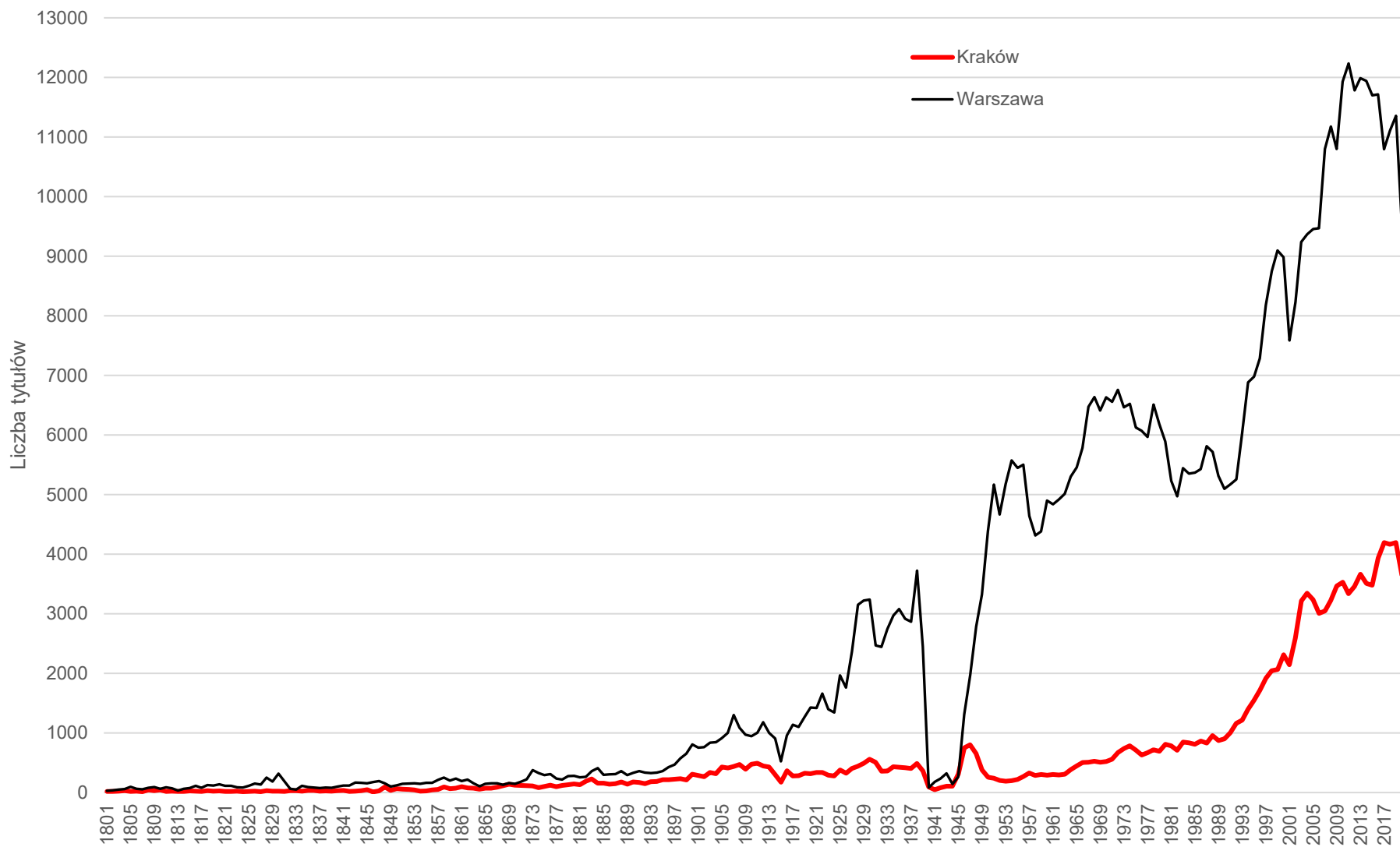




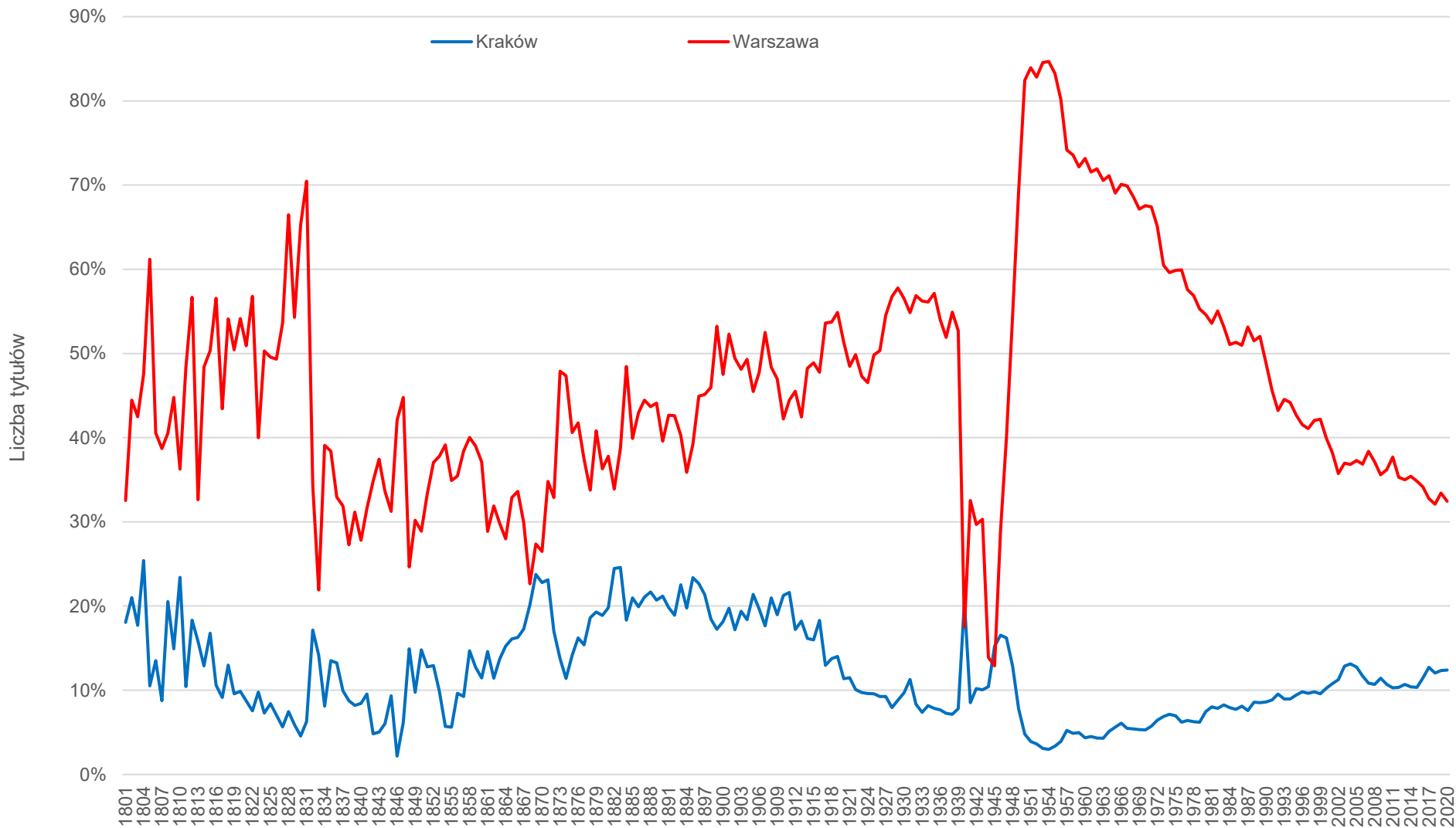
# Miasta duże i średnie bez Warszawy (proporcje)



