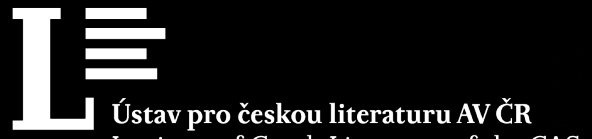
Building a European Literary Bibliography

Challenges and Strategies in Integrating and Presenting Multilingual Bibliographical Data

Ondřej Vimr & Vojtěch Malínek
Czech Academy of Sciences, Institute of Czech Literature, Czech Literary Bibliography
Linked Open Data and Literary Studies, Freie Universität Berlin, 19th November 2024



European Literary Bibliography

- Bibliographical data as an essential resource (Lahti et al. 2019, Umerle et al. 2022)
- Mission: ecosystem for **harmonising, enriching and presenting open literary bibliographical datasets**
- Developed jointly by Czech Literary Bibliography and Polish Literary Bibliography https://literarybibliography.eu

vufind®
Search. Discover. Share.

Historical Overview

2019 – first demoversion containing of Czech and Polish Literary Bibliographies records

2021-2023 – DARIAH.lab project of IBL PAN

2022 – European Literary Bibliography 1.0 realised

2022 – inclusion of Spanish and Finnish data

2023 – work on BOSLIT data

2024 – European Literary Bibliography 2.0 under preparation + Czech National Bibliography data processed

ELB - Challenges

- Bibliographical data aggregation (literature, CC0/BY, live data)
- Multilingual semistructured data (diverse formats, national vs. universal identifiers, subject heading ontologies, etc.)
- Data enrichment and LOD-ification (geolocation, WikiData interconnection, etc.)
- Data presentation

Data Sources

Bibliographical (catalogue) data

- Spanish National Library
- Finnish National Library
- Czech Literary Bibliography (CLB)
- Polish Literary Bibliography (PBL)
- Bibliography of Scottish Literature in Translation (BOSLIT)

Bibliodata formats

```
=LDR  00821nam a2200265 c 4500
=001  a5933628
=005  20190125005347.6
=008  170403s2016\\\\sp\\\\\\|\|||\000\p\spa\\
=015  \\$aMON16
=017  \\$aZ 1435-2016$bOficina Depósito Legal Zaragoza
=020  \\$a978-84-946134-0-1
=035  \\$a(OCoLC)1027700728
=040  \\$aSpMaBN$bspa$cSpMaBN$erdc
=080  \\$a821.134.2-14"19/20"(082.2)
=245  00$aAntiaérea II :$bpoesía encontrada en Zaragoza$c[Ajo ... et al.]
=250  \\$a1ª ed.
=260  \\$aZaragoza$bPregunta$c2016
=300  \\$a104 p.$c18 cm
=336  \\$aTexto (visual)$2isbdcontent
=337  \\$asin mediación$2isbdmedia
=380  \\$iMajor genre$aLiterature$leng
=381  \\$iMajor genre$aLyrical poetry$leng
=650  \7$aPoesías españolas$yS.XX$2embne
=650  \7$aPoesías españolas$yS.XXI$2embne
=650  \7$aPoetry$2ELB-g
=650  \7$aSpanish literature$2ELB-n
=655  \7$aPoesías$2tgfbne
=655  \7$aAntologías$2tgfbne
=700  0\$aAjo$4aut
=995  \\$aBiblioteca Nacional de España
```


```
{
    "creator":
        [
        "Casado Velarde, Manuel",
        "Vilarnovo, Antonio"
        ],
    "date":
        [
        "2003"
        ],
    "format":
        [
        "application/pdf"
        ],
    "identifier":
        [
        "https://dialnet.unirioja.es/servlet/oaiart?codigo=798509",
        "(Revista) ISSN 0210-8178"
        ],
    "language":
        [
        "spa"
        ],
    "publisher":
        [
        "Universidad de Extremadura: Servicio de Publicaciones"
        ],
    "rights":
        [
        "[…]"
        ],
    "source":
        [
        "Anuario de estudios filológicos, ISSN 0210-8178, Vol. 26,
2003, pags. 5-11"
        ],
    "title":
        [
        "Eugenio Coseriu: \"In memoriam\""
        ],
    "type":
        [
        "text (article)"
        ]
    },
```

