



# LARGE BIBLIOGRAPHIES IN THE DIGITAL HUMANITIES. DATA ACQUISITION, PROCESSING, APPLICATIONS AND LIMITATIONS (1)

prof. dr hab. Adam Pawłowski  
Institute of Information and Media Sciences, University of Wrocław

# LECTURE PROGRAMME

- Bibliographic data as a subject of research
- Data for research (sources)
- Potential hypotheses and objectives of the research
- Data processing methods
- Sample results and interpretation

# BIBLIOGRAPHIC DATA AS A SUBJECT OF RESEARCH

Interreg, Digital Humanities for the Future,  
University of Olomouc, 16-17.01.2023



Uniwersytet  
Wrocławski

**CLARIN**  
Common Language Resources and  
Technology Infrastructure



Politechnika Wrocławska

# BIBLIOGRAPHIC DATA

**Each bibliography is some sort of compilation of publications.**

**The basic division includes bibliographies:**

- general (national);
- special (limited range, scope, type of publication);
- field (subject).

**Commonly, a bibliography is an appendix bibliography to some document.**

# BIBLIOGRAPHIC DATA

**Each compilation of publications is a source of data.**

**Traditionally, the addressee of this information was the individual user.**

**Large bibliographies are de facto structured databases and can be the subject of autonomous research conducted by text mining methods.**

# BIBLIOGRAPHIC DATA — FORMATS

**There is a dominant family of MARC formats, or Machine-Readable Cataloging**

<https://www.loc.gov/marc/bibliographic/ecbdlist.html>

The first version was developed at the Library of Congress in the 1960s, and has been the standard in the US since 1971.

<https://www.loc.gov/loc/lcib/0605/avram.html>

The most widely used version is MARC 21.



## INFORMATION BULLETIN



Search this site

[→ Information Bulletin Home](#)[→ Past Issues](#)[→ About the LCIB](#)

## Related Resources

- [News from the Library of Congress](#)
- [Events at the Library of Congress](#)
- [Exhibitions at the Library of Congress](#)
- [Wise Guide to loc.gov](#)

## Henriette Avram, 'Mother of MARC,' Dies

By MATT SCHUDEL

Henriette D. Avram, whose far-reaching work at the Library of Congress replaced ink-on-paper card catalogs and revolutionized cataloging systems at libraries worldwide, died April 22 of cancer at Baptist Hospital in Miami. Avram, 86, had recently moved to Florida from her home in California, Md.



The practical effect of her complicated mathematical formulations was to make library collections more readily accessible to scholars and the general public. Her work greatly expanded interlibrary loan programs throughout the nation and allowed people to sit at computers and look through automated card catalogs at libraries everywhere.

After working at the National Security Agency during the early years of the computer age, Avram joined the Library of Congress in 1965. With no background in library work, she was assigned to develop an automated cataloging format where none had existed.

Combining two complex fields, computer programming and intricate cataloging practices, she and a small team completed the MARC Pilot Project—for Machine Readable Cataloging—in 1968. The system quickly became the preferred format for libraries throughout the country and, ultimately,

around the globe.

# BIBLIOGRAPHIC DATA — FORMATS

Interesting fact:

**Thus, most of mankind's written output is described by a standard that is now almost 60 years old.**



# MARC 21 FORMAT - EXAMPLE OF A MANUAL

## 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*

# MARC 21 FORMAT - EXAMPLE OF A MANUAL

## 041 - LANGUAGE CODE (R)

### Indicators

First - Translation indication

# - No information provided

0 - Item not a translation/does not include a translation

1 - Item is or includes a translation

Second - Source of code

# - MARC language code

7 - Source specified in subfield \$2

### Subfield Codes

\$a - Language code of text/sound track or separate title (R)

\$b - Language code of summary or abstract (R)

\$c - Languages of separate titles (VM) [OBSOLETE]

\$c - languages of available translation (SE) [OBSOLETE]

\$d - Language code of sung or spoken text (R)

\$e - Language code of librettos (R)

\$f - Language code of table of contents (R)

\$g - Language code of accompanying material other than librettos and transcripts (R)

\$h - Language code of original and/or intermediate translations of text (R) [REDEFINED]

\$h - Language code of original (R)

\$i - Language code of intertitles (R)

\$j - Language code of subtitles (R)

\$k - Language code of intermediate translations (R)

\$m - Language code of original accompanying materials other than librettos (R)

\$n - Language code of original libretto (R)

\$p - Language code of captions (R)

\$q - Language code of accessible audio (R)

\$r - Language code of accessible visual language (non-textual) (R)

\$t - Language code of accompanying transcripts for audiovisual materials (R)

\$2 - Source of code (NR)

\$6 - Linkage (NR)

\$7 - Data provenance (R)

\$8 - Field link and sequence number (R)

# DANE USTRUKTURYZOWANE I NIEUSTRUKTURYZOWANE

**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)

# UNSTRUCTURED DATA

Unstructured data do not include a repetitive arrangement of fields. They are difficult to process, but sometimes rich in information.

A perfect model example of unstructured data is text in natural language.

# EXAMPLE OF A RECORD

## Notation in the appendix bibliography:

Binmore Ken (2017), Teoria gier. Tłum. Iwona Konarzewska. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

98 chars

## Bibliographic description:

BINMORE, Ken (1940- )

*Teoria gier* / Ken Binmore ; tłumaczenie Iwona Konarzewska. – Łódź : Wydawnictwo Uniwersytetu Łódzkiego, 2017. – 206, [1] strona : ilustracje, fotografie, wykresy ; 21 cm. – (Krótkie Wprowadzenie ; 8)

Tytuł oryginału: *Game theory : a very short introduction*.