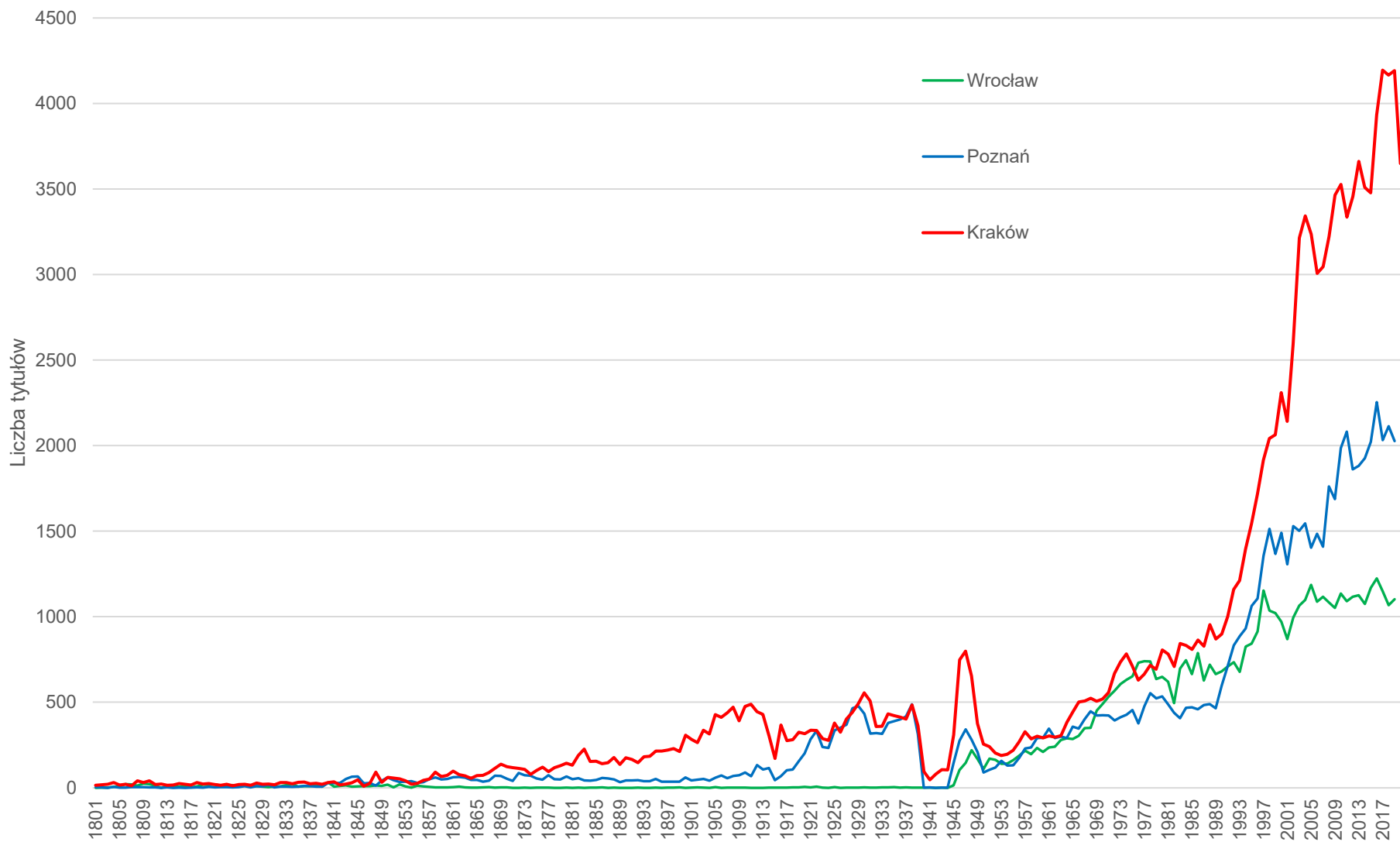
# Warszawa i Kraków (liczby tyt.)



