Global Public Books as an Exp Cultural Divers

Timothy J. Dickey, Ph.D.

Post-Doctoral Researcher

OCLC Research

OCLC™ The world's libraries. Connected.

Overview

Introduction to the Project

Preliminary Results:

- Publication profiles

- Language data

- Translation data

Next steps

Introduction: Genesis and Background of the Project
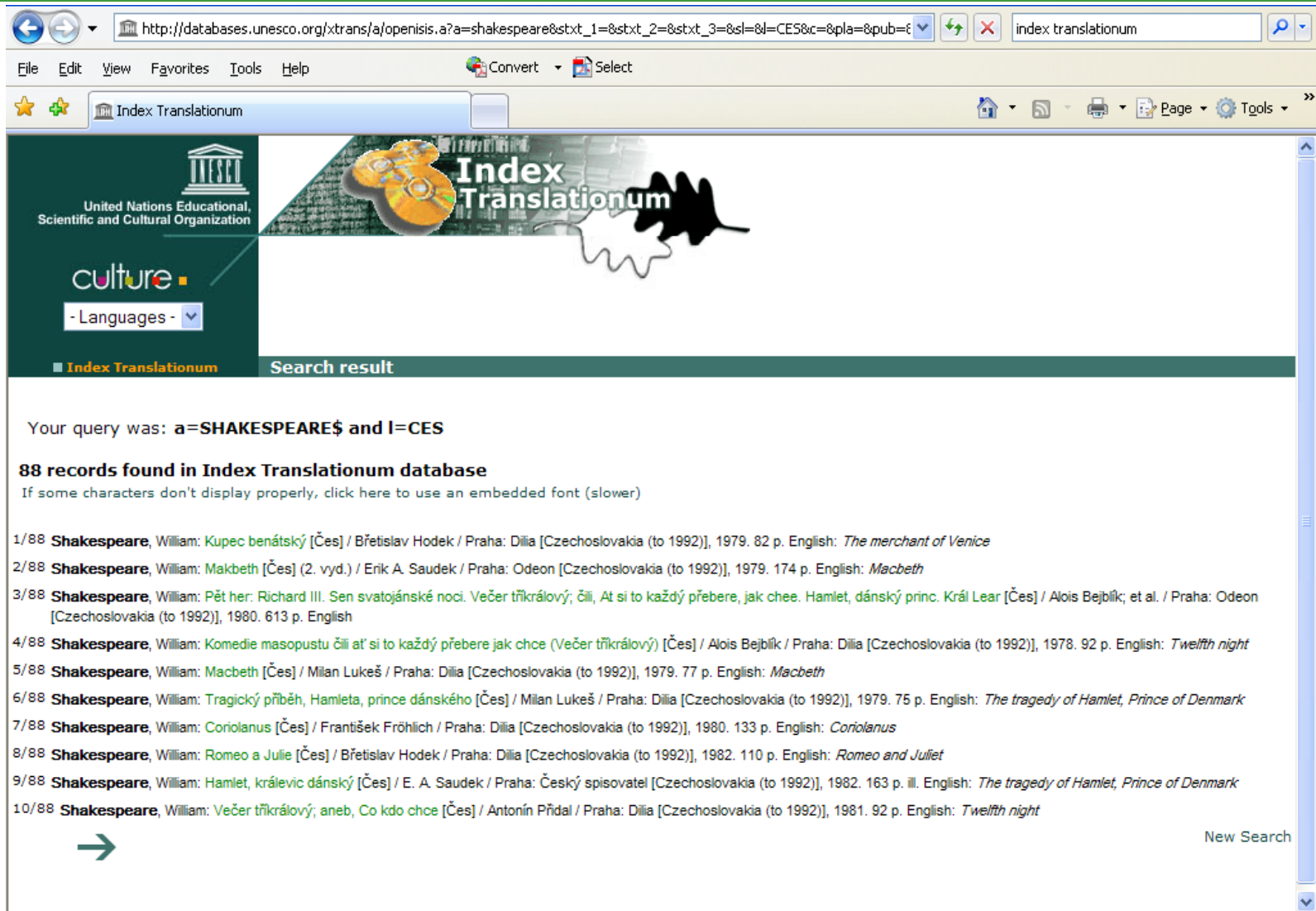
OCLC™ The world's libraries. Connected.

