



DIGITAL
HUMANITIES
CENTRE



Research funded by the National
Programme for the Development of
Humanities



NARODOWY PROGRAM
ROZWOJU HUMANISTYKI



Maciej Maryl, Piotr Wciślik
*Turning the Polish Literary
Bibliography into a
Research Tool: Challenges,
Standards, Interoperability*

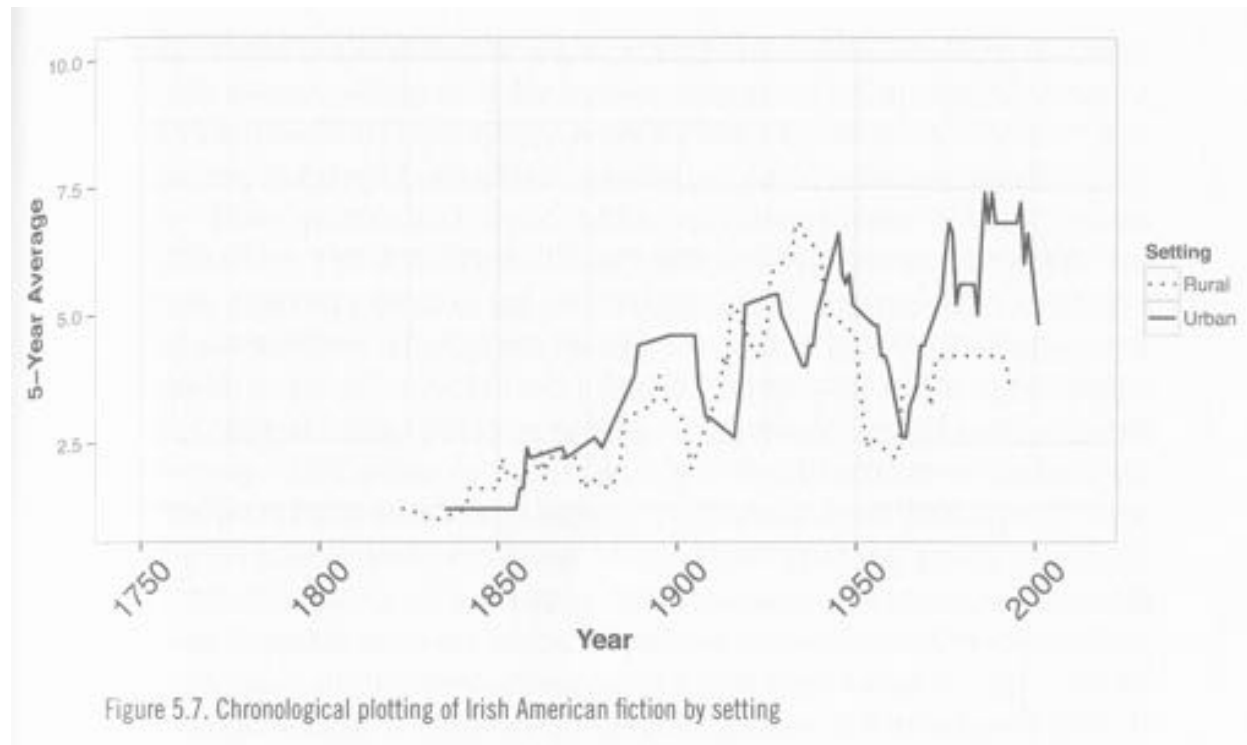
Research based on bibliographical records



Jockers in: Shea 2008.



Research based on bibliographical records



Jockers 2013, 45.



Research based on bibliographical records



Figure 19. Australian novels by gender of author, percentages, 1945 to 2009 (five-yearly averages)

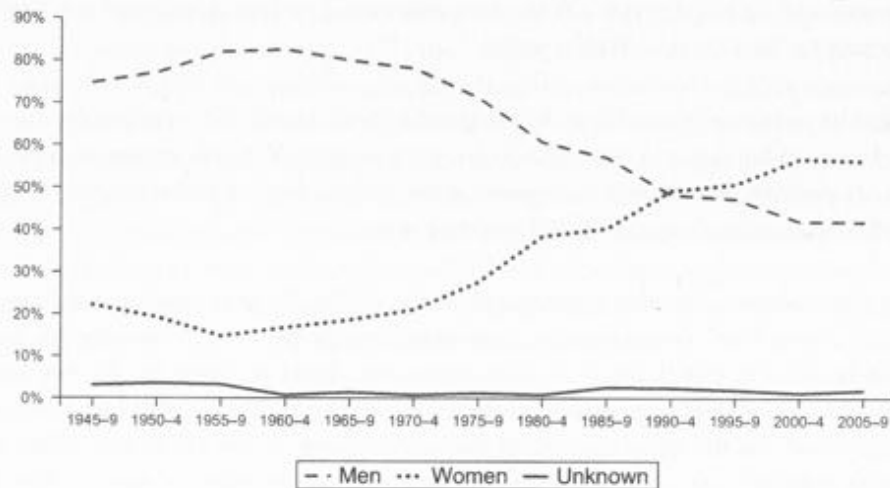
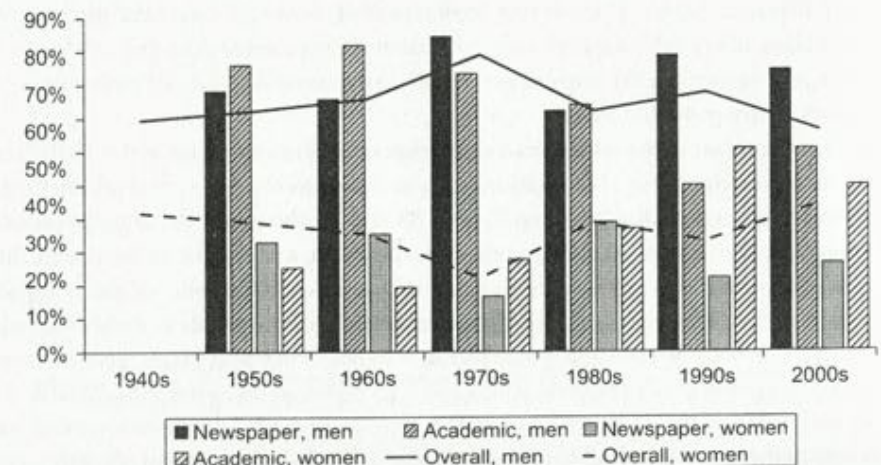