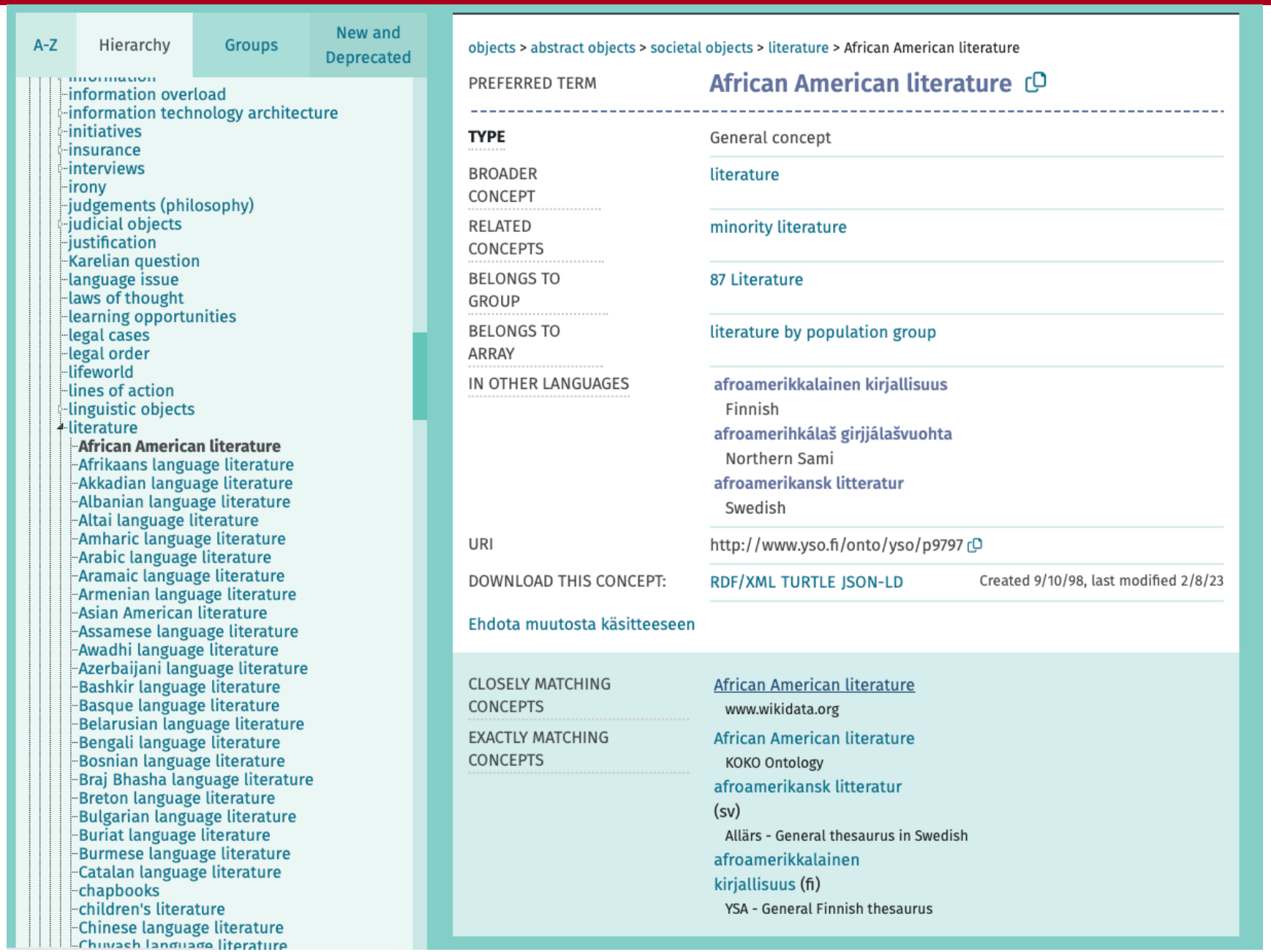
Data Sources

Bibliographical (catalogue) data

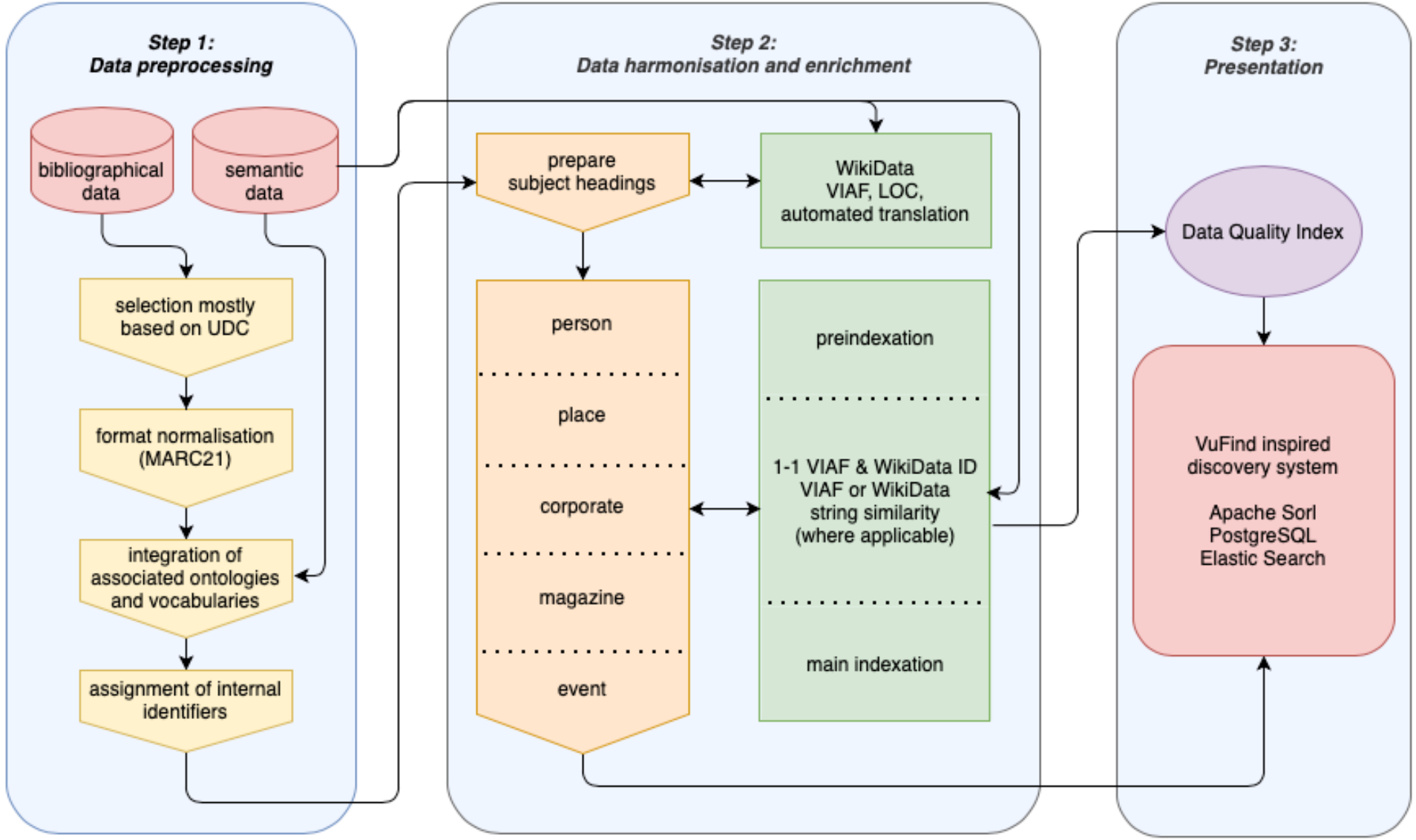
- Spanish National Library
- Finnish National Library
- Czech Literary Bibliography (CLB)
- Polish Literary Bibliography (PBL)
- Bibliography of Scottish Literature in Translation (BOSLIT)

Semantic data

- Library reference ontologies (see example)
- Library of Congress ID (subject headings) (see here)
- Virtual International Authority File (VIAF.org)
- WikiData (wikidata.org)



Finnish ontology (example)



Workflow chart

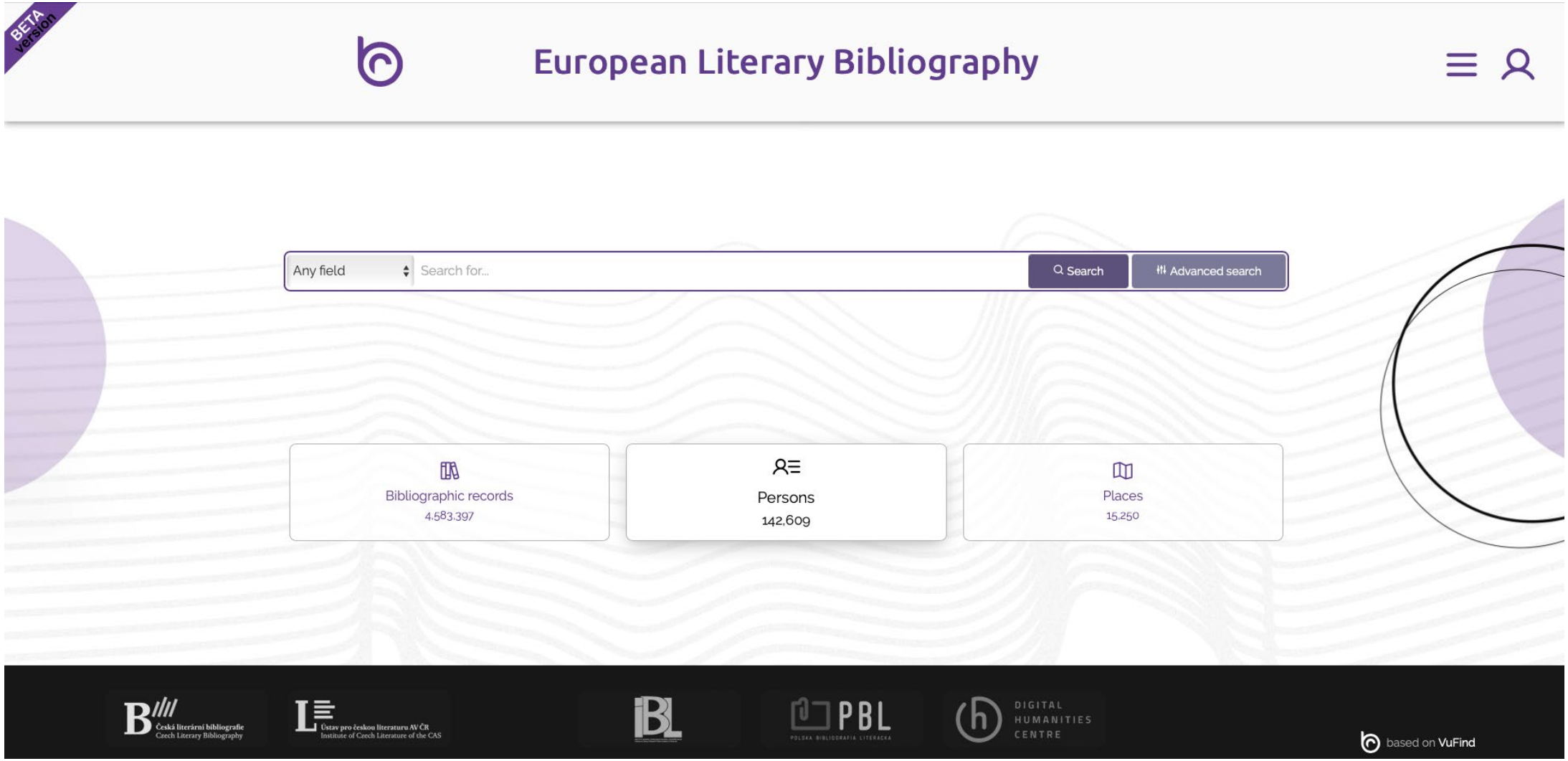
```
LDR 01151nam a2200325 c 4500
001 a6078069
003 SpMaBN
005 20210326123920.6
008 171214s2016\\\\sp\\\\\\\\\\\\000\f\spa\\
015 \\|a MON16
017 \\|a B 15417-2016 |b Oficina Depósito Legal Barcelona
020 \\|a 978-84-01-01893-0
035 \\|a (OCoLC)1055846156
040 \\|a SpMaBN |b spa |c SpMaBN |e rdc
041 1\|a spa |h fin
080 \\|a 821.511.111-31"19"
100 1\|a Waltari, Mika |d (1908-1979) |4 aut |0 XX965743
240 1 |a Sinuhe, egyptiläinen |l Español
245 1 |a Sinuhé, el egipcio |c Mika Waltari ; traducción de Manuel
250 \\|a 1ª ed.
260 \\|a [Barcelona] |b Plaza & Janés |c 2016
300 \\|a 600 p. |c 23 cm
336 \\|a Texto (visual) |2 isbdcontent
337 \\|a sin mediación |2 isbdmedia
380 \\|i Major genre |a Literature |l eng
381 \\|i Major genre |a Fiction |l eng
594 \\|a Título original: Sinuhe, egyptiläinen
655 \7|a Novelas históricas |2 tgfbne
700 1\|a Bosch Barrett, Manuel |d (1895-1961) |e traductor |4 trl
995 \\|a Biblioteca Nacional de España
```