Genesis and Background of the Project

Interested in measures of cultural diversity:

- **UNESCO Institute for Statistics**

- **IFLA Statistics and Evaluation Committee**

- **ISO Global Library Statistics Project**

- **Book and Media Studies School, Leiden**

- *Index Translationum*

    - **UNESCO Division of Cultural Expressions & Creative Industries**

    - **1932 - present**

Genesis and Background of the Project

Genesis and Background of the Project

WorldCat™ and Global Book Data:

- **150 million records, 1.48 billion holdings**

  - **Strongest in monographic records**

- **71,000 libraries**

- **112 countries**

- **> 50% non-English cataloging**

- **> 470 languages**

Project Objectives

- **Mine data from WorldCat's monographic records**

- **Parse the data:**

  - **Country of publication**

  - **Year of publication**

  - **Language use as a "measure of cultural diversity"**

- **Seek patterns in the data**

Project Objectives

Data limits:

- **Textual material, non-serial**

- **Publication date numeric, < 2010**

Data extracted:

- **Published titles**

- **Holdings – indigenous libraries and worldwide**

- **Languages**

- **Translation data**

Project Objectives

"Pre-Test" of the Procedures

- Profiles of book publishing in six countries:
  - Bolivia, Chile, Germany, Poland, South Africa, Thailand

- Expansions after technique refinement:
  - China, Columbia, Finland, France, India, Indonesia, Italy, Kenya, Nigeria, Russia, Ukraine, Venezuela
  - Australia, Belgium, Denmark, Egypt, Greece, Iran, Iraq, Ireland, Japan, Korea(s), Mexico, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland

Project Limitations

- **"As reflected in WorldCat"**
  - **Collection patterns**
  - **Cataloging practice**
  - **Definition of a book**
  - **Definition of a COUNTRY**
  - **Definition of a publishing date**

National Publishing Profile

OCLC™ The world's libraries.
Connected.

Book Publishing Profiles: Publishing Output

Country	Publications	Holdings	Translations
Bolivia (1 OCLC member lib.)	58,838	302,309	780
Chile (387 member libraries)	265,948	900,330	7,351
Germany (378 members)	12,843,615	70,341,725	381,141
Poland (27 members)	1,225,446	3,108,677	174,738
South Africa (1,082 members)	299,574	2,014,327	10,063
Thailand (39 members)	181,003	463,104	5,700

Book Publishing Profiles: Publishing Output

Country	Publications	Holdings	Translations
China (1,160 member libs.)	3,208,114	8,160,062	238,315
Finland (71 members)	905,252	1,476,734	122,150
France (253 members)	4,948,620	29,127,570	211,341
India (37 members)	1,121,311	6,636,530	52,107
Russia (45 members)	1,883,418	8,151,097	94,319
Ukraine (3 members)	236,166	710,372	9,061