# Warszawa i Kraków (proporcje)

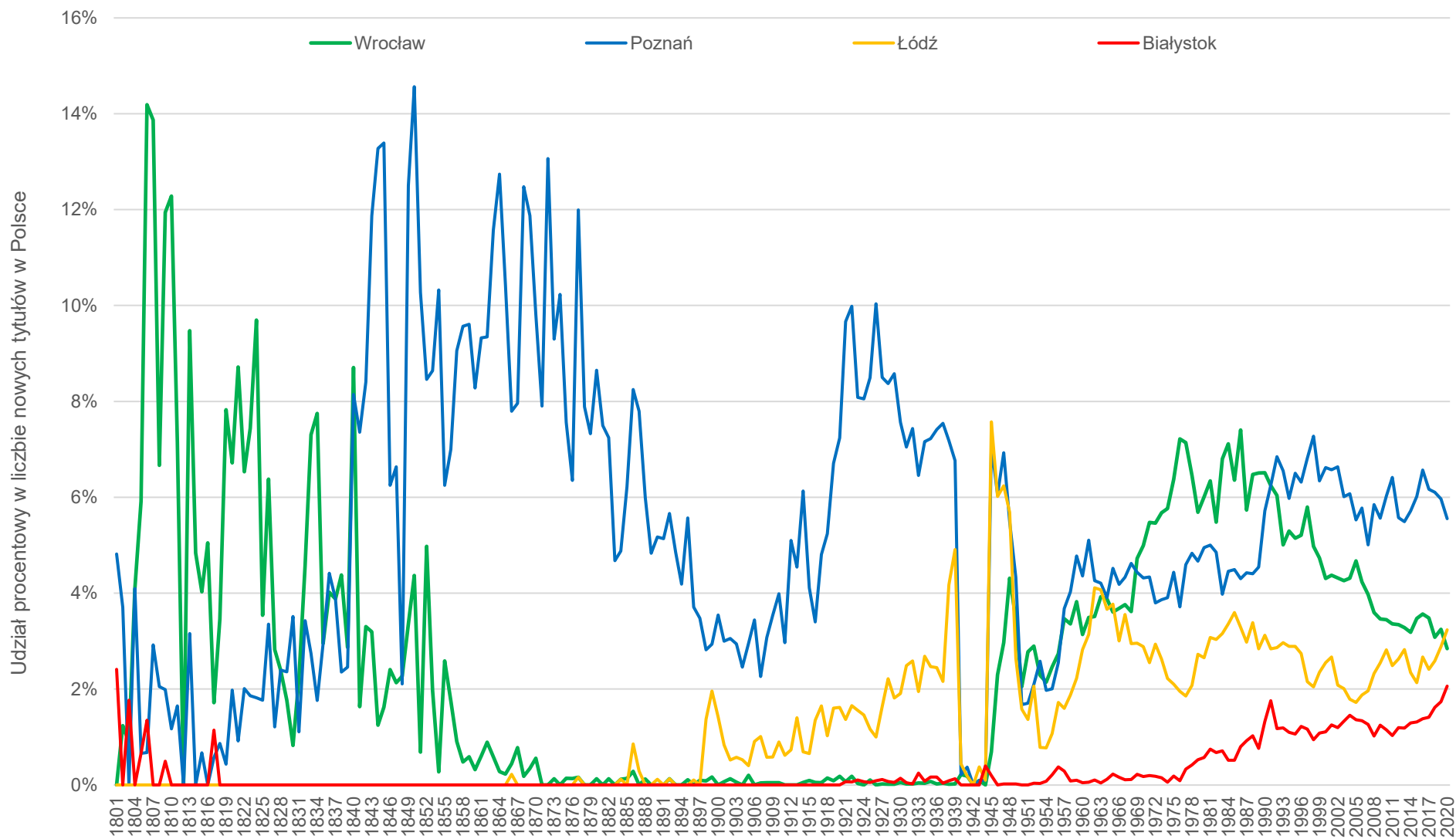