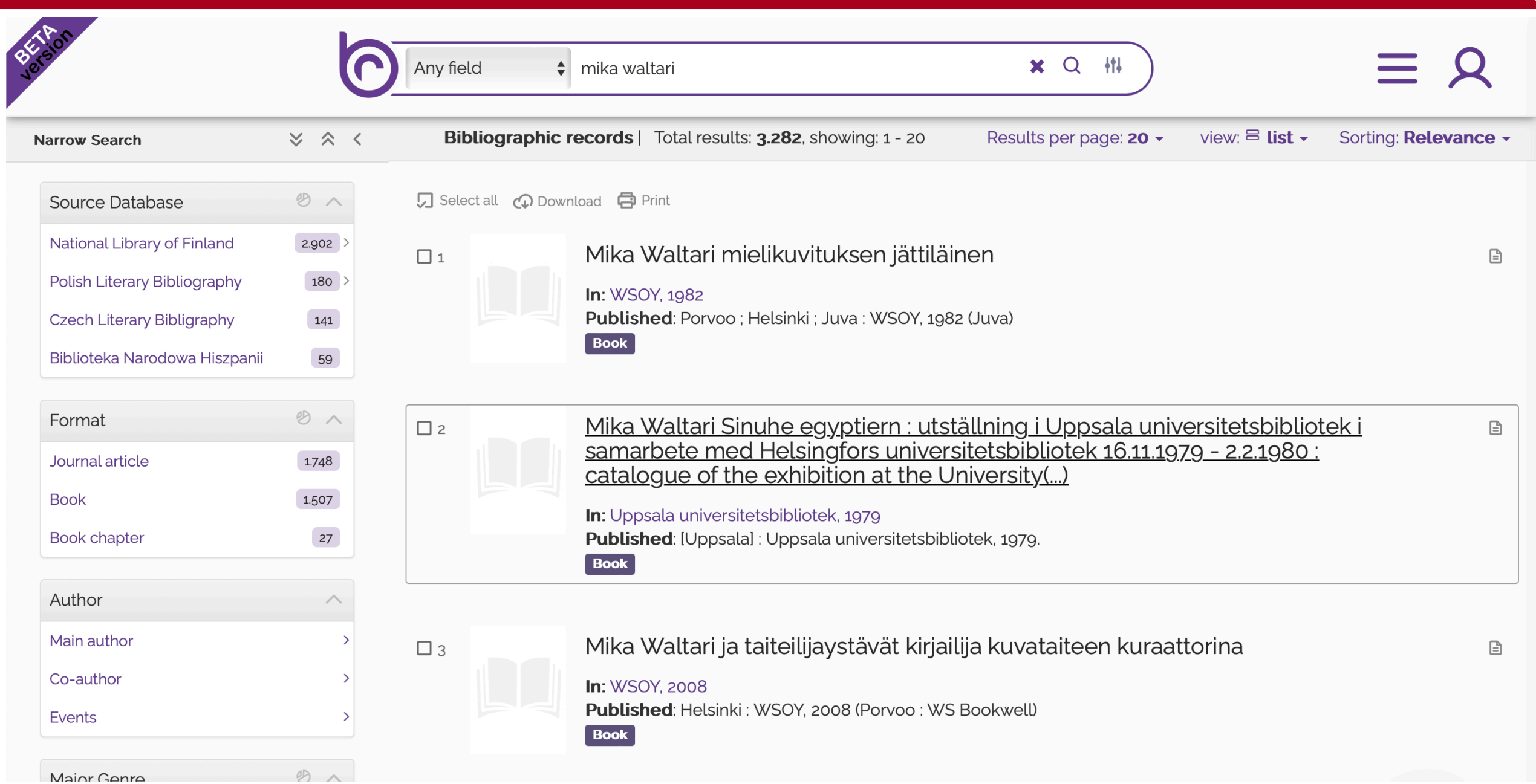

{
- 381: [1 item],
- sourceFile: "es_ksiazki__08-02-2024.mrk",
- editTime: "2024-05-28 16:07:01",
- LEADER: "01151nam a2200325 c 4500",
- majorFormat: "Book",
- rawId: "a6078069",
- id: "es.a6078069",
- language: {3 items},
- isbn: [1 item],
- OCoLC: [1 item],
- ctrlNum: "(OCoLC)1055846156",
- subject: {3 items},
- wikiQ: {
  - Q193111: {
    - mainAuthor: 1
  },
  - Q1492: {
    - publicationPlace: 1
  },
  - Q93427686: {
    - coAuthor: 1
  }
},
- orgin_labels: [3 items],
- persons: {3 items},
- titleShort: "Sinuhé, el egipcio",
- title: "Sinuhé, el egipcio",
- StatmentOfResp: [],
- edition: {2 items},
- places: {2 items},
- corporates: {2 items},
- publicationDateStr: [1 item],
- publicationYear: [1 item],
- publishedIn: [1 item],
- major_genre: [1 item],
- sourceDB: {1 item}

}

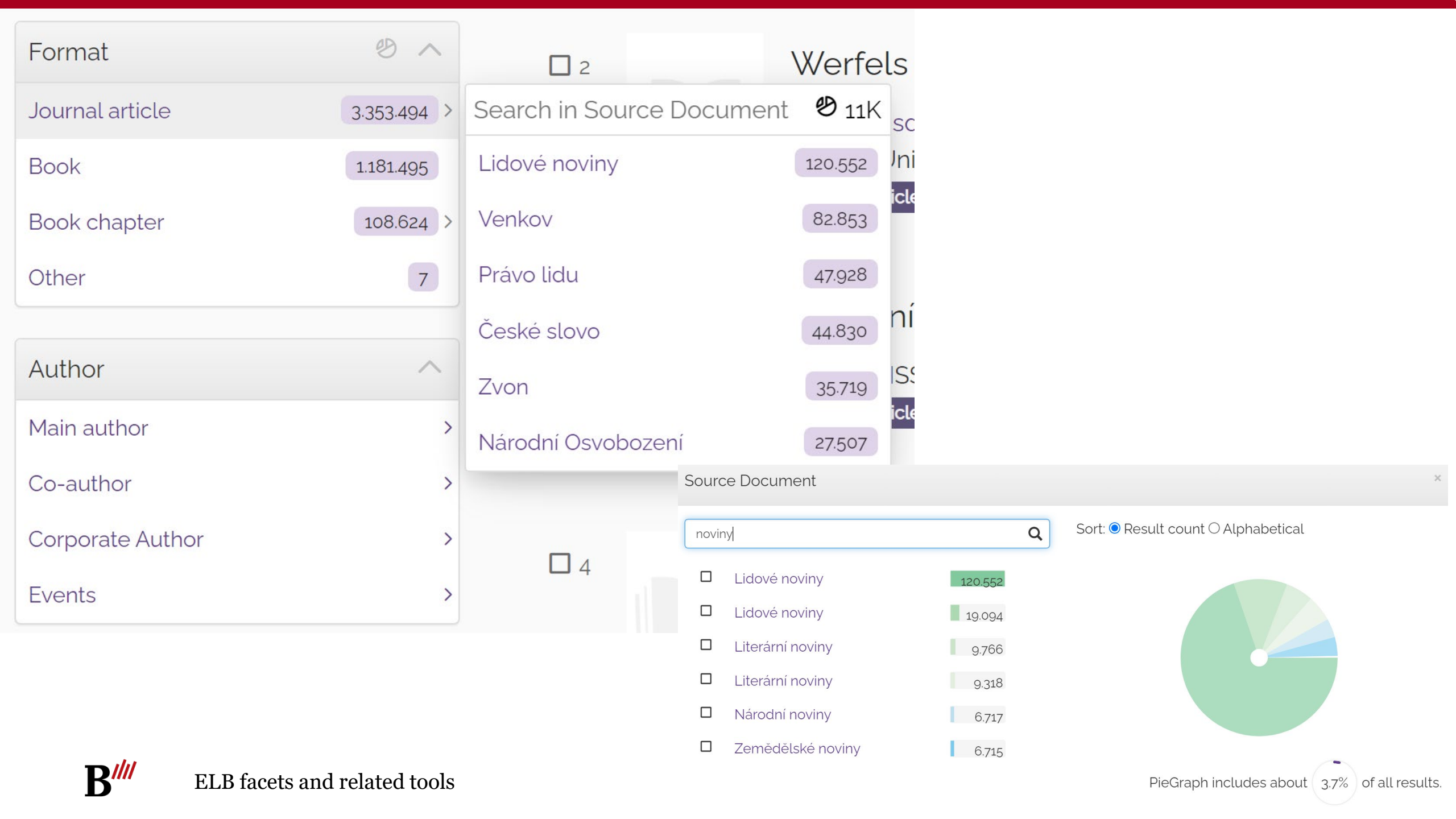
Own data structure (enriched, LODified): [record here]

Discovery system





”Traditional view of bibliographic records” for Mika Waltari.



ELB facets and related tools

Mika Waltari Mika Toimi Waltari, Kristian Korppi, Leo Rainio, Leo Arne, M. Ritvala, Nauticus