Book Publishing Profiles: Publishing Output
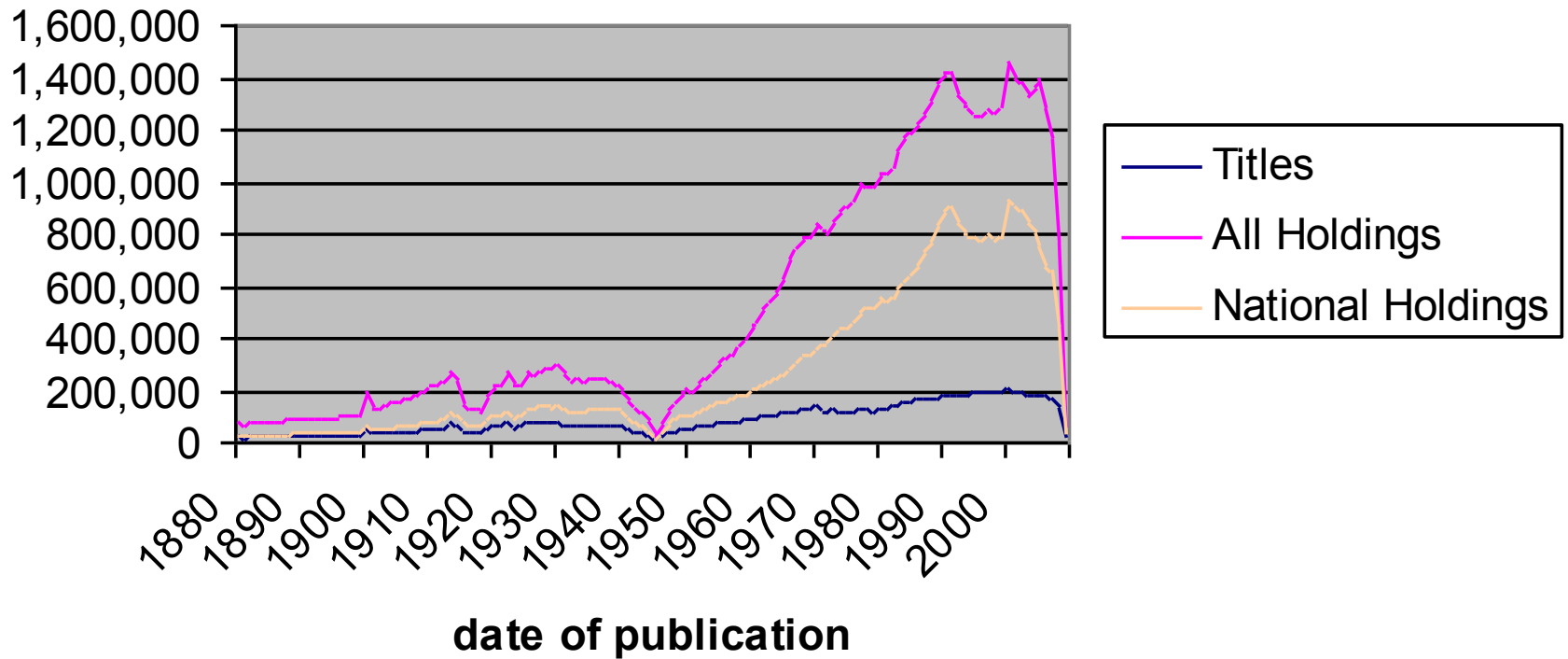
Germany: Book Publication