Figure 22. Men and women in the top twenty most critically discussed Australian authors, percentages, 1945 to 2006 (by decade)



Bode 2014, 132 i 134



Research based on bibliographical records

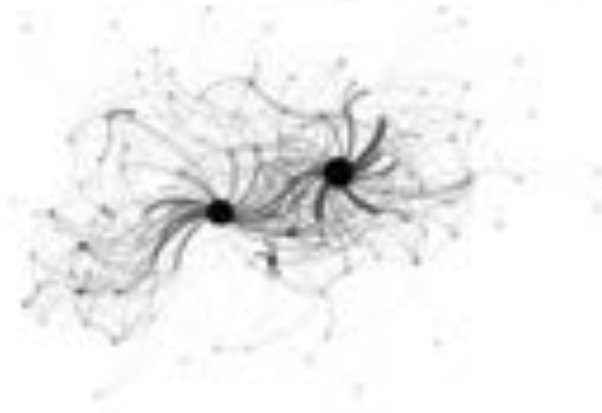
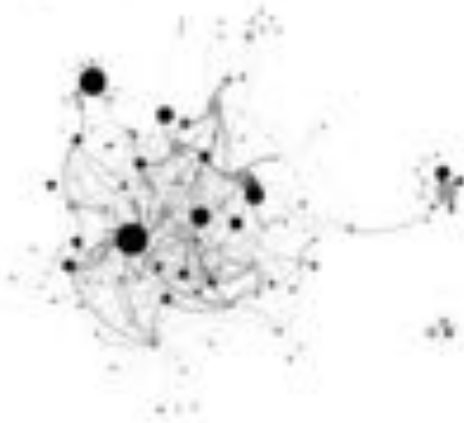


Long and So 2013a, 150.

Research based on bibliographical records



Fig. 1 Journal-Poet affiliation networks for the U.S. (1918-1922), Japan (1928-1931), and China (1915-1945). (© Hoyt Long)



274

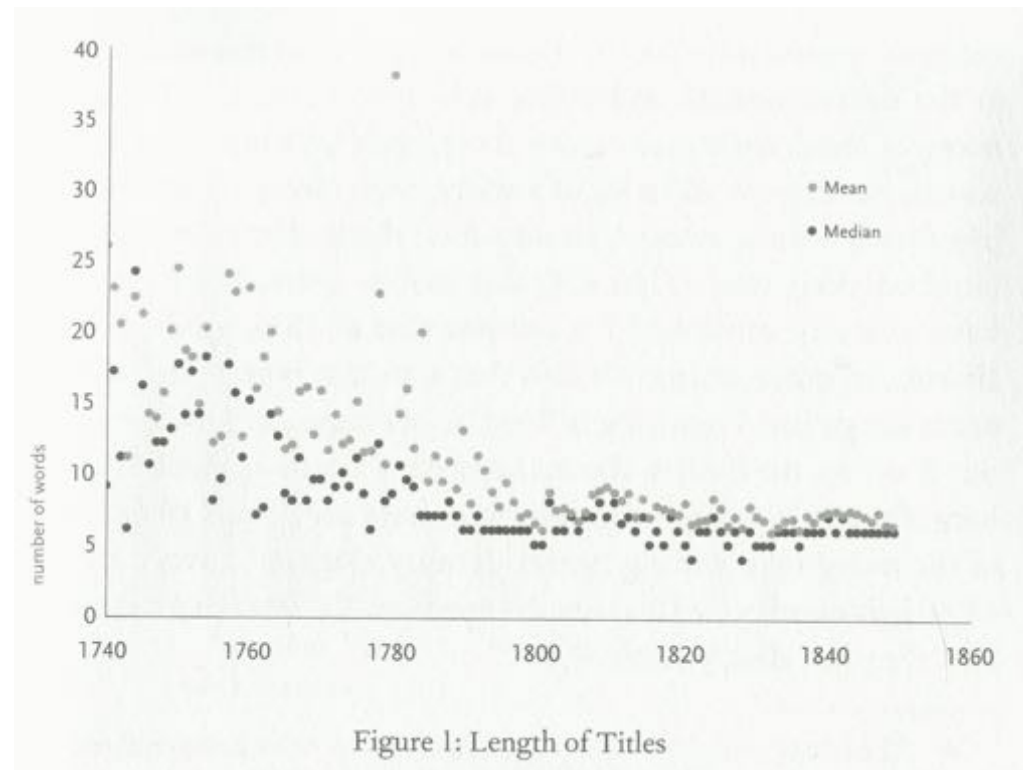
LEONARDO, Vol. 46, No. 3, p. 274, 2013

doi:10.1142/LEON_a_10579 © 2013 IASST

Long and So 2013b, 274.



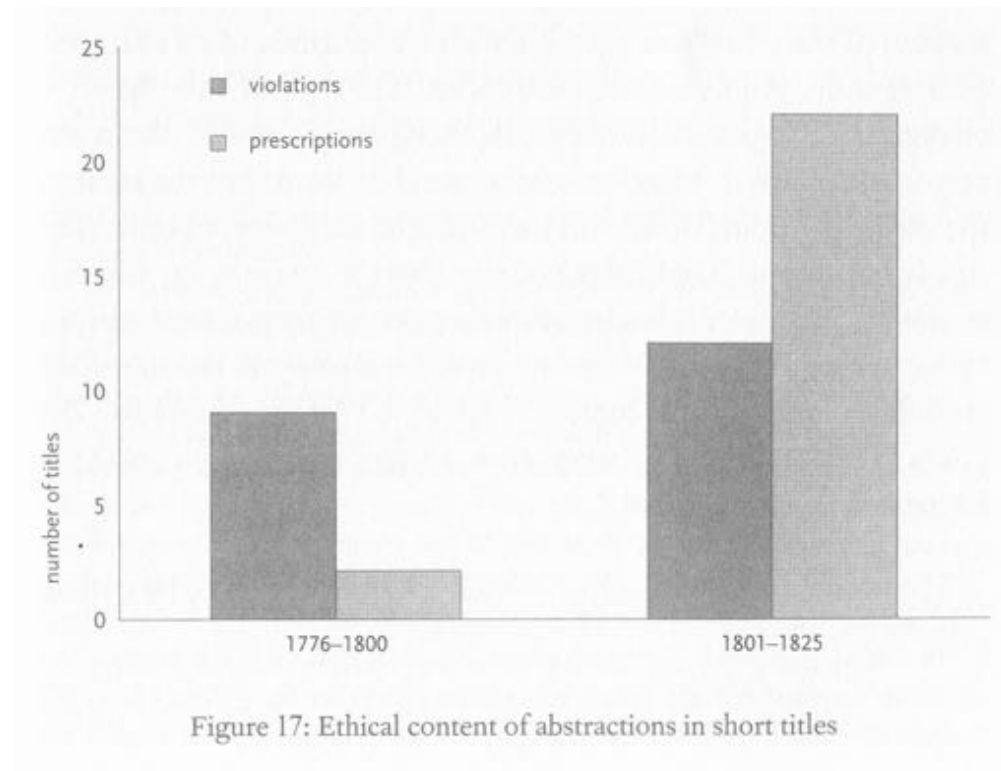
Research based on bibliographical records