Bibliografia na stronach 195-200. Indeks.

Dostępne także jako e-book. – Publikacja sfinansowana ze środków Wydawnictwa Uniwersytetu Łódzkiego

ISBN 978-83-8088-594-3

ISBN 978-83-8088-595-0 (e-ISBN)

Typ: Publikacje popularnonaukowe

Gatunek: Opracowanie

Czas powstania dzieła: 2007

Temat: Teoria gier

Ujęcie / dziedzina: Filozofia i etyka

519.83

640 chars

# RECORD AS RETRIEVED BY THE API

```
{
  "id": "5675461",
  "createdDate": "2017-08-07T13:50:59.000+02:00",
  "updatedDate": "2017-11-06T14:54:28.000+01:00",
  "language": "polski",
  "subject": "Teoria gier",
  "subjectPlace": "",
  "subjectTime": "",
  "subjectWork": "",
  "isbnIssn": "9788380885943 9788380885950",
  "author": "Binmore, Ken (1940- ). Konarzewska, Iwona. Wydawnictwo Uniwersytetu Łódzkiego.",
  "placeOfPublication": "Łódź : Polska",
  "location": "",
  "title": "Teoria gier / Game theory : a very short introduction, Krótkie Wprowadzenie ; 8",
  "udc": "519.83 02",
  "publisher": "Wydawnictwo Uniwersytetu Łódzkiego. Wydawnictwo Uniwersytetu Łódzkiego",
  "kind": "książka",
  "domain": "Filozofia i etyka",
  "formOfWork": "Książki Publikacje popularnonaukowe",
  "genre": "Opracowanie",
  "timePeriodOfCreation": "2007",
  "audienceGroup": "",
  "demographicGroup": "",
  "nationalBibliographyNumber": "PB 2017/27081",
  "publicationYear": "2017",
  "languageOfOriginal": "angielski",
  "fixedFields": {
    "label": "LANG",
    "value": "pol",
    "display": "Polish",
    "id": "24",
    "label": "COUNTRY",
    "value": "pl",
    "display": "Polska",
    "id": "89",
    "label": "CAT DATE",
    "value": "2017-08-07",
    "id": "28",
    "label": "CREATED",
    "value": "2017-08-07T11:50:59Z",
    "id": "83",
    "label": "MARCTYPE",
    "value": "107",
    "label": "REVISIONS",
    "value": "9",
    "id": "85",
    "label": "SUPPRESS",
    "value": "b",
    "id": "31",
    "label": "SKIP",
    "value": "0",
    "id": "25",
    "label": "REC TYPE",
    "value": "b",
    "id": "80",
    "label": "MAT TYPE",
    "value": "a",
    "display": "Book",
    "id": "30",
    "label": "COPIES",
    "value": "0",
    "id": "27",
    "label": "PDATE",
    "value": "2017-08-30T10:55:06Z",
    "id": "98",
    "label": "BIB LVL",
    "value": "m",
    "display": "Monograph",
    "id": "29",
    "label": "AGENCY",
    "value": "1",
    "id": "86",
    "label": "UPDATED",
    "value": "2017-11-06T13:54:28Z",
    "id": "84",
    "label": "RECORD #",
    "value": "5675461",
    "id": "81",
    "label": "LOCATION",
    "value": "multi",
    "id": "26",
    "varFields": {
      "fieldTag": "a",
      "marcTag": "100",
      "ind1": "1",
      "ind2": "",
      "subfields": {
        "tag": "a",
        "content": "Binmore, Ken",
        "tag": "d",
        "content": "(1940- )",
        "tag": "e",
        "content": "Autor"
      }
    },
    "fieldTag": "b",
    "marcTag": "700",
    "ind1": "1",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "Konarzewska, Iwona",
      "tag": "e",
      "content": "Tłumaczenie"
    },
    "fieldTag": "b",
    "marcTag": "710",
    "ind1": "2",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "Wydawnictwo Uniwersytetu Łódzkiego",
      "tag": "e",
      "content": "Wydawca"
    },
    "tag": "4",
    "content": "pbl",
    "fieldTag": "d",
    "marcTag": "380",
    "ind1": "",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "Książki"
    },
    "fieldTag": "d",
    "marcTag": "380",
    "ind1": "",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "Publikacje popularnonaukowe"
    },
    "fieldTag": "d",
    "marcTag": "650",
    "ind1": "",
    "ind2": "4",
    "subfields": {
      "tag": "a",
      "content": "Teoria gier"
    },
    "fieldTag": "d",
    "marcTag": "655",
    "ind1": "",
    "ind2": "4",
    "subfields": {
      "tag": "a",
      "content": "Opracowanie"
    },
    "fieldTag": "d",
    "marcTag": "658",
    "ind1": "",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "Filozofia i etyka"
    },
    "fieldTag": "g",
    "marcTag": "015",
    "ind1": "",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "PB 2017/27081"
    },
    "fieldTag": "i",
    "marcTag": "020",
    "ind1": "",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "9788380885943"
    },
    "fieldTag": "i",
    "marcTag": "020",
    "ind1": "",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "9788380885950"
    },
    "tag": "q",
    "content": "e-ISBN"
  },
  "fieldTag": "j",
  "marcTag": "080",
  "ind1": "",
  "ind2": "",
  "subfields": {
    "tag": "a",
    "content": "519.83"
  },
  "fieldTag": "l",
  "marcTag": "998",
  "ind1": "",
  "ind2": "",
  "subfields": {
    "tag": "a",
    "content": "ik"
  },
  "fieldTag": "n",
  "marcTag": "504",
  "ind1": "",
  "ind2": "",
  "subfields": {
    "tag": "a",
    "content": "Bibliografia na stronach 195-200. Indeks"
  },
  "fieldTag": "n",
  "marcTag": "530",
  "ind1": "",
  "ind2": "",
  "subfields": {
    "tag": "a",
    "content": "Dostępne także jako e-book"
  },
  "fieldTag": "n",
  "marcTag": "536",
  "ind1": "",
  "ind2": "",
  "subfields": {
    "tag": "a",
    "content": "Publikacja sfinansowana ze środków Wydawnictwa Uniwersytetu Łódzkiego"
  },
  "fieldTag": "p",
  "marcTag": "260",
  "ind1": "",
  "ind2": "",
  "subfields": {
    "tag": "a",
    "content": "Łódź",
    "tag": "b",
    "content": "Wydawnictwo Uniwersytetu Łódzkiego",
    "tag": "c",
    "content": "2017",
    "fieldTag": "r",
    "marcTag": "300",
    "ind1": "",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "206, [1] strona",
      "tag": "b",
      "content": "ilustracje, fotografie, wykresy",
      "tag": "c",
      "content": "21 cm"
    },
    "fieldTag": "s",
    "marcTag": "490",
    "ind1": "1",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "Krótkie Wprowadzenie",
      "tag": "v",
      "content": "8"
    },
    "fieldTag": "s",
    "marcTag": "830",
    "ind1": "1",
    "ind2": "0",
    "subfields": {
      "tag": "a",
      "content": "Krótkie Wprowadzenie",
      "tag": "v",
      "content": "8"
    },
    "fieldTag": "t",
    "marcTag": "245",
    "ind1": "1",
    "ind2": "0",
    "subfields": {
      "tag": "a",
      "content": "Teoria gier /",
      "tag": "c",
      "content": "Ken Binmore ; tłumaczenie Iwona Konarzewska"
    },
    "fieldTag": "u",
    "marcTag": "246",
    "ind1": "1",
    "ind2": "",
    "subfields": {
      "tag": "i",
      "content": "Tytuł oryginału",
      "tag": "a",
      "content": "Game theory",
      "tag": "b",
      "content": "a very short introduction",
      "tag": "f",
      "content": "2007"
    },
    "fieldTag": "y",
    "marcTag": "008",
    "ind1": "",
    "ind2": "",
    "content": "170807s2017 pl aod |001 0 pol nam i",
    "fieldTag": "y",
    "marcTag": "040",
    "ind1": "",
    "ind2": "",