Legend:
- Titles
- All Holdings
- National Holdings

date of publication

Book Publishing Profiles: Publishing Output

Germany: Book Publication

Legend: Titles, All Holdings, National Holdings

x-axis: date of publication (1480–1870)
y-axis: 0 – 70,000
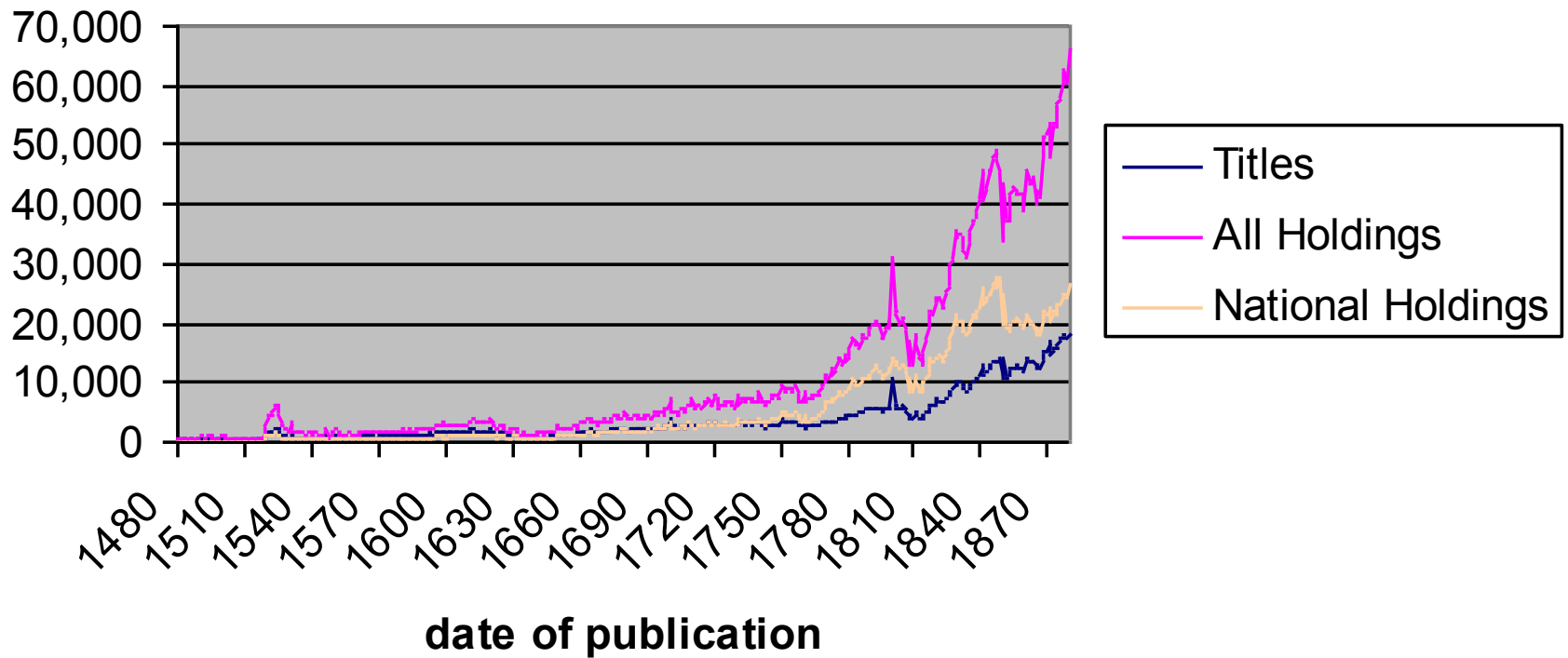
Book Publishing Profiles: Publishing Output