# Kraków, Poznań, Wrocław (liczby tyt.)

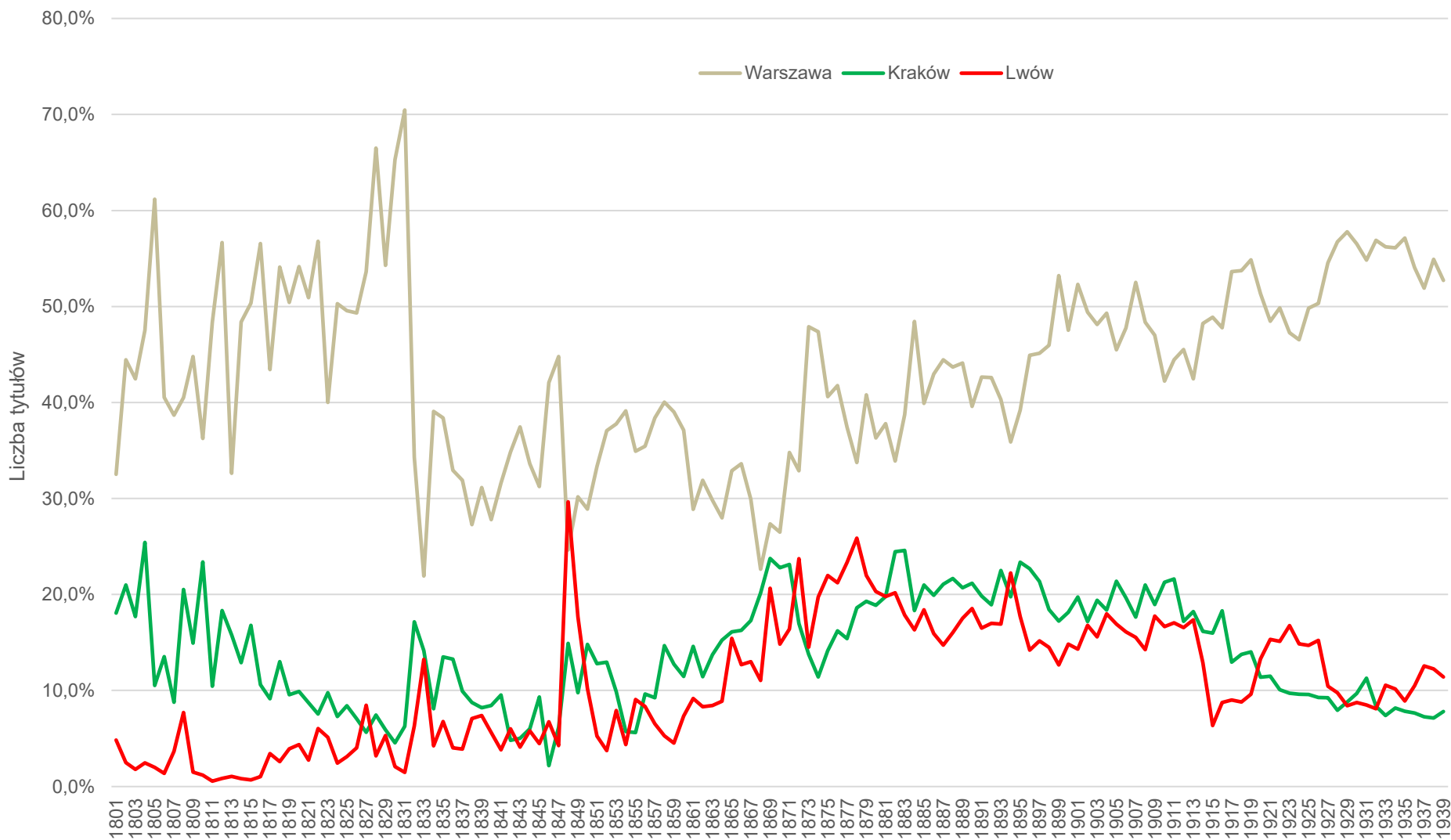