Moretti 2013, 183.



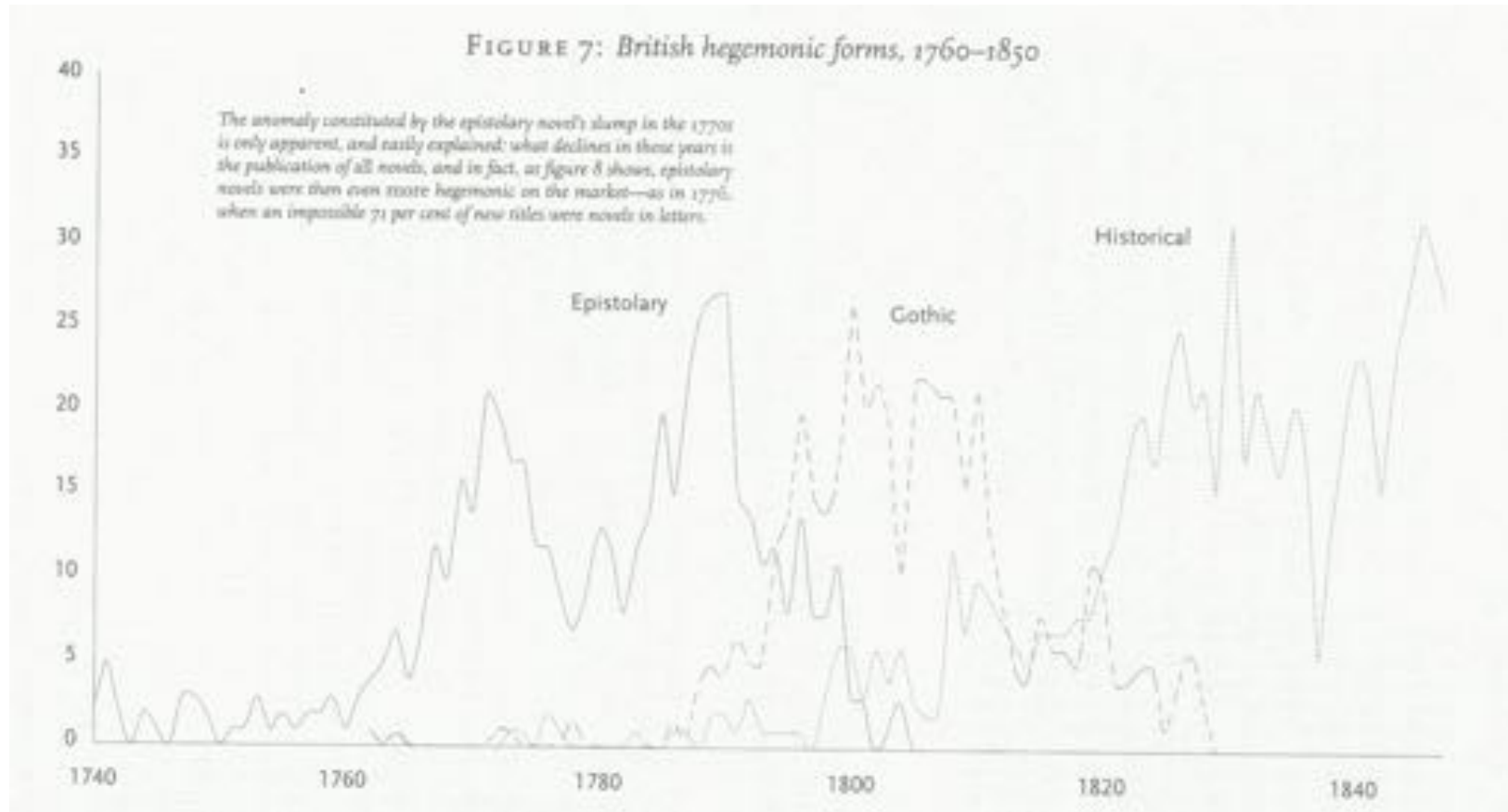
Research based on bibliographical records



Moretti 2013, 201.



Research based on bibliographical records



Moretti 2005, 15.



Research based on bibliographical records



Moretti 2005, 19.