    "subfields": {
      "tag": "a",
      "content": "WA N",
      "tag": "c",
      "content": "WA N",
      "fieldTag": "y",
      "marcTag": "041",
      "ind1": "1",
      "ind2": "1",
      "subfields": {
        "tag": "a",
        "content": "pol",
        "tag": "h",
        "content": "eng"
      },
      "fieldTag": "y",
      "marcTag": "046",
      "ind1": "",
      "ind2": "",
      "subfields": {
        "tag": "k",
        "content": "2007"
      },
      "fieldTag": "y",
      "marcTag": "336",
      "ind1": "",
      "ind2": "",
      "subfields": {
        "tag": "a",
        "content": "Tekst",
        "tag": "b",
        "content": "txt",
        "tag": "2",
        "content": "rdamedia"
      },
      "fieldTag": "y",
      "marcTag": "337",
      "ind1": "",
      "ind2": "",
      "subfields": {
        "tag": "a",
        "content": "Bez urzęduzenia pośredniczącego",
        "tag": "b",
        "content": "n",
        "tag": "2",
        "content": "rdamedia"
      },
      "fieldTag": "y",
      "marcTag": "338",
      "ind1": "",
      "ind2": "",
      "subfields": {
        "tag": "a",
        "content": "Wolumin",
        "tag": "b",
        "content": "nc",
        "tag": "2",
        "content": "rdacarrier"
      },
      "fieldTag": "y",
      "marcTag": "920",
      "ind1": "",
      "ind2": "",
      "subfields": {
        "tag": "a",
        "content": "978-83-8088-594-3"
      },
      "fieldTag": "y",
      "marcTag": "920",
      "ind1": "",
      "ind2": "",
      "subfields": {
        "tag": "a",
        "content": "978-83-8088-595-0 (e-ISBN)"
      },
      "fieldTag": "y",
      "marcTag": "999",
      "ind1": "",
      "ind2": "",
      "subfields": {
        "tag": "a",
        "content": "zkd",
        "tag": "b",
        "content": "eow",
        "tag": "x",
        "content": "33",
        "tag": "y",
        "content": "17"
      },
      "fieldTag": "y",
      "marcTag": "084",
      "ind1": "1",
      "ind2": "1",
      "subfields": {
        "tag": "a",
        "content": "02",
        "tag": "b",
        "content": "00000nam a2200517 i 4500",
        "marc": {
          "leader": "00000nam a2200517 i 4500",
          "fields": {
            "001": "b56754619",
            "008": "170807s2017 pl aod |001 0 pol nam i",
            "015": {
              "ind1": "",
              "ind2": "",
              "subfields": {
                "a": "PB 2017/27081"
              },
              "020": {
                "ind1": "",
                "ind2": "",
                "subfields": {
                  "a": "9788380885943"
                },
                "020": {
                  "ind1": "",
                  "ind2": "",
                  "subfields": {
                    "a": "9788380885950"
                  },
                  "q": "e-ISBN"
                },
                "040": {
                  "ind1": "",
                  "ind2": "",
                  "subfields": {
                    "a": "WA N",
                    "c": "WA N"
                  },
                  "041": {
                    "ind1": "1",
                    "ind2": "",
                    "subfields": {
                      "a": "pol",
                      "h": "eng"
                    },
                    "046": {
                      "ind1": "",
                      "ind2": "",
                      "subfields": {
                        "k": "2007"
                      },
                      "080": {
                        "ind1": "",
                        "ind2": "",
                        "subfields": {
                          "a": "519.83"
                        },
                          "084": {
                            "ind1": "",
                            "ind2": "",
                            "subfields": {
                              "a": "02"
                            },
                              "100": {
                                "ind1": "1",
                                "ind2": "",
                                "subfields": {
                                  "a": "Binmore, Ken",
                                  "d": "(1940- )",
                                  "e": "Autor"
                                },
                                  "245": {
                                    "ind1": "1",
                                    "ind2": "0",
                                    "subfields": {
                                      "a": "Teoria gier /",
                                      "c": "Ken Binmore ; tłumaczenie Iwona Konarzewska"
                                    },
                                      "246": {
                                        "ind1": "1",
                                        "ind2": "",
                                        "subfields": {
                                          "i": "Tytuł oryginału",
                                          "a": "Game theory",
                                          "b": "a very short introduction",
                                          "f": "2007"
                                        },
                                        "260": {
                                          "ind1": "",
                                          "ind2": "",
                                          "subfields": {
                                            "a": "Łódź",
                                            "b": "Wydawnictwo Uniwersytetu Łódzkiego",
                                            "c": "2017",
                                            "300": {
                                              "ind1": "",
                                              "ind2": "",
                                              "subfields": {
                                                "a": "206, [1] strona",
                                                "b": "ilustracje, fotografie, wykresy",
                                                "c": "21 cm"
                                              },
                                              "336": {
                                                "ind1": ""
                                              }
                                            }
                                          }
                                        }
                                      }
                                    }
                                  }
                                }
                              }
                            }
                          }
                        }
                      }
                    }
                  }
                }
              }
            }
          }
        }
      }
    }
  }
}
```