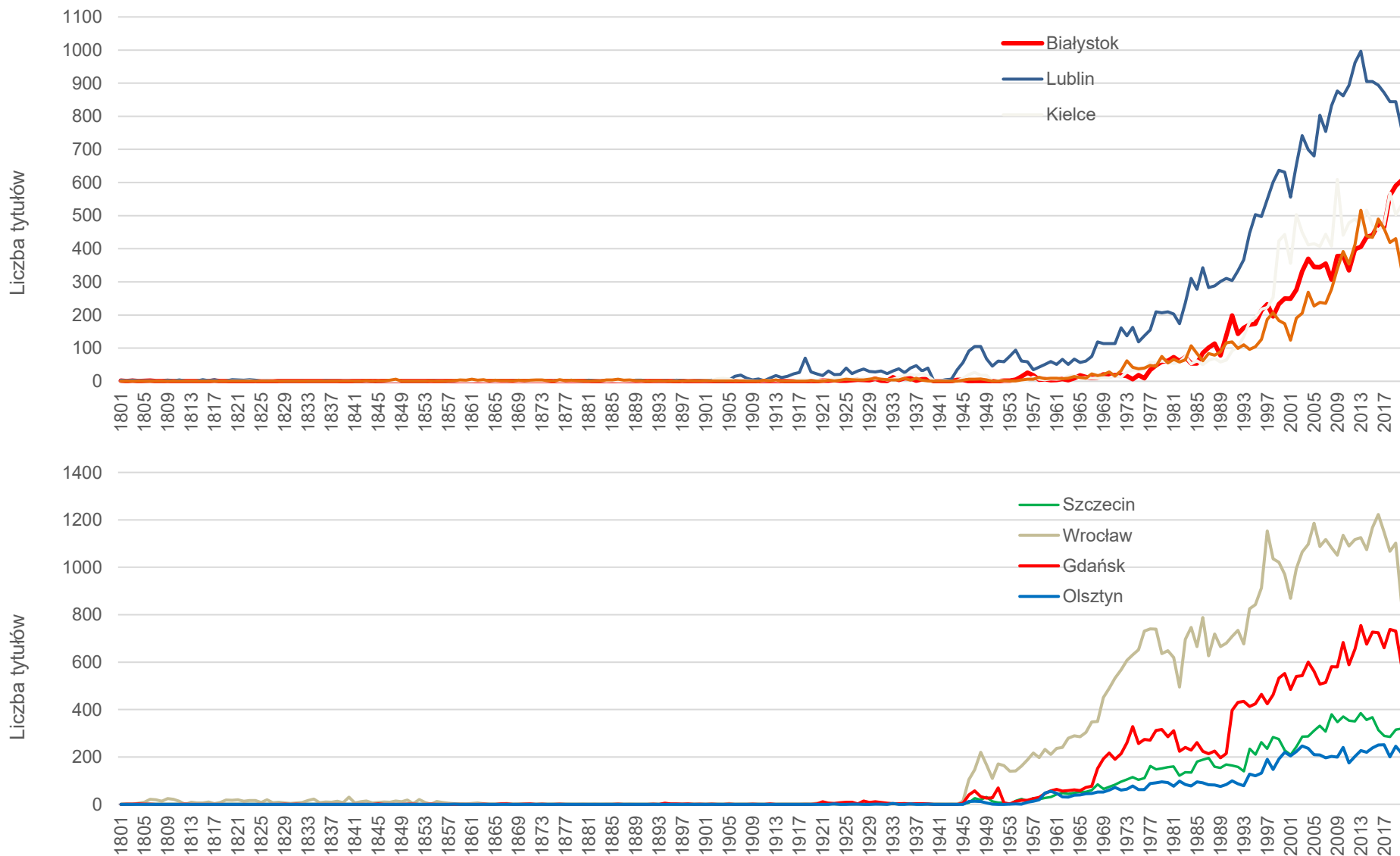
# Łódź, Białystok, Poznań, Wrocław (proporcje)