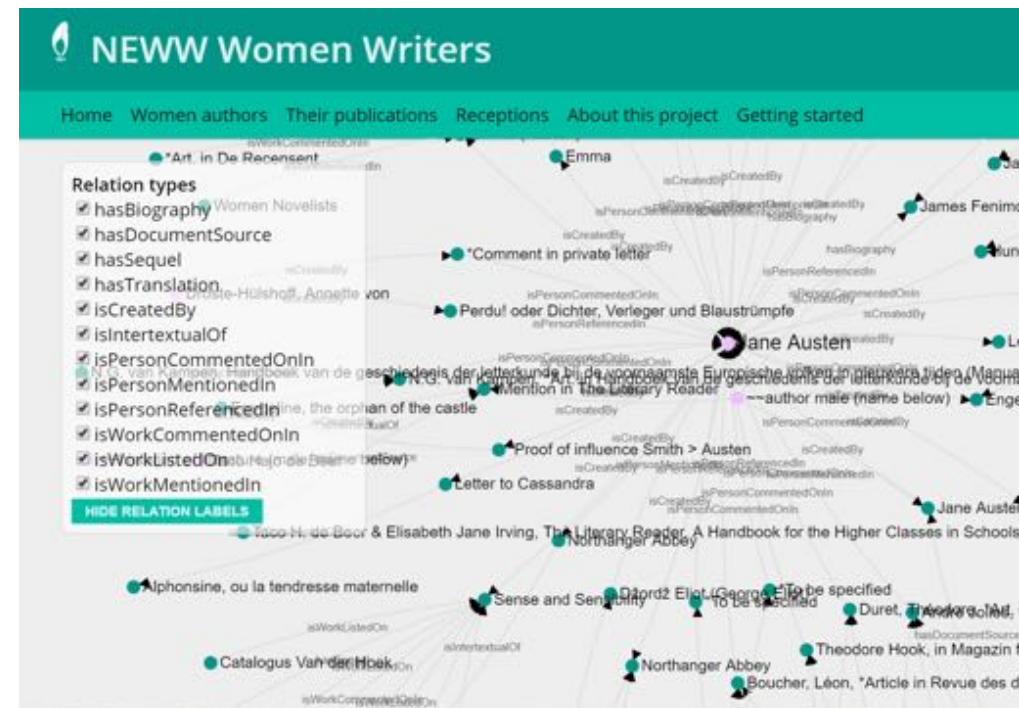


Research based on bibliographical records



- Online bibliographies with research components:

- AustLit
- NWBIB
- Orlando
- NEWW (Women Writers VRE)



Towards literary "Big Data"



- **Big and dirty or small and clean?**
- **Better data → better answers**
- **Bigger data → bigger problems**
- **Working with data collected as a reference resource, not research material**
- **Goal: forging a database ready to answer unpredicted questions**



Polish Literary Bibliography (PBL)

- **Annotated bibliography of Polish literature (since 1954)**
- **Register of all articles, notes and other materials concerning literary life**
- **Bibliographical records concerning Polish Literature:**
 - **Authors**
 - **Works**
 - **Reviews**
 - **Articles**
 - **Contests and awards**
 - **TV, radio, theatre and cinema adaptations**
 - **Events**
 - **Exhibitions**
 - **Semantic annotations (e.g. 'literary theory'; 'exile literature'; 'psychology'; 'education'; 'censorship')**

- **Project:**

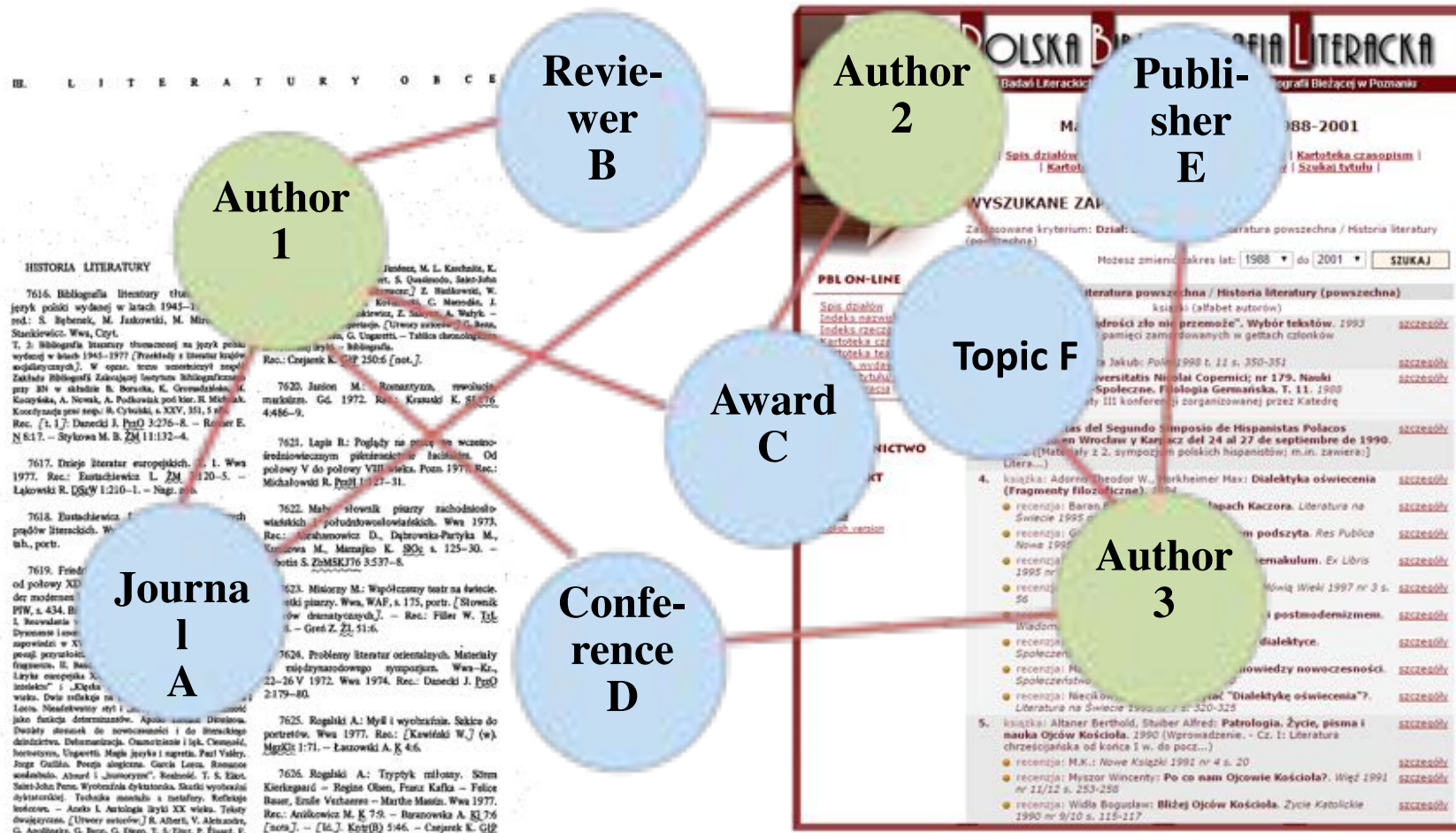
- Polish Literary Bibliography as a research tool on contemporary Polish culture
- 3-year project funded by the National Programme for the Development of Humanities (2015-2018)
- ICT partner: PSNC (Poznań Supercomputing and Networking Center)

- **Goals:**

- Creating a database of ~4 million bibliographical records on Polish literature (1939-2004)
- Integration of heterogenous datasets
- Integration with Linked Open Data Cloud
- Data mining and visualisation tools for modeling processes of literary life



Remediating bibliography



How to get there?