France: Book Publication

Legend:
- Titles
- Holdings

Y-axis: 60,000 / 50,000 / 40,000 / 30,000 / 20,000 / 10,000 / 0

X-axis (date of publication): 1600, 1620, 1640, 1660, 1680, 1700, 1720, 1740, 1760, 1780, 1800, 1820, 1840, 1860
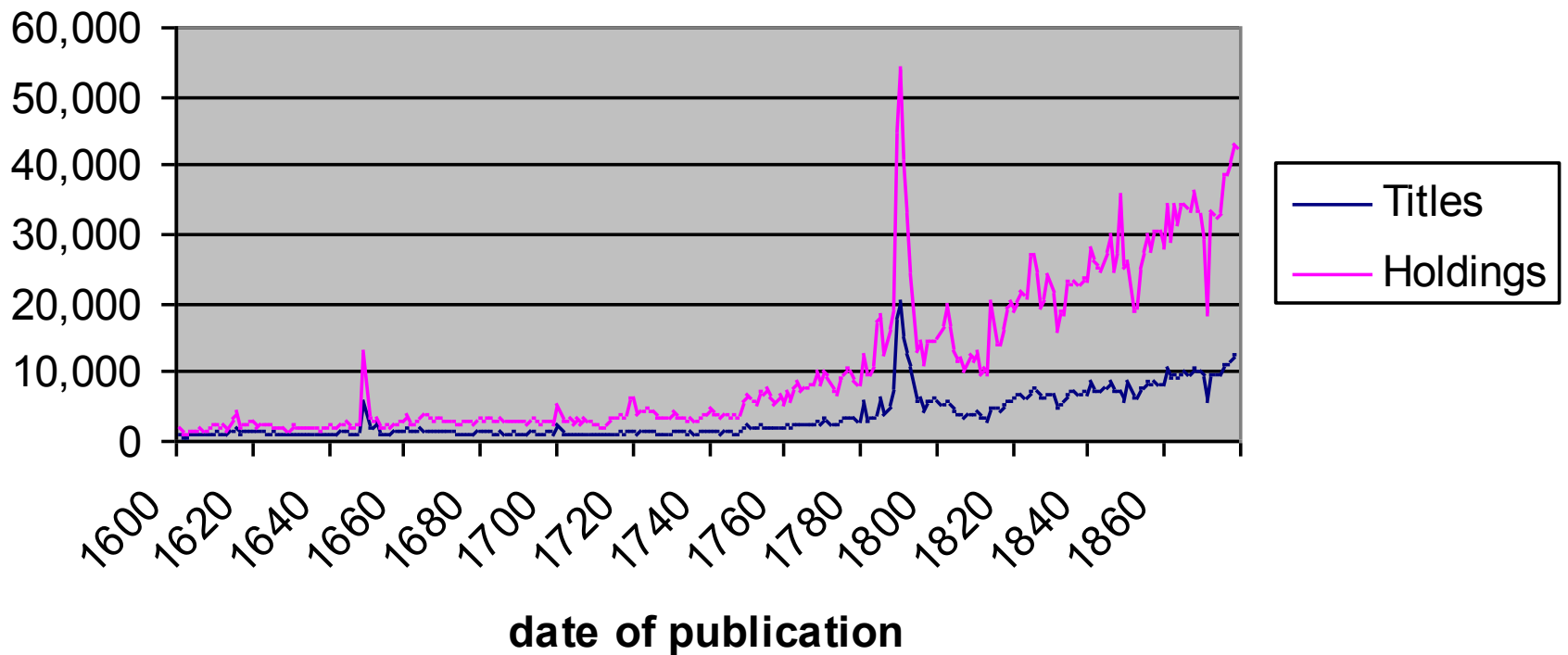
Poland: Book Publication