9324 chars



# RESEARCH DATA (SOURCES)



# RESEARCH DATA

## **1. The digital catalog of the National Library (BN)**

- period covered: 1800-2010
- volume: ca 1.9 million records (after filtering ca 1.25 million).

## **2. The digital catalogs of the Deutsche Nationalbibliothek**

- period covered: 1800-2010
- volume: ca 7.5 million records (after filtering 5 million).

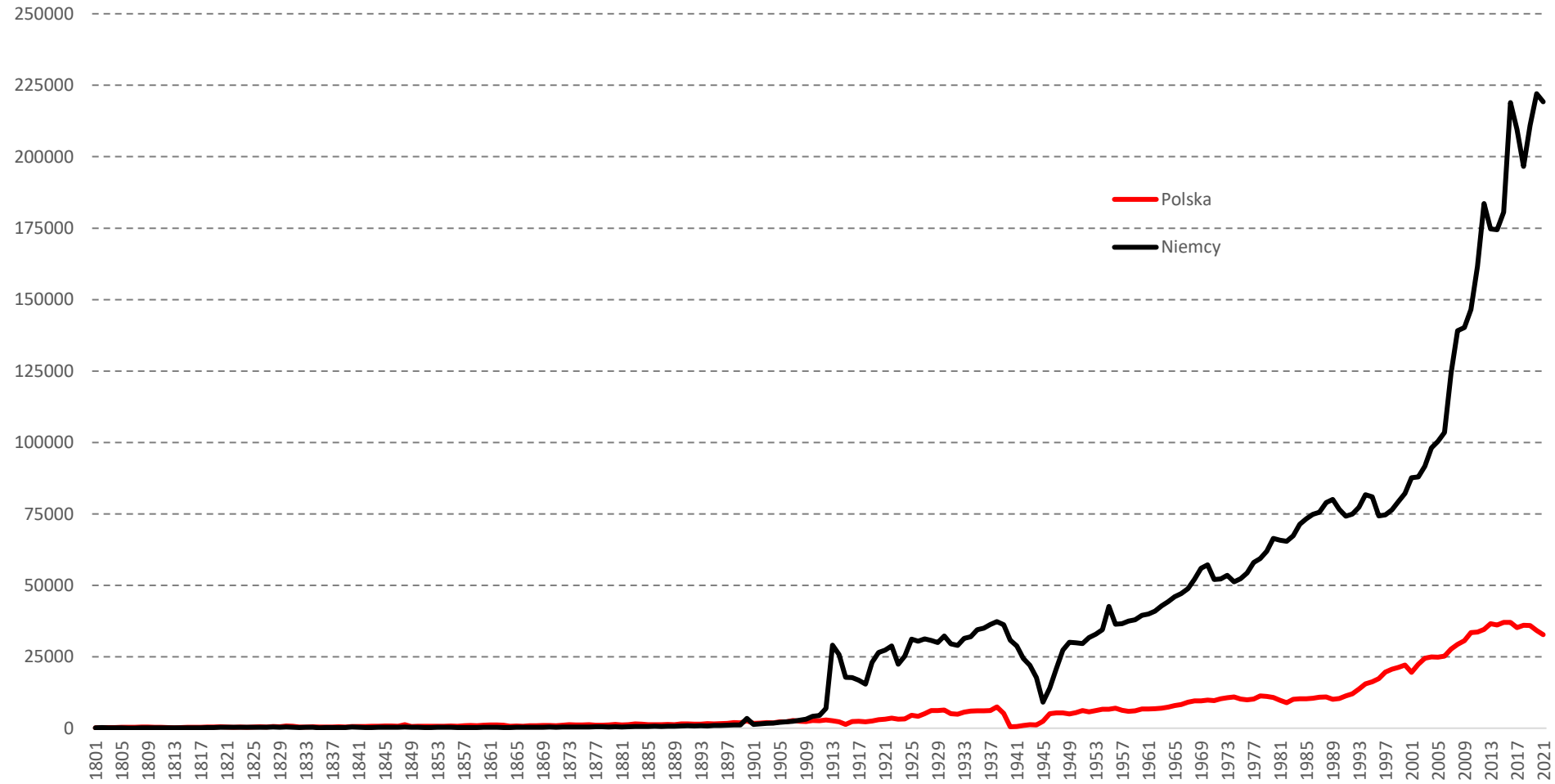
## **3. The type of content studied**

- book, that is, what has an ISBN number.



# TOTAL CIRCULATION OF BOOKS

Number of books published PL / DE



# REASONS FOR DIFFERENCES BETWEEN PL AND DE

1. Different potential of both countries (demographics, economics, science)
2. Different cataloging systems
  - e.g. PhDs have ISBN in DE, but are not included in PL
  - each medium / version gets a separate number in DE, in PL it is not always the case
  - strong pressure for creating digital and audio versions in DE, less pressure in PL

# EXAMPLES FROM DATABASES (VARIOUS LANGUAGES)

1. MARC 21 in Czech databases
2. MARC 21 in Norwegian databases
3. MARC 21 in Polish databases
4. MARC 21 in German databases

From external files.

# EXAMPLES FROM OLD DATABASES

34

MACKEWICZ — MACZYŃSKI.

czyna chorób hemoroidalnych, z wskazaniem środków najpewniejszych i najspieszniejszych wyleczenia i zapobieżenia takowemu bez żadnego uszkodzenia, z angielskiego, podług siódmej poprawnej edycji. Warszawa, S. Orzelbrand, druk M. Chmielewskiego, 1849, w 12ce, str. 128. (Przedruk edycji lwowskiej.)

— Obacz: Hermbstädt.

Mackiewicz J. Szkice badań choroby. Warszawa, 1849; obacz: Swoboda.

Macleod, obacz: Zieliński W.

Macpherson Jakób (urodz. 1738 † 1796), ob. Ossyan.