- **Choosing the right data standard**
- **Remediating the input tool (and habits)**
- **Converting the existing online database**
- **Retrodigitising printed volumes**
- **Integration with other bibliographies**



Choosing the data standard



Schema.org INSTEAD of bib-dedicated ontologies and data models (FRBR, RDA, BIBFRAME)

BIB-DEDICATED ONTOLOGIES

- **Not well equipped to handle theatre, cinematographic, radio and television instances of literary works (except FRBRoo)**
- **Too complex to handle by metadata producers (including FRBRoo)**
- **Validity still TBC by community of practice**

SCHEMA.ORG

- **The internet standard**
- **Robust enough to handle PBL data model**
- **Rendering Work/Expression/Manifestation somewhat difficult, but doable**
- **Not unprecedented in the library domain**



Schema.org: standard vocabulary for various encodings



the legend of tarzan

All Videos Images News Shopping More + Search tools

About 21,500,000 results (0.48 seconds)

Showtimes for The Legend of Tarzan

All times are in CT

Thu, Jul 7 **Fri, Jul 8** Sat, Jul 9 Sun, Jul 10

All times Morning Afternoon Evening Night

Muvico Rosemont 18 and MuvuXL - Map

Standard	11:10am	2:00pm	4:45pm	7:30pm	10:15pm
3D	11:55am	2:40pm	5:25pm	8:05pm	10:50pm

AMC Showplace Village Crossing 18 - Map

Standard	11:10am	1:50pm	4:25pm	7:25pm	9:10pm	10:10pm
3D	10:10am	12:45pm	3:25pm	6:20pm		

AMC Showplace Galewood 14 - Map

Standard	12:10pm	5:30pm	10:50pm		
3D	2:50pm	8:10pm			

More showtimes


The Legend of Tarzan (2016) - IMDb

www.imdb.com/title/tt0918940/ • IMDb •

★★★★+ Rating: 7.1/10 - 7,170 votes

Directed by David Yates. With Alexander Skarsgård, Rory J. Saper, Christian Stevens, Christoph Waltz. Tarzan, having acclimated to life in London, is called ...

Alexander Skarsgård · Full Cast & Crew · Rory J. Saper · Trivia



The Legend of Tarzan

2016 · Action/Adventure · 5h 50m

Play trailer on YouTube

7.1/10 IMDb 35% Rotten Tomatoes 43% Metacritic

It's been nearly a decade since Tarzan (Alexander Skarsgård), also known as John Clayton III, left Africa to live in Victorian England with his wife Jane. Danger lurks on the horizon as Leon Rom (Christoph Waltz), a treacherous envoy for King Leopold, devises a scheme that lures the couple to the Co... More

Release date: July 1, 2016 (USA)
Director: David Yates
Distributed by: Warner Bros.
MPAA rating: PG, PG-13
Production companies: Village Roadshow Pictures, RatPac-Dune Entertainment, Riche Productions

Critic reviews [View 5+ more](#)



Data conversion



	A	B	C	D	E	F	G	H	I	J
6	Klasyfikacja	Rodzaj Zapisu	Książka o filmie		Book	about	Movie	name		
7			Książka o twórcy		Book	about	author	roleName	intangible	name
8			Książka o stworze		Book	about	CreativeWork	workExample		
9			Książka twórcy		Book	author	Person	roleName	intangible	name
10			Książka w formie czecowym		Book	about	Role	roleName	intangible	name
11			książka wywiad		Book	genre				
12			utwór		CreativeWork	workExample	Book			
13			[książka która jest wydaniem utworu]		Book	exampleOfWork				
14		Dział			Book	about	Role	roleName	intangible	name
15		Twórca			ZOB. KARTOTEKA TWÓRCÓW					
16	Podstawowe dane książki	Autor			ZOB. KARTOTEKA AUTORÓW					
17		Tytuł			Book	name				
18		Tytuł Oryginału			Book	translationOfWork	Book	name		
19		Język			Book	inLanguage	Language			
20		Osoby współtworzące			ZOB. KARTOTEKA WSPÓŁTWÓRCÓW					
21		Funkcja	Adapt. i reżyseria		Zob. Inżynier adaptacji i reżyser					
22			Adaptator		CreativeWork	adaptationOfWork	CreativeWork	author	Person	
23			Analizy		Chapter	name	contributor	Person		
24			Autor tekstu składowego		Chapter	name	contributor	Person		
25			Autor utworu w którego powstaniu adaptacja		CreativeWork	adaptationOfWork	CreativeWork	author	Person	
26			Autor noty bibliograficznej		Chapter	name	contributor	Person		
27			Autor noty biograficznej		Chapter	name	contributor	Person		
28			Autor noty edytorskiej		Chapter	name	contributor	Person		
29			Autor posłowa		Chapter	name	contributor	Person		
30			Autor przedmowy		Chapter	name	contributor	Person		
31			Autor wstępu		Chapter	name	contributor	Person		
32			Autor wyboru		Book	editor	Person	Role	roleName	
33			Dialogi		CreativeWork	dialogues	Person			
34			Drzeworytami zdobite		Book	hasPart	VisualWork	artform	illustrator	Person
35			Fotografie		Book	hasPart	VisualWork	artform	illustrator	Person
36			Grafika		Book	hasPart	VisualWork	artform	illustrator	Person
37			Ilustrator		Book	hasPart	VisualWork	illustrator	Person	



Documentation of mapping to schema.org

Data conversion challenges



- **Work/Expression/Manifestation**
 - Selective application to reference commentary (reviews, polemics etc.)
 - Not used to reference translations: the problem of translations from secondary source
- **Contributor roles**
 - 86 different roles generated through autopsy
 - Overlaps with the note field
 - List simplified and flexible: general/obligatory and specific/voluntary
- **Linking the data with the LOD cloud**
 - places
 - people
 - subjects



Remediating the tool



IBL - Publikowanie roczników PBL - [Zapisy]

Struktura Kartoteki Zapisy Redakcja Parametry Zakończ pracę

Inny zapis Rok: 2002 Numer zapisu: 962585 Znacznik Uwagi

Klasyfikacja

Rodzaj zapisu: kult
Dział: Hasła osobowe (literatura polska) - L
Dział:
Twórcy: Leśmian Bolesław
Kartoteka twórców