date of publication

Legend:
- Titles
- All Holdings
- National Holdings
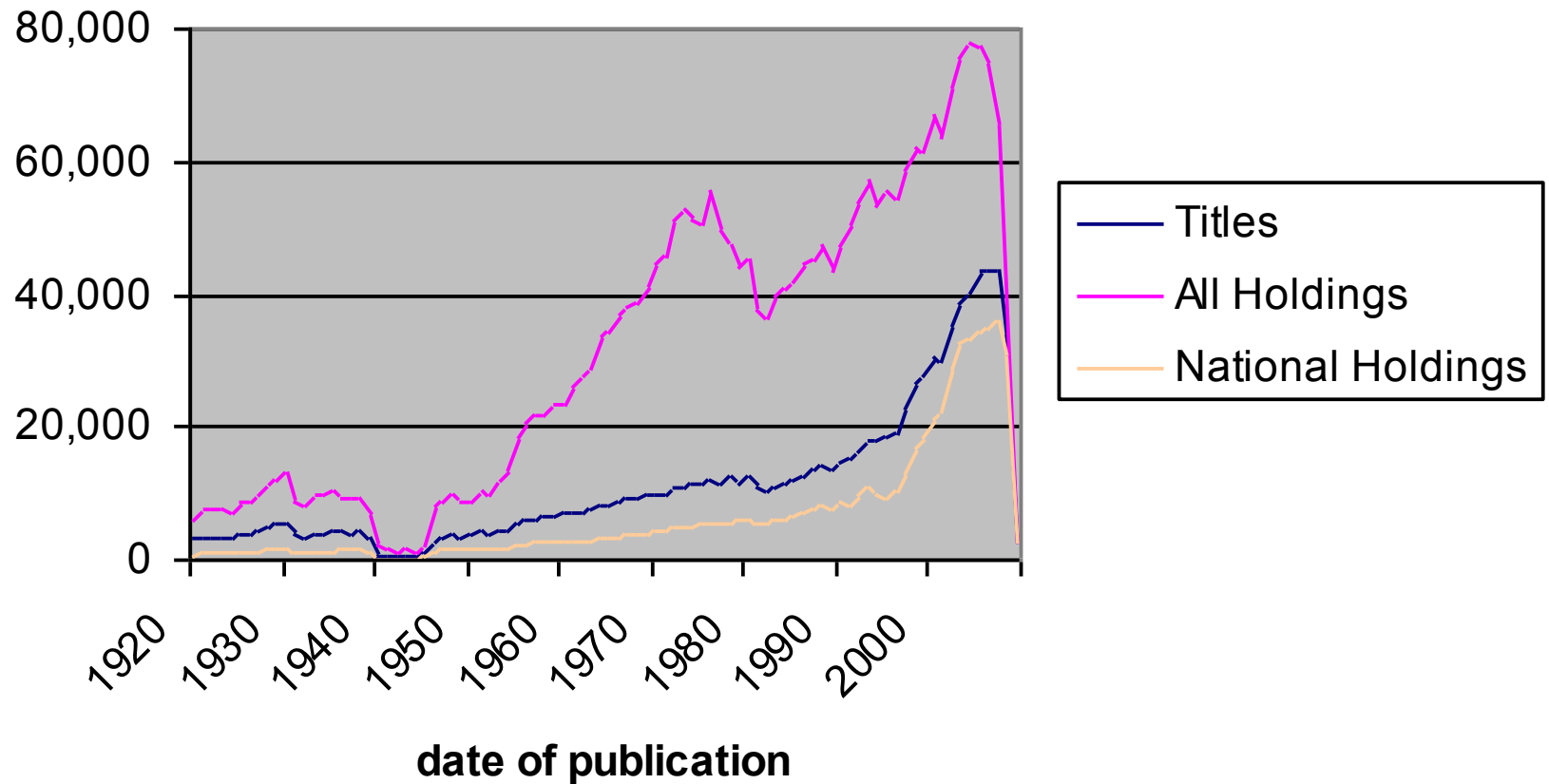
South Africa: Book Publication
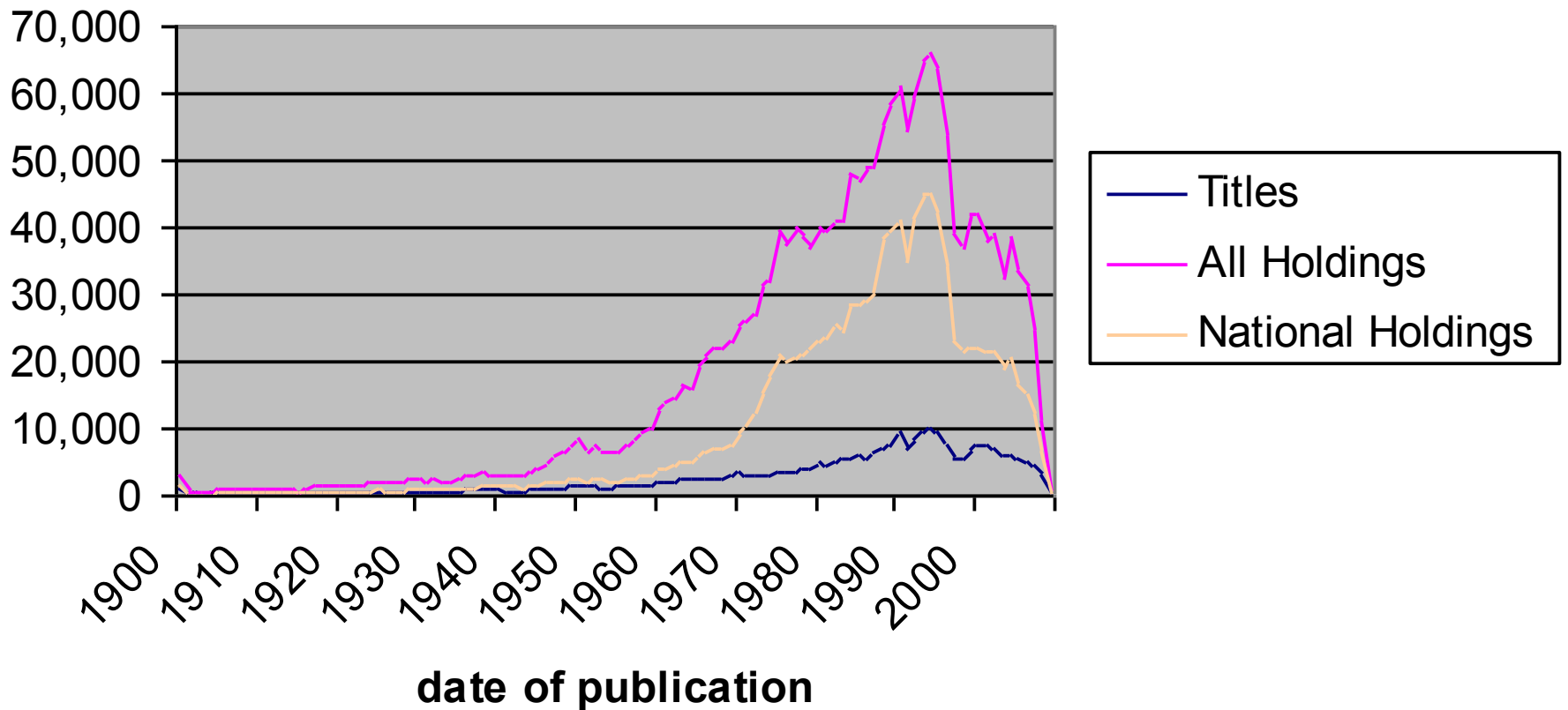
Book Publishing Profiles: Language Data