MACQUER Filip (urodz. 1720 † 1770). Dzieje Rzeczypospolitej Rzymskiej od założenia Rzymu aż do Cesarzów krótko porządkiem lat opisane, z francuskiego na polskie przełożone, potrzebnymi przypiskami objaśnione (przez X. J. Albertandego), a w tem nowym wydaniu mianowicie co do przypisków poprawione, przez tegoż, który pierwotne tłumaczenie przed 40 laty wykonał. Z przydatkiem potrzebnej wiadomości o klasycznych autorach, co w tym przeciągu czasu kwitniali. Tomów II. Warszawa, Zawadzka wdowa, 1806, w 8ce; Tom I, str. XVII i 617; Tom II, str. 11, 598 i LVII. 36 złp.

(Pierwsze wydanie: Warszawa, 1768. +)

Macquet Aug., obacz: Dumas Al.

MACRI Michel Angiolo. Memorie istorico critiche intorno alla vita ed alle opere di Monsignore fratre Paolo Piromalli Domenicano, arcivescovo di Naxivan, aggiuntavi la Siderografia. Napoli, nella tipogr. della Società filomatia, 1824. (Piromalli, urodz. około 1592 † 1667, był w Polsce za Władysława IV.)

MACUDZIŃSKI W. (Dr.). Wiadomość o Bóbrce, podał Dr.... Jasio, druk Ludwika Dyoniziego Stoegera, 1871, w 8ce, str. 24. + +

MACZAKIEWICZ X. A. Mowa przy złożeniu w grobie Niedźwiedzkim przywiezionych z Krakowa zwłok ś.p. J. W. z hr. Wodzieńskich Tekli hr. Małachowskiej wojewodziny krakowskiej, przez.... Plebana w Górze Koscielniczkiej w d. 23 Grudnia 1829 r. miana. Na korzyść ubogich. Kraków, drukiem Gieszkowskiego, 1830, w 8ce, str. 12. + + +

— Nabożeństwo do św. Wiktorii Panny i maczenniczki, z dodaniem innych modlitw ułożone i wydane dla pożytku i wygody parafian Komuszy, przez X.... miejscowego proboszcza. Warszawa, w drukarni XX. Missyonarzy u św. Krzyża, 1856, w 8ce, str. 21.