00:00 00:00

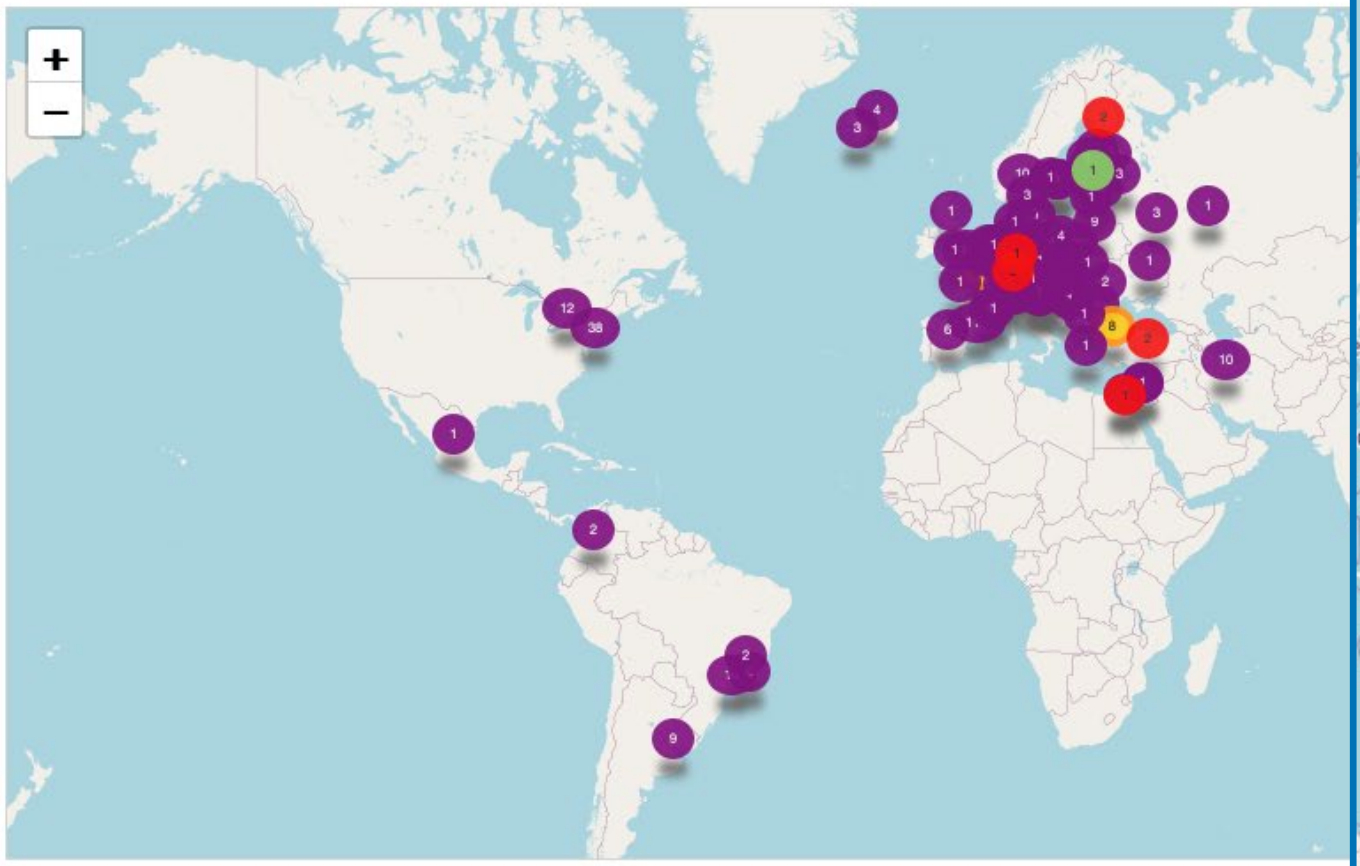
Finnish writer

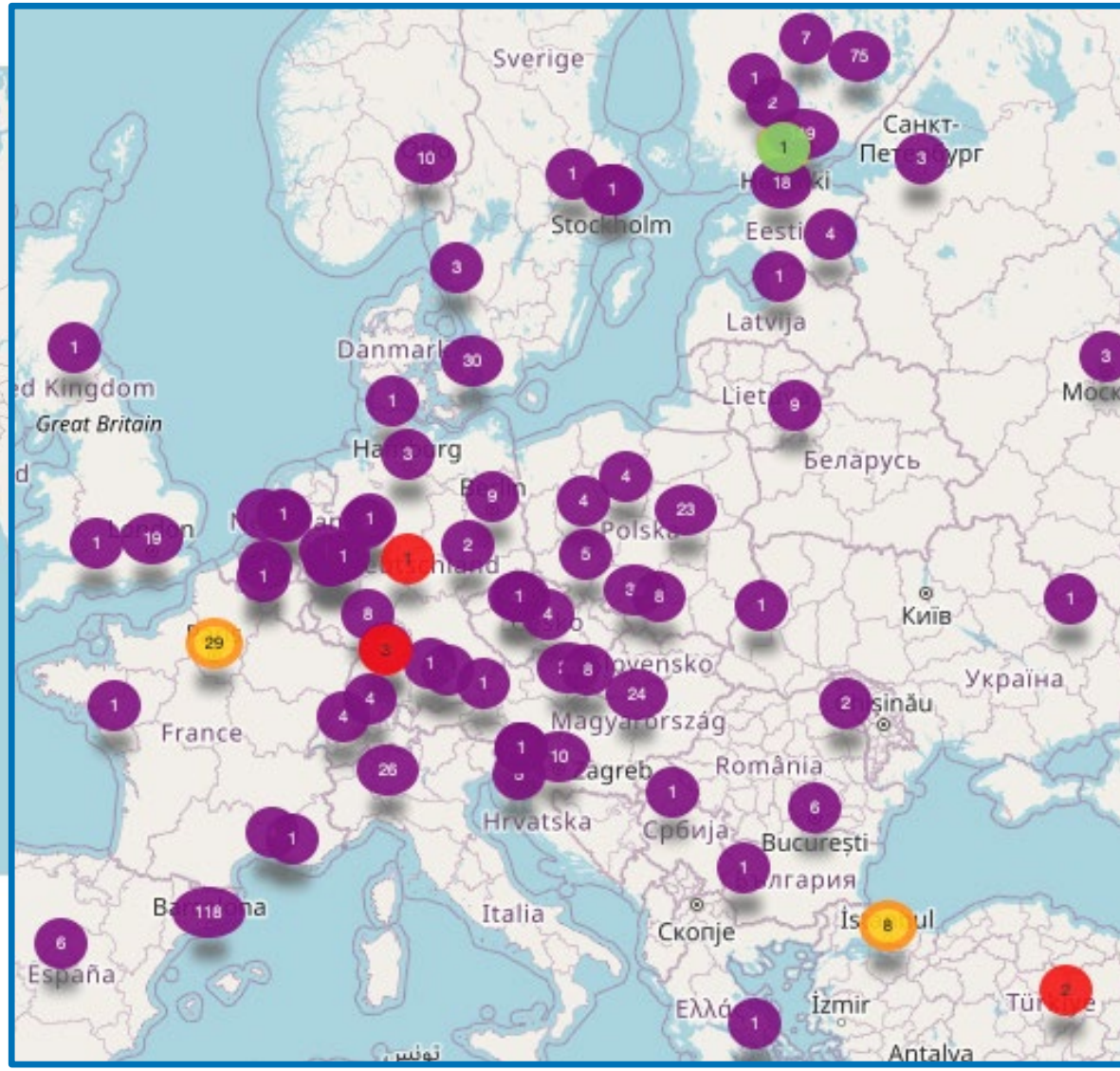
1099 597

Sex or Gender:	male
Date of Birth:	1908-09-19
Place of Birth:	Гельсингфорс (at present: Helsinki)
Date of Death:	1979-08-26
Place of Death:	Хельсинки (at present: Helsinki)
Country of citizenship:	Finland
Childrens:	Satu Waltari
Occupation:	writer translator journalist novelist screenwriter poet playwright
Genres:	historical novel crime novel
Viaf ID:	111008219

Landing page for an individual PERSON (Mika Waltari), upper part.

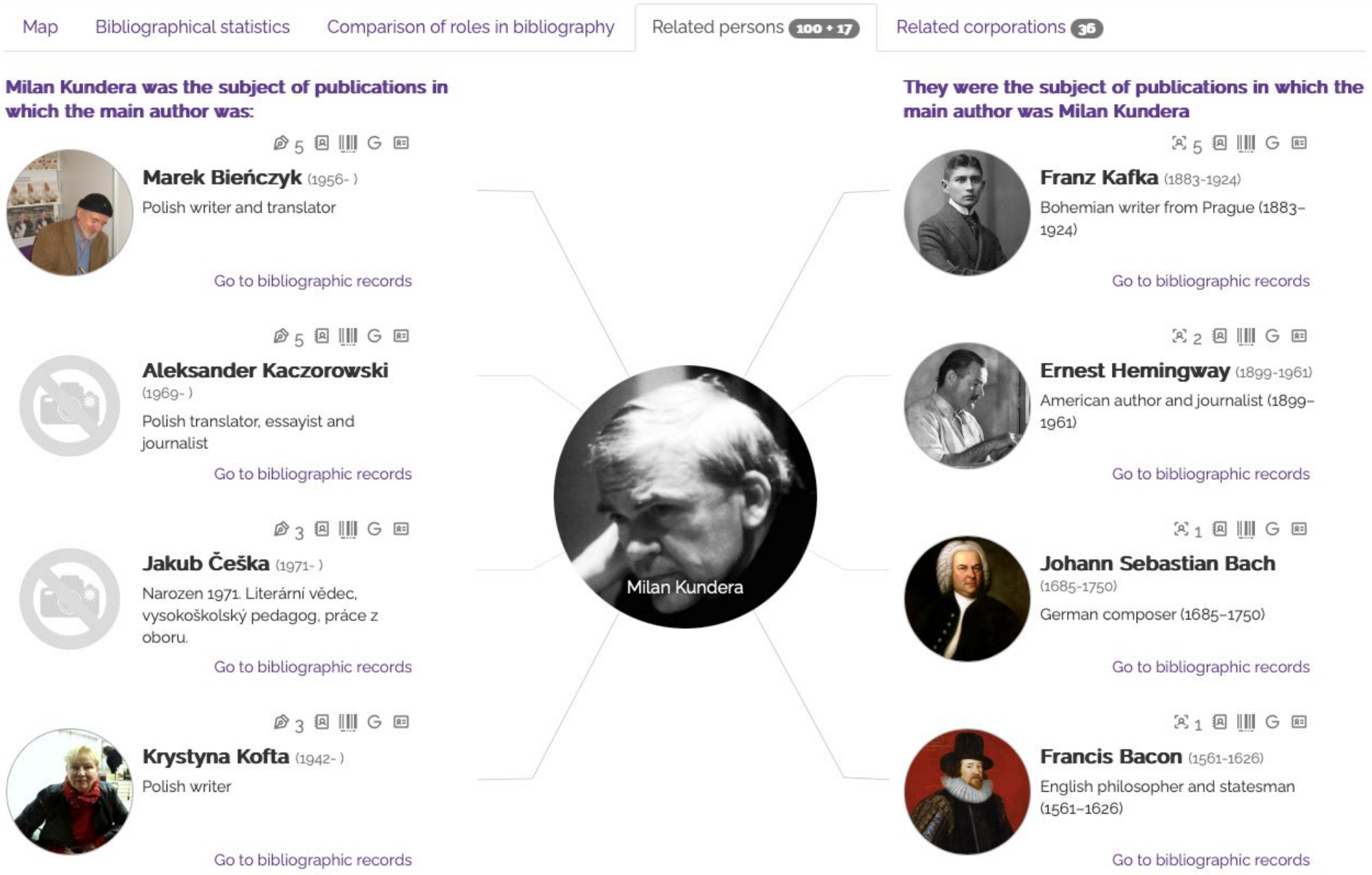
Map | Bibliographical statistics | Comparison of roles in bibliography | Related persons 99 + 4 | Related corporations 37