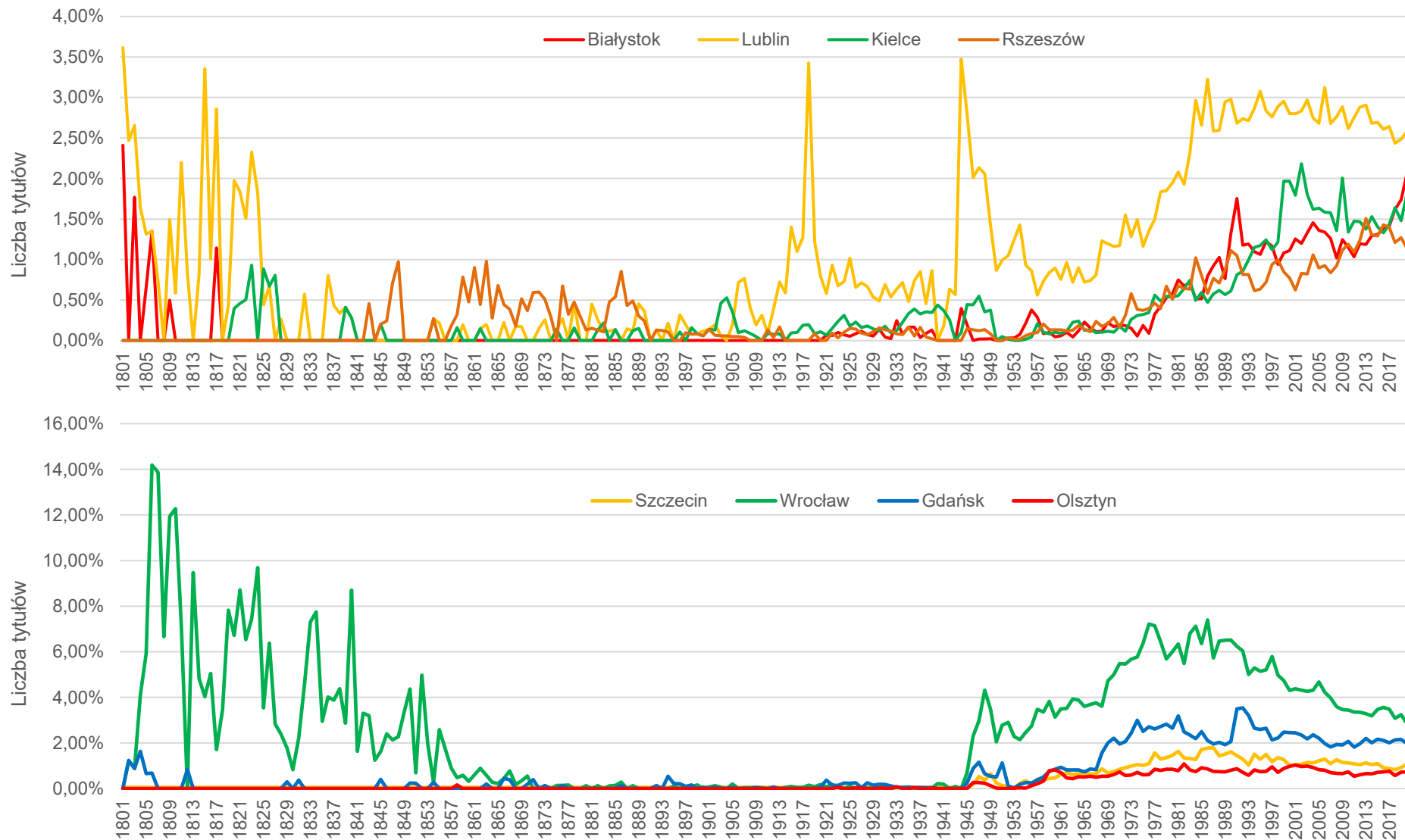
# Warszawa, Kraków, Lwów (proporcje do 1939)