MACZUSKI Andrzej. O przyjaźniach i przyjaciółach, Andrzeja Maczuszkiego Pismo wydane w Dobromilu według wszelkiego do prawdy podobieństwa między r. 1530 a 1540, na jaw ogłosza, z dodatkami tekstów oryginalnych S. F. Zukowski. Wilno, Zawadzki, 1817, w 8ce, str. XX i 203, z 2ma rycinami. 7 złp. +

(Są tu przekłady z Cycerona, Plutarcha, Teofrasta, Hekatora.)

Mączka, obacz: Otto.

Mączyska Maryanna z Dzanottów, ob. Łancucki Józef; Mączyński Józef i Wojciech.

MĄCZYŃSKI Józef (urodz. 1805 † 14 Października 1862). Archiprezbiterowie i Kapłani kościoła Maryackiego w Krakowie. (Odrdruck z Dodatku Czasu.) Kraków, druk Czasu, 1857, w 8ce, str. 47. 2 złp. +

— Cracovie et ses environs. Description historique, géographique et pittoresque de cette ville et de ses contrées, illustrée de plusieurs plans et litogr. Cracovie, J. Czech, 1846, w 12ce, str. IV i 373. (Skrócenie przez Fel. Wrotnowskiego.) 18 złp. +

— Cudami wstawiony Obraz Najświętszej Panny Maryi w baszcie bramy Floryjańskiej w Krakowie. Kraków, czioczkami Czasu, 1859, w 8ce, str. 15. +

— Kościuszkowskie czasy, z księgi wspomnień Wojciecha Maczyńskiego pułkownika komend. Milicji Rzeczypospolitej Krakowskiej. (Odrdruck z Dodatku miesięcznego Czasu, Sierpień.) Kraków, druk Czasu, 1857, w 8ce, str. 72. +

— Kraków dawny i teraźniejszy, z przeglądem jego okolic, opisał.... Z rycinami. Kraków, druk i nakład J. Czecha, 1854, w 12ce, str. XXIV i 236. (Z pamfletem przeciw Biskupowi L. Żętkowskiemu.) +

— Pamiętka z Krakowa, opis tego miasta i jego okolic, przez.... Z rycinami i planami. Tomów III. Kraków, J. Czech, 1845, w 12ce; Tom I, kart 2, str. 227, karta 1, rycin 3; Tom II, kart 2, str. 402, i rycin 5; Tom III, kart 2, str. 313 i 2, rycin 5. 24 złp. +

— Kilka podań i wspomnień krakowskich ze zbioru wiadomości o Krakowie. (Odrdruck z Czasu.) Kraków, Czech, 1855, w 8ce, str. 120. +

— (Treść: Duchy na Krzemionkach; Podziemie zamek krakowski; Stopka królowej Jadwigi; Pamiętka po królowej Jadwidze; Wieść kościoła Panny Maryi; Statua N. Panny przed kościołem OO. Kapucynów; Widmo w Krzyżstoforach; Cudowna obrona Krakowa w r. 1768; Wąsy Króla Jana III.) 4 złp.

— Pojawienie się Eleonory Lubomirskiej spoczywającej w grobach kościoła św. Trójcy OO. Dominikanów w Krakowie. Wydane na dochód podźwignienia z grusów tegoż kościoła, ze zbioru podań. (Odrdruck z Czasu.) Kraków, Czech, 1853, w 16ce, str. 33. +

— Powinszowanie imienin W. J. P. Maryannie z Dzanottów Maczeńskiej d. 8 Lutego 1817 r., fol. 1/2 ark. (Wiersz).

— Rady i zwady niewiast o wielożenstwie. Ze zbioru wiadomości o Krakowie. (Osobne odbicie z Dodatku do Czasu na miesiąc Kwiecień.) Kraków, czioczkami drukarni Czasu, 1855, w 8ce, str. 63. 50 cent. +

MĄCZYŃSKI — MADEJSKI.

35

— Szkoła kucharek czyli Łatwy sposób wyuczenia się różnych potraw mięsnych i roślinnych, na wzór dzieła Carème kuchmistrza Napoleona, ułożył L. Z. Kraków, J. Czech, 1843, w 12ce, str. XIV i 207. +

(Oryginalnie ułożył z notatki matki swej. Dla żartu zaś położył pierwsze litery nazwiska znanego paleontologa.)

— Toż, inna edycja. Tamże, 1843, str. 219. 4 złp.

— Wiadomość historyczna o cudownym obrazie N. Maryi Panny Różańca św. w kościele OO. Dominikanów w Krakowie. Kraków, druk Czasu, 1863, w 8ce, str. 63, z 2 rycin. litogr. 1 złp. 6 gr. +

— Włościanie z okolic Krakowa w zarysie, z rycinami Michała Stachowicza. Kraków, nakład i druk J. Czecha, 1858, w 12ce, str. 122 i 12 rycin kolorow. +

— Wspomnienia historyczne Krakowa na każdy dzień roku. Kraków, J. Czech, 1834, w 8ce, str. 30. +

— Założenie Biskupstwa krakowskiego. (Odbitka z XXIX. Tomu Rocznicu Towarzystwa naukowego.) Kraków, drukiem Uniwersytetu, 1862, w 8ce, str. 32. +

— Zbierecz umysłowych rozrywek. Tomów II. (Ułożył J. Maczyński z przedruków powiastek z różnych czasopismów i tłumaczeń.) Kraków, J. Czech, 1839, w 12ce.