Display on map:

- Publication places 97
- Subject places 9
- Residence places 3 order may not be correct.
- 3 Mixed relation points

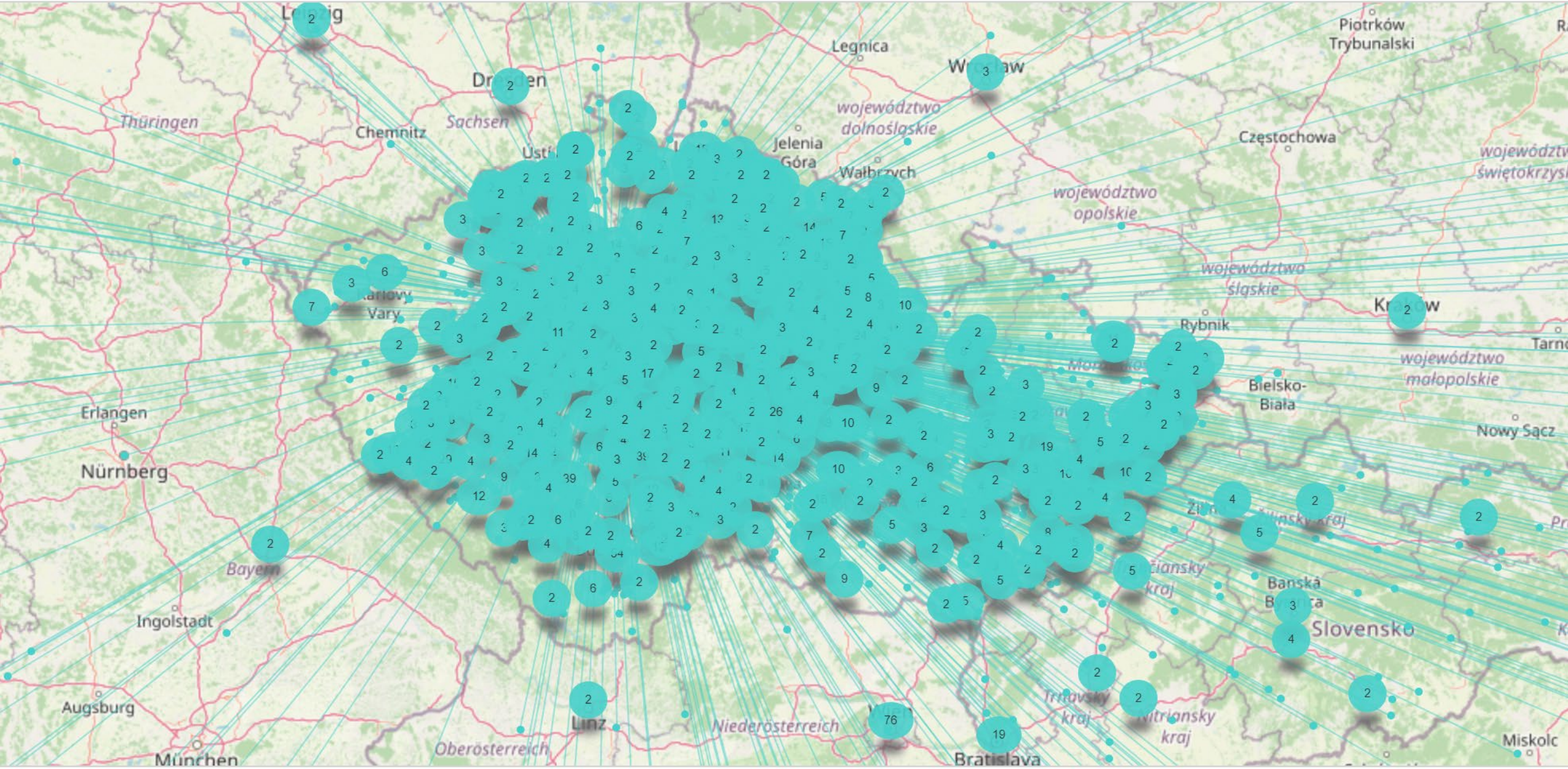


Bibliographical records related to PERSON (Milan Kundera) according to the other's roles.

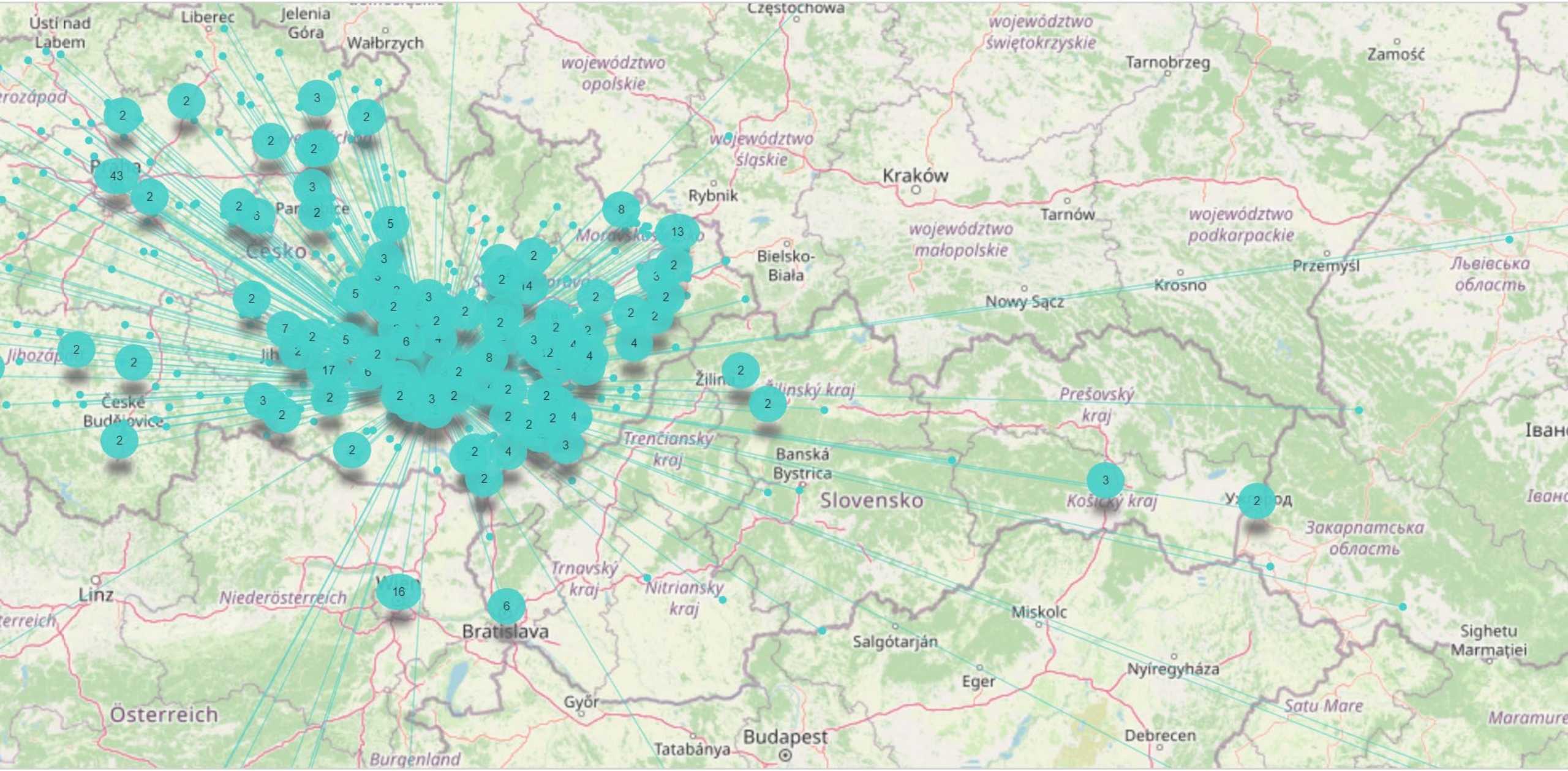
Map | Bibliographical statistics | Related persons 151

- Place of publication of works with subject place Barcelona 233
- Subject place of works publicated in Barcelona 331
- Birthplace of newcomers 24
- Place of death of "emigrants" 18
- **to do**: publication places of works of persons live-related (born/died/residance) with Barcelona 24

Bibliographical records related to PLACE (Barcelona) according to their respective capacity.