Podstawowe dane

Autor, Nazwisko Imię Kryptonim Kartoteka autorów
Fornek Bożena
Tytuł: Był Leśmian w Zamościu
Odnosi się do zapisu: Szukaj zapisu Usunięcie odwołanie

Adnotacje

nt. imprez i konkursów w Zamościu, którym patronuje Bolesław Leśmian

Wytwórcie filmowe

Osoby współtworzące

Funkcja	Nazwisko, imię, uwagi	Kartoteka osób współtworzących

Opis:

Hasła przekrojowe i klucze do indeksu rzeczowego

Hasła i klucze	Dodatkowe słowa kluczowe
Miejscowości w Polsce Zamość	

Hasła i klucze: Dodatkowe słowa kluczowe

Opis: Czasopismo: ZamKwKrt Zamojsko Kwartalnik Kulturalny
Rok: 2002 Numer: nr 3/4 Strony: s. 113 Kartoteka źródeł
Zapis utworzony przez: BEATAK 03.09.16 Zmodyfikowany przez:



Remediating the tool (annotations)



Tytuł ? Przypnij to pole

tytuł równoległy
 tytuł podstawowy przełladu
 tytuł oryginalny

wpisz tytuł

Przypnij to pole

Autorstwo i typy utworów zawartych w antologii Przypnij to pole

Typ publikacji
Wybrano:

Zawiera wpisz określenie zawartości oraz osoby

Przypnij to pole

Osoby współtworzące ? Przypnij to pole

Rodzaj współtwórstwa
Wybrano:

(po wybraniu z listy automatycznie słowa te podstawily się w tym polu); wpisz osoby



Retrodigitisation



- 1939-1945 (3 vols., 1250 pages)
- 1944-88 (51 vols., 37 661 pages)
- 1976-1989 - *samizdat* (1 vol., text file)

SPIS TREŚCI

WSTĘP	VII	Tematy i motywy	87
SPIS ŹRÓDEŁ BIBLIOGRAFICZNYCH	IX	Wydawstwa	89
WYRAZ SKAŃCÓW	XXVIII	ORGANIZACJA NAUKI O LITERATURZE	
I. TEORIA LITERATURY			
Opis	2	Biuro	92
Metodologia badań literackich	4	Do s. 1939	92
Teoria procesu historycznoliterackiego	4	Wojna 1939-1945	93
Teoria dzieła literackiego	5	Nauka współczesna, od r. 1945	93
Stylistyka	5	Dydaktyka literatury	93
Formalizm	5	Kody znakowe	95
Psychologia	6	Nagrody naukowe	95
Komparacja	6	Odszacowanie	95
Genealogia	6	Współpraca z zagranicą	95
Teoria krytyki literackiej	10	Wzajemny prac naukowych	100
Psychologia literatury	10	Filologia	100
Sociologia literatury	10	Filologia polska	100
Zagadnienia przekładu	11	Instytut filologiczny	102
Literatura w innych naukach	12	Poliska Akademia Nauk, Wydział I	103
		Instytut Instytucji i Uniwersytetu Osobowe	103
		Nauka polska za granicą	105
		BAMA SZCZECIŃSKIE (Instytut i seminarium)	
II. LITERATURA POLSKA			
HISTORIA LITERATURY			
Opis	13	Instytut naukowy i sztuki	105
Literatura staropolska	19	Antologia i zbiory	108
Opis	21	Baszki naukowe	109
Rozmowy	22	III. LITERATURA OBCE	
Porównanie	23	Całość ogólna	425
Młoda Polska	24	Literatura angielska	426
Dwudziestolecie międzywojenne	25	Lit. amerykańska (Ameryka)	426
Literatura lat 1939-1945	27	Lit. hiszpańska	426
LITERATURA WSPÓŁCZESNA			
Opis	28	Lit. amerykańska	431
Opis	33	Lit. arabska	431
Krytyka	33	Lit. argentyńska	442
Przebieg	34	Lit. australijska	443
Przebieg	34	Lit. austriacka	443
Współczesna literatura polska za granicą	36	Lit. belgijska	445
Życie literackie polskie za granicą	37	Lit. brazylijska	445
Życie literackie polskie za granicą	37	Lit. brytyjska	445
Życie literackie polskie za granicą	37	Lit. bułgarska	456
Życie literackie polskie za granicą	37	Lit. chińska	459
Życie literackie polskie za granicą	37	Lit. chińska	459
ZAGADNIENIA SPECJALNE			
Bibliografia, lewicowstwo, katalogi, słowniki	66	Lit. czechosłowacka	460
Charakterystyka	70	czeska	460
Kontakty literackie z zagranicą	78	albanijska	465
Literatura dla dzieci i młodzieży	82	Lit. duńska	466
Literatura ludowa	83	Lit. egipska starożytna	467
Literatura regionalna	85	Lit. egipska współczesna	467
Pamiętnikarstwo	85	Lit. fińska	467
		Lit. francuska	468