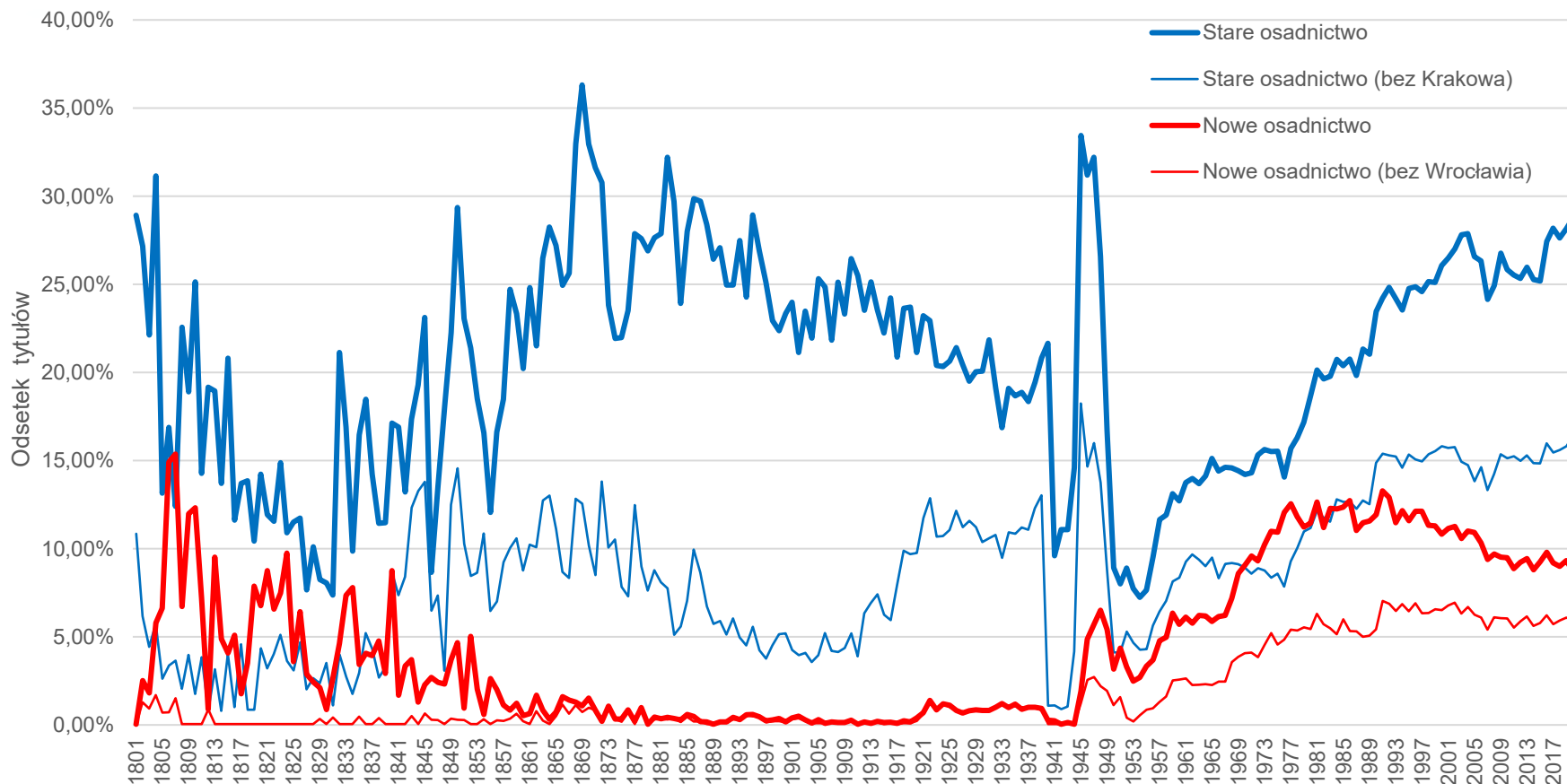
# Wschód i zachód (?) Polski (liczby tyt.)



# Wschód i zachód (?) Polski (liczby tyt.)



# Próba wyjaśnienia różnic: zakorzenienie



# Podsumowania i wnioski

# Podsumowanie

- Oficjalne zaistnienie danych bibliograficznych repertuarze badań NLP, AI i humanistyki cyfrowej
- Dowód wielkiego potencjału wielkich danych bibliograficznych
- Namacalny efekt korzystnej synergii współpracy informatyki i humanistyki

ale także

- Dowód uznania dla rzesz bezimiennych, słabo opłacanych bibliotekarzy z działów opracowania, którzy przez dziesięciolecia skrupulatnie wykonywali swoją pracę

# Bariery i problemy

- Brak zrozumienia i uznania dla badań wielkich bibliografii
- Brak zakorzenienia w tradycyjnych dyscyplinach
- Powolne tworzenie się tożsamości humanistyki cyfrowej i humanistycznych nurtów *data science*



# Trochę autoreklamy



# Dziękujemy za uwagę!