Birthplaces of the literary active people who died in [Prague](#).



Birthplaces of the literary active people who died in Brno.

Pre-hamonisation and unification bibliographic record labels linked to this record:

Name (a)	Dates range (d)	Has VIAF	Other ID	Appearances	Certainty of fit
Cervantes Saavedra, Miguel de	(1547-1616)	✓		6.274	
Cervantes Saavedra, Miguel de	(1547-1616)	✓	jn19990001383	430	
Cervantes Saavedra, Miguel de	(1547-1616)	✓		182	
Cervantes Saavedra, Miguel de	(1547-1616).	✓		8	
Cervantes Saavedra, Miguel de	(1547-1616)	✓		3	
Cervantes.				3	
Figura, Michael				3	
we				1	
Cervantes Saavedra, Miguel de	(1547-1616)	✓	aun2006372705	1	

Above you can see a list of all the variants of how this record was written in the bibliographic records.
The **Certainty of fit** column shows how "confident" we are that this label should be here.

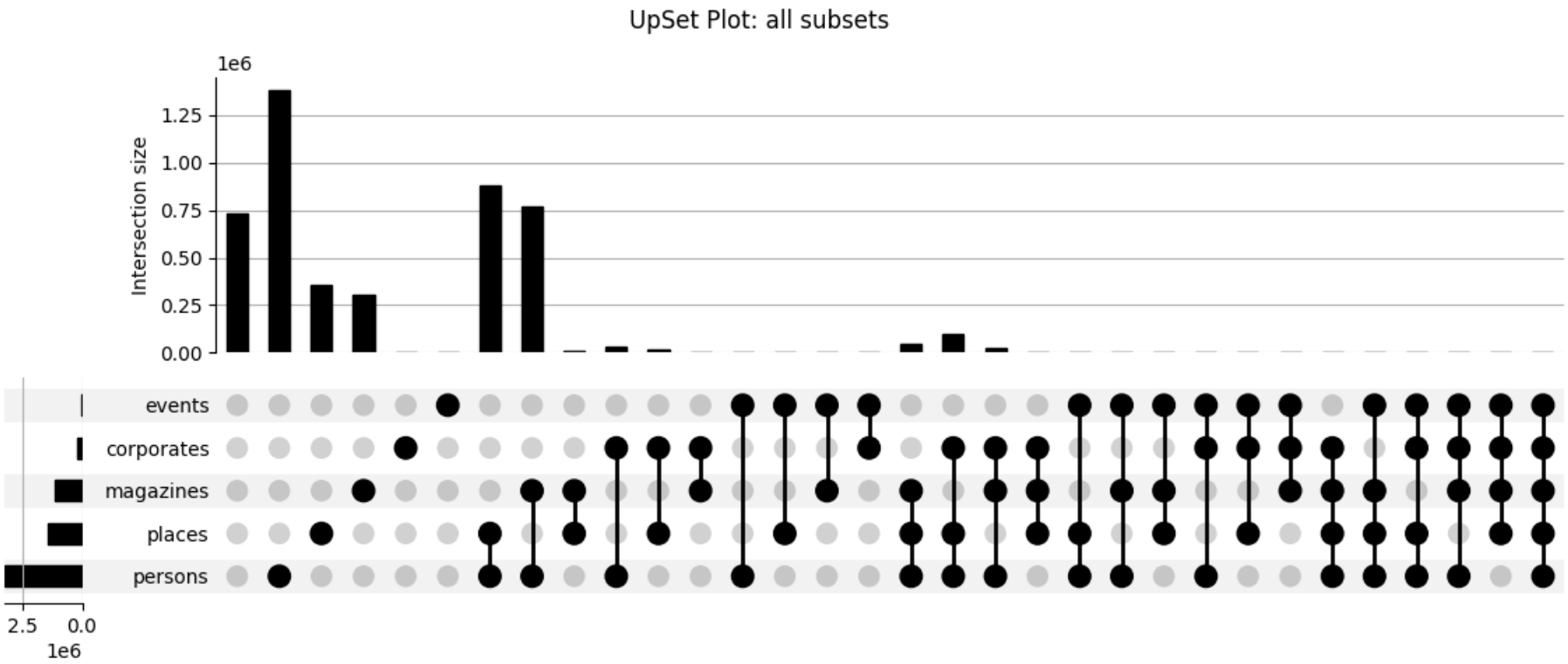
Labels on VIAF:

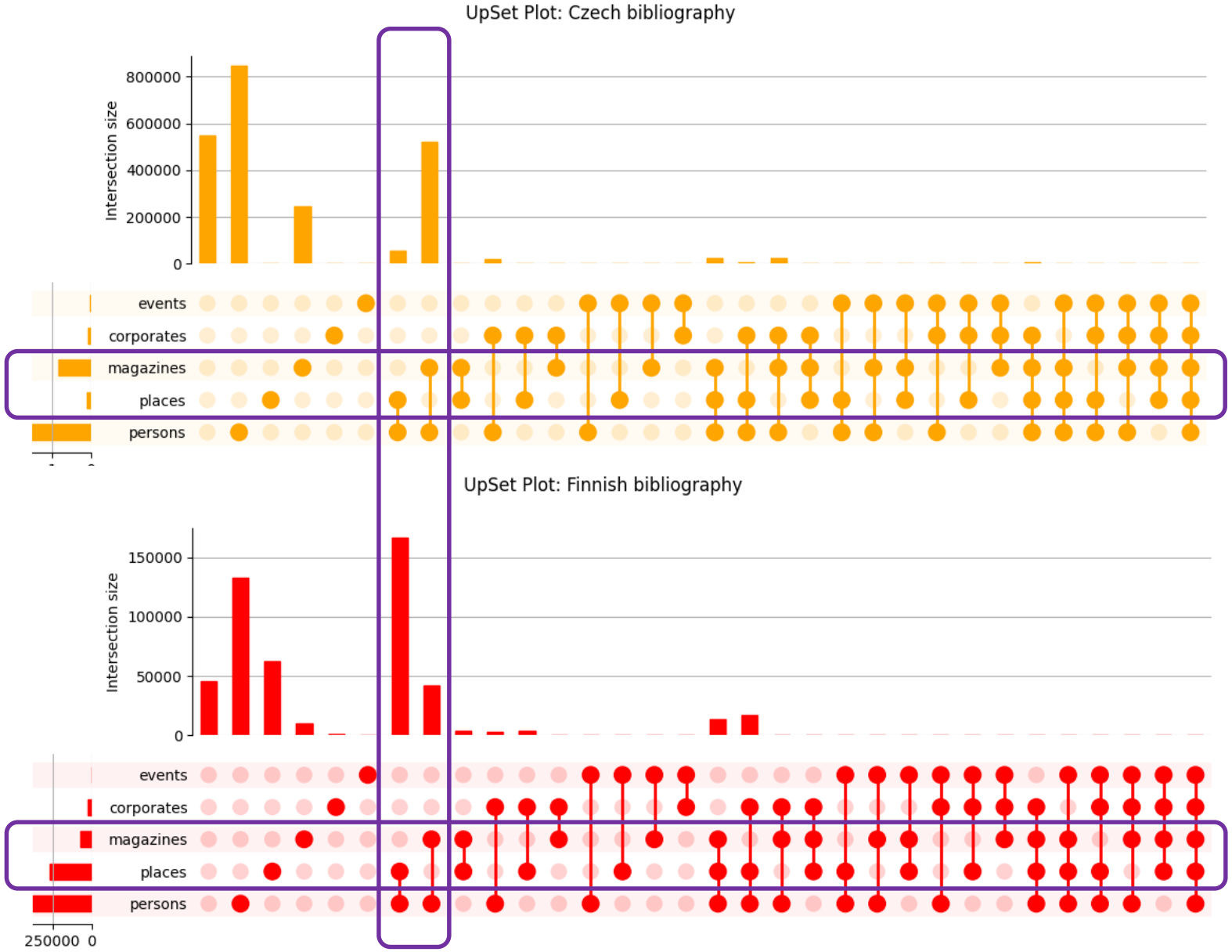
Cervantes Saavedra, Miguel de 1547-1616	31
Cervantes Saavedra, Miguel de	5
Cervantes, Miguel de, 1547-1616	2
Cervantes y Saavedra, Miguel de, 1547-1616	1
Сервантес Сааведра, М. де 1547-1616 Мигель де	1
Cervantes Saavedra, M. de 1547-1616 Miguel de	1
سرفانتس، ميغيل دي، 1547-1616	1
סרונטס, מיגל דה, 1547-1616	1
Сервантес Сааведра, Мигель Де, 1547-1616	1
Miguel de Cervantes Saavedra	1
de Cervantes Saavedra, Miguel	1
Cervantes Saavedra, M. de	1
ثربانتس سابدرا، ميجل دى، 1547-1616	1
Cervantes de Saavedra, Miguel, 1547-1616	1

Data identification and links for **4,611,498** bibliographical entries

category	records with data	unique items	links total
Person	3,745,450 **(81%)**	279,259	6,381,219
Place	1,440,803 **(31%)**	8,787	1,784,478
Corporate	1,376,301 **(30%)**	7,206	1,598,087
Magazine	2,508,598 **(54%)**	882	2,667,997
Event	62,007 **(1%)**	314	56,796

Intersections of sets and their sizes




UpSet Plot: Czech bibliography
Intersection size
800000
600000
400000
200000
0
events
corporates
magazines
places
persons
UpSet Plot: Finnish bibliography
Intersection size
150000
100000
50000
0
events
corporates
magazines
places
persons
250000
0

Issues and Challenges

- Limits of non-interventionist approaches
  - Inclusion of more datasets
  - Definition of literature (primary & secondary)
  - Curate or not curate
  - Live data and updates (original datasets, wiki, VIAF)
- Technological sustainability
- Data format backward compatibility
- Wikidata corruption (including hacking)

Helsinki Helsingfors, Helsingia, Helsingin kaupunki, Stadi, Hesa

capital and most populous city of Finland

Country:	Finland
Coordinates (WGS-84):	60,17556,24,93417
Country:	at present: Finland 1918-05-16: Finnish Socialist Workers' Republic 1917-12-06: Russian Empire 1809-09-05: Sweden 1550
Historical names:	at present: ru Хельсинки 1917-12-05: ru Гельсингфорс long long time ago

Source of information Wikidata

More information on Wikipedia

Bibliographical records related to PLACE (Helsinki) demonstrating corrupt data (historical place name).