924 INDEKS BZECZORY

<p>Szkoły podstawowe i zasadnicze polskie w kraju</p> <p>Szkoła Podstawowa im. M. Kosciuszkiej w Izbicy 2762</p> <p>Szkoła Podstawowa im. G. Morciana w Rędnym 3592</p> <p>Szkoła Podstawowa nr 42 im. K. I. Gałczyńskiego w Lublinie 1930</p> <p>Szkoła Podstawowa nr 2 im. L. Króczyńskiego w Krakowie 2584</p> <p>Szkoła Podstawowa nr 248 im. K. K. Baczynskiego w Karszewie 948</p> <p>Szkoła Podstawowa nr 10 im. H. Sienkiewicza w Opolu 4576</p> <p>Szkoła Podstawowa nr 10 w Krakowie 5486</p> <p>Szkoła Podstawowa nr 11 w Krakowie 5486</p> <p>Szkoła Podstawowa nr 80 im. K. Makuszyńskiego w Poznaniu 3394</p> <p>Szkoła Podstawowa nr 30 im. S. Wyspiańskiego w Katowicach 5486</p> <p>Szkoła Podstawowa nr 7 w Rzeszowie 2762</p> <p>Szkoła Podstawowa nr 191 im. J. L. Kaszewskiego w Warszawie 2959</p> <p>Zasadnicza Szkoła Osielbowa nr 2 w Krakowie 5486</p> <p>polskie na gruncie</p> <p>Szkoła im. M. Kosciuszkiej w Londynie 8816</p> <p>Szkoła Przemysłowców ojczyznym im. H. Sienkiewicza 8816</p> <p>Szkoła Sobotnia w Brecler 8816</p> <p>Szkoły średnie i wyższe zob. Uniwersytety, szkoły średnie i wyższe</p> <p>Szekholm 3300</p> <p>Szwajcaria 4576, 3663</p> <p>film 8880</p> <p>kontakty kulturalne z Polską, polonika 189, 333, 616, 702, 2044, 2045, 2511, 4771, 8425, 8851-8852</p> <p>literatura 127, 7038, 7275-7283 zob. również</p> <p>Realizacje polskie obcych utworów literackich teatr 6585, 8538, 8851, 8852</p> <p>Życie literackie polskie 421 III, 5328</p> <p>Życie nauki 84, 616, 702</p> <p>Szwajcaria</p> <p>kontakty kulturalne z Polską, polonika 357, 534, 686, 3300, 5120, 8663, 8694, 8852; zob. również: Przekłady na język szwajcarski</p> <p>literatura 3300, 7227, 7284-7314; zob. również</p> <p>Realizacje polskie obcych utworów literackich radio 9068</p> <p>teatr 8538, 8853</p> <p>Życie literackie 5730</p> <p>Życie literackie polskie 5487</p> <p>zob. również: Szekholm</p> <p>Śląsk 33, 366, 571, 2957, 3299, 3392, 3513, 4046, 4320, 4880, 5280, 5536, 8429, 8441, 8897</p> <p>członkostwo 797, 478, 491</p> <p>literatura 327, 348, 359, 361, 325, 454, 559, 560, 563, 567-568, 583, 748, 761, 796, 854, 3271, 3784</p> <p>teatr 397, 5486, 8447, 8541, 8546, 8550, 8731</p> <p>Życie literackie 397, 357, 368, 3880</p> <p>Życie nauki 623</p> <p>Śląsk Czerwony</p> <p>literatura 360, 754, 848, 1260</p>	<p>Śląsk Dolny 594</p> <p>członkostwo 496</p> <p>Śląsk Górny</p> <p>literatura 754</p> <p>teatr 8405, 8533, 8741</p> <p>Śląsk Opolski 330, 643</p> <p>literatura 754</p> <p>teatr 8404 I, 8489, 8492</p> <p>Śnieżka</p> <p>muzeum 717</p> <p>Śrem</p> <p>Życie literackie 397</p> <p>Świdnica 3869</p> <p>Świdwin</p> <p>Życie literackie 312 II, 5041</p> <p>Świnie</p> <p>Życie literackie 405</p> <p>Świętojaire 8795</p> <p>Życie literackie 406</p> <p>Tablice pamiątkowe zob. Pomniki</p> <p>Tadżykistan</p> <p>kontakty kulturalne z Polską, polonika 2413, 2416, 4501, 8430; zob. również: Przekłady na język tadżycki</p> <p>Tallin</p> <p>teatr 8868</p> <p>Teren 2496</p> <p>Życie literackie 540</p> <p>Terespol</p> <p>Życie literackie 4665</p> <p>Torun</p> <p>biblioteki 4754</p> <p>teatr 8857</p> <p>Tortu</p> <p>teatr 8540, 8868</p> <p>Toskani</p> <p>Życie literackie 5732</p> <p>Tatarska literatura 8289</p> <p>Teatr 8528-8868</p> <p>teatr amatorski 319, 319 VIII, 8404 I, 8425, 8430, 8434, 8449, 8492, 8493, 8507, 8508, 8516, 8517, 8525, 8720-8745, 8782, 8816-8818, 8820-8821, 8846; balet 8474; cyrk 313 VI; dla dzieci i młodzieży 8526, 8533, 8539, 8540, 8746-8782, 8897, 9034, 9035, 9037-9045, 9059; estrada 362, 8528, 8593; jednego aktora 8515, 8525, 8555, 8728; kabaret 8493, 8563, 8581, 8591, 8606, 8607, 8664-8665, 8707, 8708, 8741-8744, 8798-8802, 8813, 9032; kamienia del'antre 8514, 8875; lalki 556, 5738, 8405 III, 8419, 8521, 8658, 8746-8782, 8822, 8825, 8826, 8837, 8849, 8859; lufowcy 105, 556; musical 8451; obłądźcy 3227, 8584, 8634, 8641, 8657, 8791, 8792; obywateli 8497, 8498; opera 8453, 8516; pantomima 8450, 8451, 8515, 8521, 8536, 8537, 8713-8715, 8802, 8804; prr; sztuki 376, 8579, 8625; studentki 305 VI, 402, 491, 8405 VIII, 8504, 8539, 8783-8808, 8843, 8846, 8848, 8851, 8857, 8862, 8868; szkoły 844, 8493, 8517, 8732, 8741-8744; szopka 560, 5738, 8452, 8747; telewizyjny zob. Telewizja; wojakowy 8404 I, 8430, 8740, 8745</p> <p>historia teatru 51, 104, 105, 161, 164, 197, 201, 313, 745, 1613, 2992, 3665, 4415, 4441, 4881,</p>
---	---



Retrodigitisation challenges



- **Technology**
 - Parsing
 - Lemmatization
- **Tracing methodological inconsistencies**
- **Time factor**
 - Subject-classification (new approaches to literary studies, e.g. gender studies)
 - new forms of literary life, e.g. online literary life
 - changing political geography: Yugoslavia and Soviet Union no longer there

PAWLIKOWSKI MICHAŁ
1245. *Uwagi na temat Dżana. Pierwsza i druga
tęmi Arduwii. Traduzione dal polacco e note
di Enrico Damiata. Roma 1941 Edizioni D.O.C.,
n. 34. Carta d'Europa 2.*

PEIPER TADEUSZ
1246. [P.] Fragment listki o Kolczynie
[Odrady 1943 nr 40 s. 8 [opracowanie s. nr 44
s. 7 w rubryce: Korespondencja] -- W Je-
lenka. Twórcy. 1943 r. 1 s. 103 -- 121, s. 2
s. 99 -- 136.