— Życie i śmierć Józefa Chłopickiego (Czas, dodatek miesięczny, Lipiec i Sierpień, 1858).

(Maczyński Józef.) Odpowiedź na zarzuty Kapituły katedralnej krakowskiej poczynione przez pana Józefa Maczyńskiego w Notatach jego z Wędrówki po Krakowie, w części literackiej Dziennika krakowskiego Czas umieszczane. Kraków, czioczkami Karola Budweiera, 1859, w 8ce, str. 32. +

— Obacz: Zbierecz.

(Maczyński Wojciech.) Wojciechowi Maczeńskiemu powtarzającemu ślub uroczysty z Maryanną z Dzanottów w 50 roku szczególnego małżeństwa w Krakowie d. 15 Lipca 1813 r., wnuki, w 4ce, str. 4. (Wiersz.) +

Madagaskar, obacz: Beniowski A.; Parny; Pleszczewski L. (Język).

(Madalińska Joanna.) Summariusz dowodów w sprawie Prokuratury W. K. P. na rzecz Skarbu Królestwa przeciw Joannie z Krakowskich Madalińskiej o część dóbr Mierzy, Łaszowa i Toporowa. Warszawa, drukiem J. Jaworskiego, 1855, fol. str. 36.

— Obacz: Przeszkodziński.

Madalinsky Albertine, oder das Mädchen aus Polen. Aus dem Englischen. Leipzig, 1801, w 8ce.

Madaj obacz: Wąsik.

MADEJSKI Edward. Dyetetyka dzieci oparta na wskazówkach przyrody. Przewodnik w zdrowym pielęgnowaniu dzieci od przyjścia na

świat aż do dojrzenia, określony przez Przajaciela dzieci. Lwów, Wild, druk E. Winia-rza, 1869, w 8ce, str. XVI i 150. 1 złr. 15 cent.; zniż. 60 cent. +

— Gimnastyka racjonalna, jako część dyetetyki ze stanowiska lekarskiego historyczno-krytycznie i praktycznie opracowana przez autora Dyetetyki dzieci. Dwie Części, z 96 w tekście umieszczonymi drzeworytami. Warszawa, Gebethner i Wolff, drukiem J. Ungra, 1871, w 32ce; Część I, str. VIII, 174 i rejestr.; Część II, str. 116 i rejestr. 1 rub. 20 kop. +

— Plan i cel Zakładu gimnastycznego (t. z. organopatycznego). Lwów, Wild, 1871, w 8ce, str. 15. 20 cent.

— Kilka słów o migrenie przez.... homeopatę i organopatę. Lwów, 1874, w 8ce, kart 2.

MADEJSKI Jan, obacz: Korzeniowski Leon.

MADEJSKI Leon. Nekrolog § 998 Ustawy cywilnej, przez.... (Osobne odbicie z Czasopisma prawniczego Uniwersytetu Jagellońskiego, 1869, w 8ce, karta 1, str. 7. +

— Zdania ze wszystkich galezi prawa i umiejętności politycznych, których w celu otrzymania godności obęga Praw Doktora etc., podejmuje się publiczenie bronić.... Kraków, druk Uniwersytetu, 1862, w 8ce, str. 7. +

MADEJSKI Marcelli (urodz. 1822 r.). Krakowiaki do śpiewu z towarzyszeniem fortepianu; Nr. 1: Wiosna o zaraniu, słowa L. P. 40 cent.; Nr. 2: Zapomniane skrzypki moje, słowa Edmunda Wasilewskiego. 30 cent.; Nr. 3: Oj pada listek pada, słowa J. Zachariasiewicza. 70 cent.

— Pieśni. Słowa Bohdana Zaleskiego ułożone w muzykę z towarzyszeniem fortepianu. Lwów, Kallenbach, 1852, w 4ce (nuty z tekstem). Nr. 1: Nigdyż? 30 cent.; Nr. 2: Tryolety. 45 cent.; Nr. 3: Zakończana. 45 cent.; Nr. 4: Co ja widział. 60 cent.

— Piosenki na głos „mezzo soprano“, z towarzyszeniem fortepianu, mianowicie: Wspomnienie; Modlitwy moje; Kukulka. Lwów, K. Wild, w 4ce (nuty z tekstem). 1 złr. 6 cent.

— Tajemnica. Pieśń z 3go aktu komedii: „Drzymka pana Prospera.“ Lwów, K. Wild, w 4ce. 45 cent.

— Obacz: Zaleski J. B.

MADEJSKI Marian Mikołaj. Sätze aus sämtlichen Rechts- und politischen Wissenschaften etc. Lemberg, Druck von Sznajder, 1846, w 8ce, str. 8.