Permalink: http://viaf.org/viaf/136958681

Preferred Forms

- 110 2 _ ‡a Charta 77
- 110 2 _ ‡a Charta 77
- 110 2 _ ‡a Charta 77 (Group)
- 110 2 _ ‡a Charta 77 (Group)
- 110 2 _ ‡a Charta 77 (Group)
- 110 2 _ ‡a Charta 77 (Group)
- 110 2 _ ‡a Charta 77 (Prague) (1977-1989)
- 110 2 _ ‡a Charta 77 (musikgrupp)
- 110 2 _ ‡a Charta 77 (skupina)
- 110 2 _ ‡a Charta 77 (skupina)
- 110 2 _ ‡a Charta 77 Foundation (Prague)
- 210 | | ‡a Charta 77 ‡c Prague
- 110 2 _ ‡a Charta 77 ‡g Körperschaft

Bibliographical records related to CORPORATION (Charta 77) demonstrating corrupt data (two different entities).

Literarybibliography.eu 2.0

- Under preparation
- New modules and tools
- Works on backend -> deduplications, data enrichment
- Licences
- New datasets:
  - Czech National Bibliography (books)
  - Bibliography of Scottish Literature in Translation

European Literary Bibliography
Search in bibliografic records:
Any field
Search for...
Search
Advanced search
Bibliographic records
4 643 833
Persons
199 275
Places
8 268
Corporates
7 392
Magazines
847
Events
310

Sources

The Czech Literary Bibliography,

created at Institute of Czech Literature of the Czech Academy of Sciences, a public research institution, is a basic research infrastructure for literary Czech studies and related humanities (e.g. philology, philosophy, history, art history and journalism). Its objective is to process, continuously edit and update and place online a unique set of bibliographical and other knowledge databases, seamlessly mapping out literary events and literary studies in the Czech lands and the circulation of Czech literature abroad from the beginning of the National Revival to the present, regardless of language, format or medium. Currently, Czech Literary Bibliography databases contain records on 2,250,000+ articles on literature and literary life in the Czech lands since 1770 up to the present days.

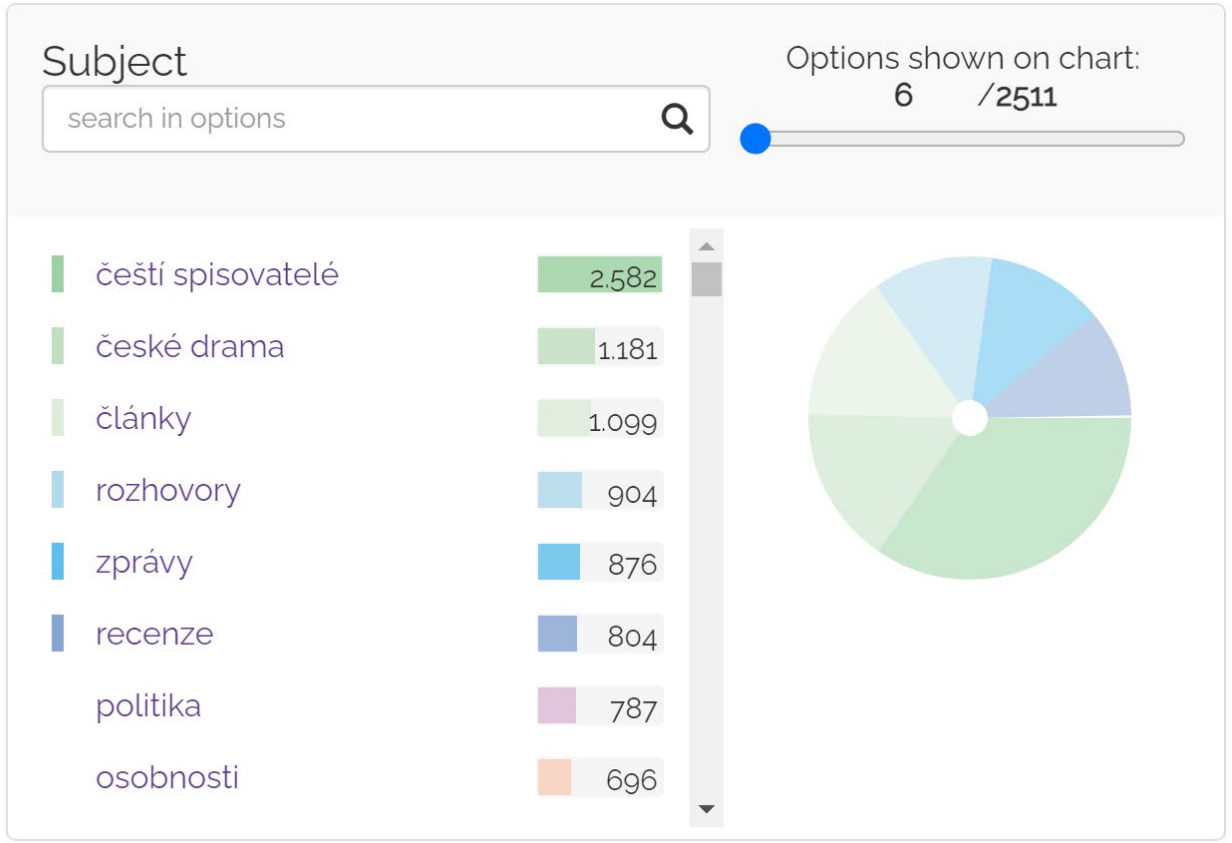
cz_articles0_08-02-2024.mrk
cz_articles1_08-02-2024.mrk
cz_articles2_08-02-2024.mrk
cz_articles3_08-02-2024.mrk
cz_articles4_08-02-2024.mrk
cz_books__08-02-2024.mrk
cz_chapters__08-02-2024.mrk
cz_ucla_2024-02-20update.mrk

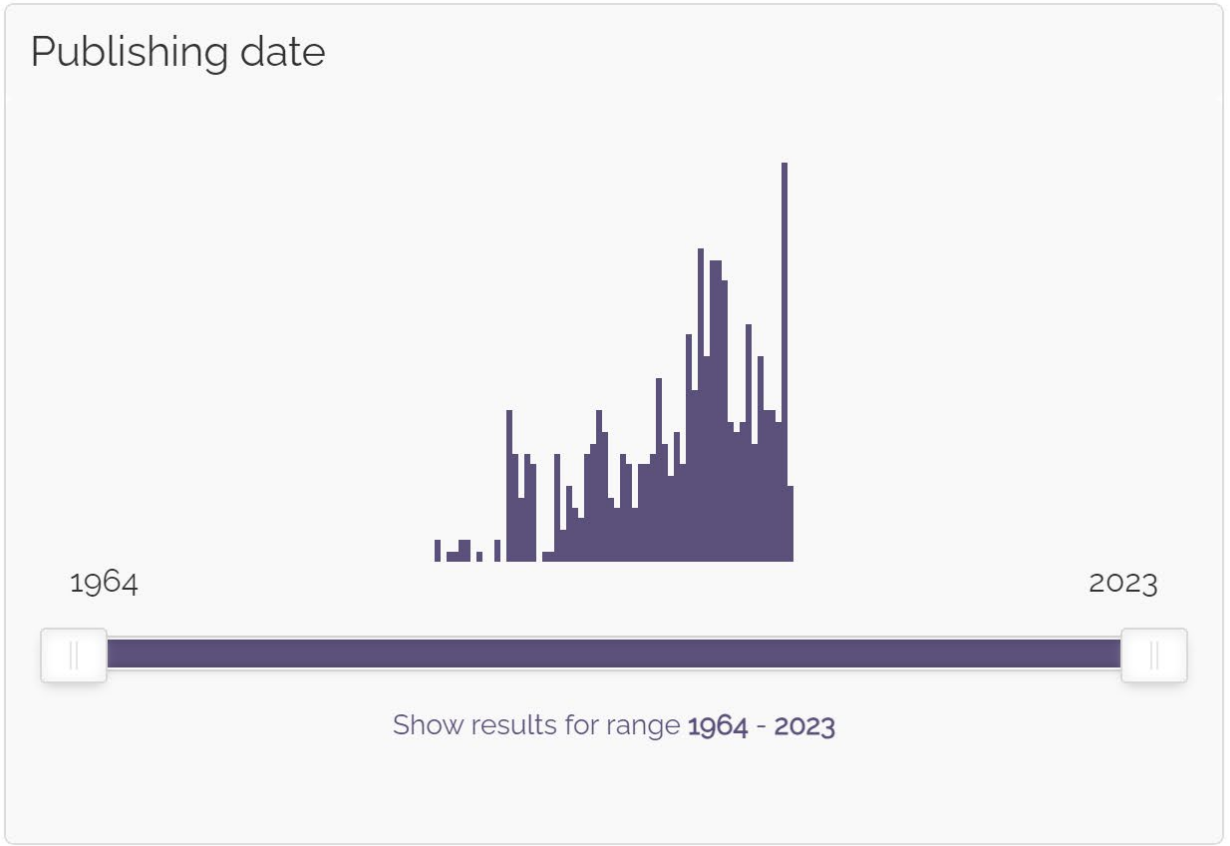
Licence: CC BY-NC-SA 4.0 When using the CLB databases, you agree to acknowledge the use of the CLB resources in all your work. Example: "During the work on this research paper [book/presentation/text...] use was made of the Czech Literary Bibliography research infrastructure – https://clb.ucl.cas.cz/ (ORJ identifier: 90243)."