1247. Helosiana (Kąkolnowo 1942). Odrady
1943 nr 35 s. 4 [w Niemcech białoruskich] --
O VII-ej symfonii Szostakowicza (Kąkolnowo,
1942). Odrady 1943 nr 46 s. 4. -- [Insc.] Har-
mofilania nad kłódką wroczliwą -- Odrady
1943 nr 36 s. 7 [krotka odpowiedź na ankietę
„Odnalezienie” = stowozka piątej do rury-
wistacji wojennej] -- Z Gabrieł Turtakowski.
Nasz Władca. 1944 nr 4 s. 13 [art. sprac-
o wyttawie „Wojna w słońcu Ojczyzny”]

PELZAR MARIAN
1248. Na granach Głaińska. Tyg. Powsz.
1943 nr 7 s. 2 [reportaż].

PELTZ ADAM
1249. [W.] Pełn rubryki zachodnich. W.
Pd. 1943 nr 12 s. 2.

Research challenges: local specificity



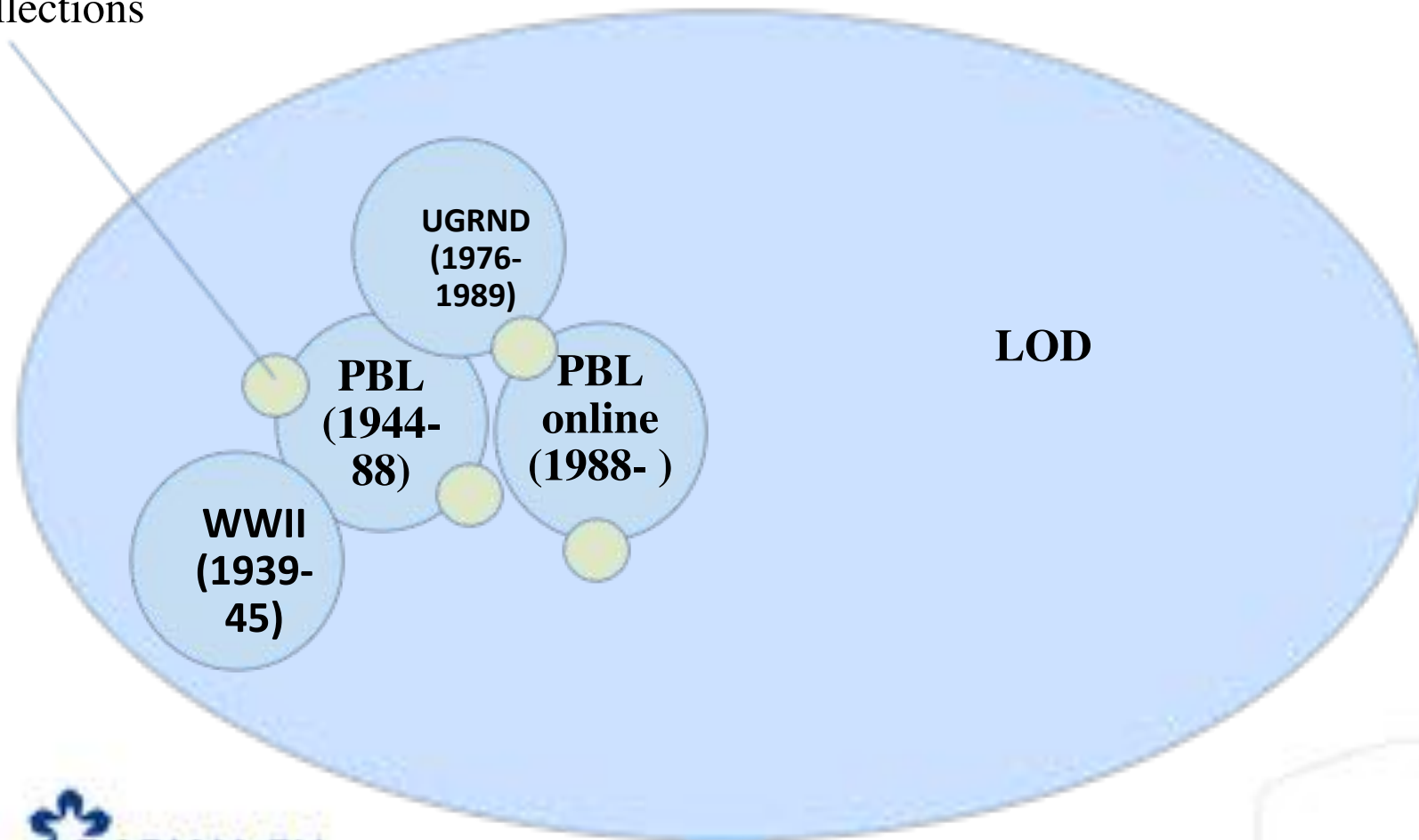
- **Data-collection methodology**
 - Selection of writers
 - Unrecorded debuts
 - Data censorship
- **Geography**
 - Changing borders
 - Literature in Exile
 - Polish literature = in Poland or in Polish?
 - Literary transfers
- **Shape of literary life**
 - Official vs. Underground publishing



PBL_lab VRE



Topical data
collections



Conscious research based on bibliographical data



In order to come up with right research questions (and answers)...

...be sure to know how your data were selected, collected, censored, abridged, standardised, annotated, digitised, corrected and linked together



References



- Bode, Katherine. 2014. *Reading by numbers. Recalibrating the literary field*. London: Anthem Press.
- Long, Hoyt and Richard So (2013a) "Network Analysis and the Sociology of Modernism" *boundary 2* 40(2):147-182
- Long, Hoyt and Richard So (2013b) "Network Science and Literary History" *Leonardo* 3(46), 274-274.
- Jockers, Matthew L. 2013 *Macroanalysis. Digital Methods & Literary History*, Chicago: Chicago UP.
- McCarty, Willard. 2008. "Modeling in Literary Studies" *A Companion to Digital Literary Studies*, ed. Susan Schreibman and Ray Siemens. Oxford: Blackwell, <http://www.digitalhumanities.org/companionDLS/> (28.02.2016)
- @maciejmaryl maryl.org maciej.maryl@ibl.waw.pl
- Moretti, Franco 2013 "Style, Inc.: Reflections on 7,000 Titles" *Distant Reading*, London: Verso.
- Moretti, Franco 2005 *Graphs, maps, trees. Abstract Models for Literary History*. London: Verso.