MADEJSKI Jerzy Stanisław (Dr.). Podręcznik sądowy ułożył i wydał.... Adjunkt sądowy w Podgórzu. Kraków, drukiem Uniwersytetu, 1870, w 8ce, str. XXXV i 72. +

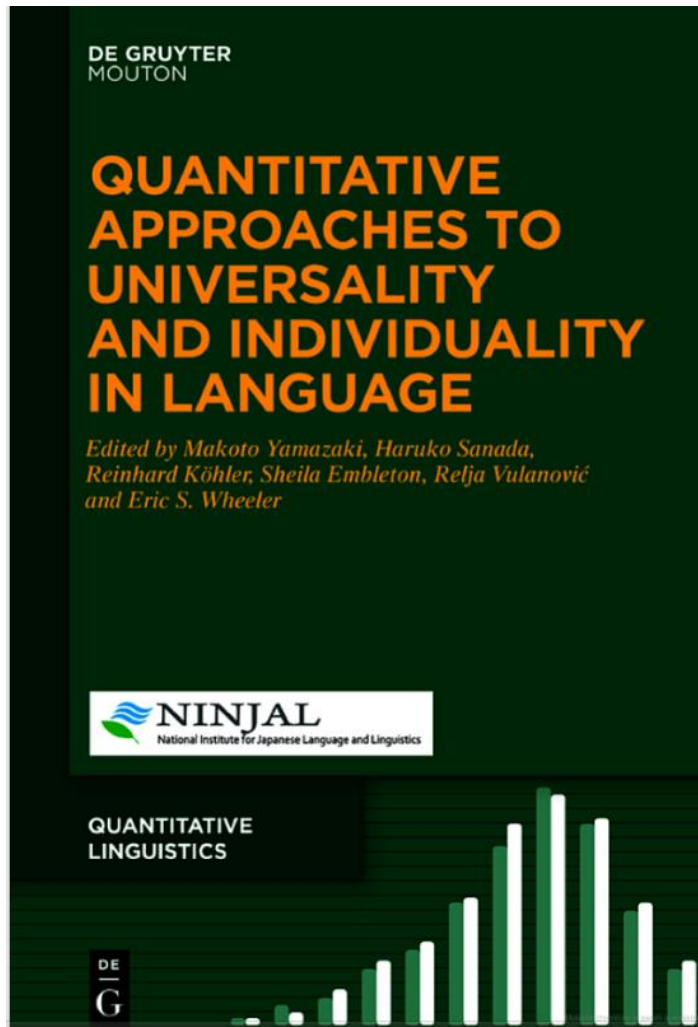
— Podręcznika sądowego Część II. Niemiecko-polski słownik prawnicy cywilny, na podstawie dzieł prawniczych Królestwa Pol-

Biografia XIX wieku. I Wydział. Tom II

www.eschke.net.pl



# MULTIPLICATION OF BOOKS (DE)



Artificial  
proliferation of  
publications.

ISBN 9783110628081  
e-ISBN (PDF) 9783110763560  
e-ISBN (EPUB) 9783110763638

**Bibliographic information published by the Deutsche Nationalbibliothek**  
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2023 Walter de Gruyter GmbH, Berlin/Boston

**Language:** English  
**Publisher:** De Gruyter Mouton  
**Copyright year:** 2023  
**Audience:** Researchers in Quantitative Linguistics, Corpus Linguistics, Computational Linguistics, Digital Humanities  
**Pages**  
Front matter: 8  
Main content: 229  
**Illustrations**  
Illustrations: 43  
Coloured Illustrations: 30  
Tables: 50  
**Keywords:** Quantitative Linguistics; Text Analysis; Authorship Attribution; Corpus Linguistics.

---

**eBook**  
Planned Publication: November 7, 2022  
ISBN: 9783110763560

**Hardcover**  
Planned Publication: November 7, 2022  
ISBN: 9783110628081

# POTENTIAL HYPOTHESES AND OBJECTIVES OF THE RESEARCH

Interreg, Digital Humanities for the Future,  
University of Olomouc, 16-17.01.2023



Uniwersytet  
Wrocławski

**CLARIN**  
Common Language Resources and  
Technology Infrastructure



Politechnika Wrocławska

# HYPOTHESES AND OBJECTIVES OF THE RESEARCH

## **Objective:**

- To create a resource and applications that allow semantic analysis of bibliographic databases.

## **Approach:**

- Use of text data in records (titles, subject headings) and other fields (year, location, toponyms, antonyms).
- Use of AI and text-mining methods.
- Generating dependency charts, geolocation of publications and concept maps.

# HYPOTHESES AND OBJECTIVES OF THE RESEARCH

## Hypotheses:

- Gender participation in intellectual life can be measured by the number of publications.
- Internal variation in a country's level of development can be measured by the number of publications and visualized by projecting the data onto a map.

## General assumption:

- Book publications are a synthesis of many specific indicators (education level, social status, wealth).



# DATA PROCESSING METHODS

Interreg, Digital Humanities for the Future,  
University of Olomouc, 16-17.01.2023



Uniwersytet  
Wrocławski

**CLARIN**  
Common Language Resources and  
Technology Infrastructure



Politechnika Wrocławska

# TOPONYM EXTRACTION FROM TITLES

- Title (text) -> list of place names
- Named entity recognition tool for Polish
  - machine learning tool, not a dictionary
  - **detection of n.e.r** – PolDeepNer
    - XMLRoberta architecture
    - starting language model HerBERT (large BERT for Polish)
    - trained to detect NE boundary detection and 82 categories
    - 20 types are place names (i.e., country, city, lake, mountain)
  - **lemmatization of detected name entities (NE)**
    - Polish is highly inflectional
    - *Niemieckiej Republice Ludowej* -> *Niemiecka Republika Ludowa*
    - tagger (recurrent neural network) + Polem for multiword NE



Uniwersytet  
Wrocławski

**CLARIN**  
Common Language Resources and  
Technology Infrastructure



Politechnika Wrocławska

# THANK YOU FOR YOUR ATTENTION