Polish Literary Bibliography

is created and compiled at the Institute of Literary Research of the Polish Academy of Sciences, a public research institution, that serves as a research infrastructure for literary studies and related humanities (e.g. philology, philosophy, history, art history and journalism). Its main objective is the curation of Polish literary bibliographical data. It registers, processes and updates a unique bibliography containing books published in Poland and abroad in Polish. For each year-book, PBL compiles approximately 850 periodical titles: national dailies, national literary, cultural, cultural and social periodicals, serials published by

pl_bn_articles2_08-02-2024.mrk
pl_bn_articles_08-02-2024.mrk
pl_bn_books2_08-02-2024.mrk
pl_bn_books_08-02-2024.mrk
pl_bn_chapters_08-02-2024.mrk
pl_bn_chapters__08-02-2024.mrk
pl_pbl_articles_08-02-2024.mrk
pl_pbl_books_08-02-2024.mrk





Updated ELB analytical module.

Map | Bibliographical statistics | Comparison of roles in bibliography | Related... | Data quality

Relation were created on the basis of entries in bibliographic records. Number of records in which the Václav Havel was found: **7610**.

Václav Havel appears in roles:

subjectPerson	6.917
mainAuthor	701
coAuthor	309

Show relations with:

Persons	37.010
sourceMagazine	3.265
Corporates	2.085
Places	1.738
Events	108

Persons appears in roles:

subjectPerson	17.431
mainAuthor	4.732
coAuthor	6.920

Total results **6225** showing **18**

4.9K 591 105

Milan Kundera

Franco-Czech poet and novelist (1929-2023)

Show bibliograhic record with Václav Havel and with Milan Kundera as

subjectPerson 365 | mainAuthor 5 | coAuthor 15

2.2K 818 211

Ludvík Vaculík

Czech publicist, politic writer and bookwriter (1926-2015)

subjectPerson 263 | mainAuthor 15 | coAuthor 88

2.2K 657 135

Pavel Kohout

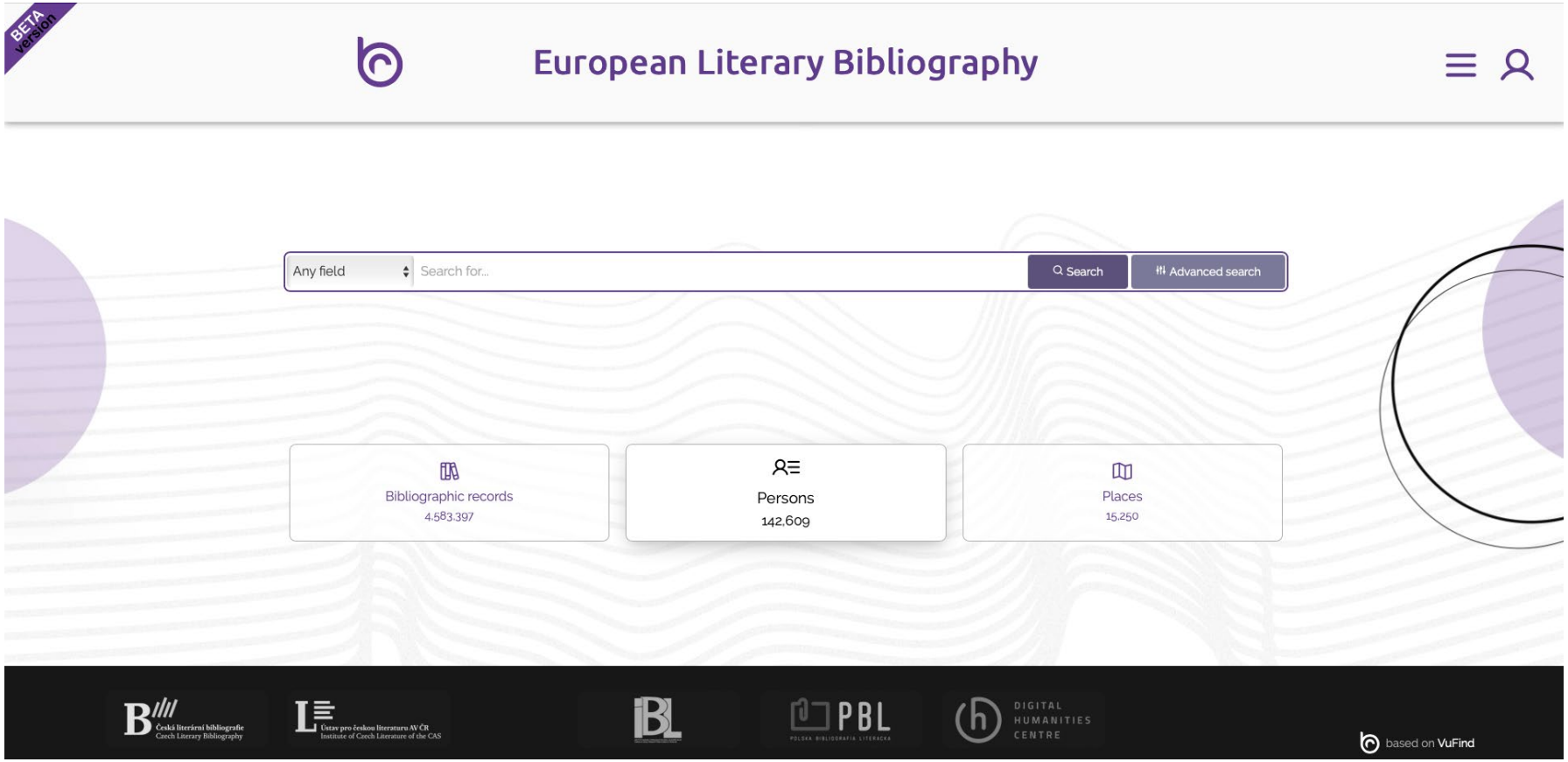
Czech poet, playwright, exil writer, film director, politic writer and bookwriter

subjectPerson 216 | mainAuthor 29 | coAuthor 57

Marc view | ELB fields | Map | Other versions


```
{
  sourceFile: "es_ksiazki__08-02-2024.mrk",
  sourceDB: {
    licence: {
      name: "CC0",
      link: "https://creativecommons.org/publicdomain/zero/1.0/legalcode",
      description: ""
    },
    name: "Biblioteca Nacional de España"
  },
  editTime: "2024-09-23 15:12:32",
  LEADER: "00949nam a2200277   4500",
  majorFormat: "Book",
  rawId: "bimo0000964468",
  prefix: "es",
  id: "es.bimo0000964468",
  language: {3 items},
  isbn: [1 item],
  OCoLC: [1 item],
  ctrlNum: "(OCoLC)431558018",
  subject: {1 item},
  wikiQ: {
    Q905: {
      mainAuthor: 1
    },
    Q1492: {
      publicationPlace: 1
    }
  },
  places: {
    publicationPlace: {
      Q270: {
        name: "Warszawa",
        ids: [],
        biblio_label: "Warszawa||",
        wikiQ: "Q270",
        nameML: "Warsaw|Varšava|Warszawa|Varsova|Varsovia"
      }
    }
  },
  mainAuthor: {
    Q868: {
      wikiQ: "Q868",
      viaf: "7524651",
      year_death: "",
      year_born: "",
      dates: "",
      nameML: "Aristotle|Aristotelés|Arystoteles|Aristoteles|Aristóteles",
      name: "Aristóteles",
      ids: {1 item},
      role: "mainAuthor",
      roles: [],
      biblio_label: "Aristóteles||7524651|"
```

Please test, provide feedback and join: https://literarybibliography.eu/



Mgr. Ondřej Vimr, Ph.D.
vimr@ucl.cas.cz

Mgr. Vojtěch Malínek, Ph.D.
malinek@ucl.cas.cz

Česká literární bibliografie
Czech Literary Bibliography

Czech Academy of Sciences, Institute of Czech Literature
Na Florenci 1420/3, 110 00 Praha 1, Czech Republic
T +420 222 828 132, +420 770 136 131 E clb@ucl.cas.cz W clb.ucl.cas.cz

The Czech Literary Bibliography Research Infrastructure is fully supported by the Ministry of Education, Youth and Sports of the Czech Republic under the large research infrastructures programme (project LM2023043).