	Publications	Translations	Language	Titles	Translations	Language Distribution
Bolivia	58,838	780	Spanish (O)	56,326	529	96.56%
			Quechua (O)	202	46	0.35%
			Aymara (O)	228	28	0.39%
			Guarani	28	10	0.05%
			Other Amerindian	269	64	0.46%
			(English)	887	69	1.52%
Chile	265,948	7,351	Spanish (O)	248,935	6,505	94.20%
			English	10,965	533	4.15%
			German	772	81	0.29%
			Mapudungun	127	33	0.05%
			(French)	935	55	0.35%
Thailand	181,003	5,700	Thai	129,461	3,806	73.70%
			English	44,233	1,514	22.76%
			Chinese	1,224	65	0.70%
			(French)	991	91	0.56%
			(Pali)	920	47	0.52%

Book Publishing Profiles: Language Data

	Publications	Translations	Language	Titles	Translations	Language Distribution
Germany	12,843,604	381,141	German	10,558,895	317,446	83.73%
			Danish	1,712	320	0.01%
			Frisian	401	51	0.00%
			Sorbian	83	22	0.00%
			Romani	32	4	0.00%
			(English)	580,149	29,072	4.60%
			(Latin)	378,698	8,041	3.00%
			(French)	92,516	6,919	0.73%
			(Italian)	16,901	1,588	0.13%
Poland	1,225,446	174,738	Polish	917,229	150,061	78.95%
			German	85,304	5,041	7.34%
			Ukrainian	2,105	282	0.18%
			Lithuanian	327	81	0.03%
			Belarusian	691	92	0.06%
			Kashubian	43	6	0.00%
			(English)	63,137	10,916	5.43%
			(Latin)	23,877	882	2.06%
			(French)	13,556	2,204	1.17%
			(Russian)	9,927	912	0.85%

German-language Publishing in Germany

A line chart titled "German-language Publishing in Germany" with x-axis "date of publication" ranging from 1880 to 2000, and y-axis from 0 to 200,000. Two lines: "Titles" (dark blue) and "Translations" (magenta).
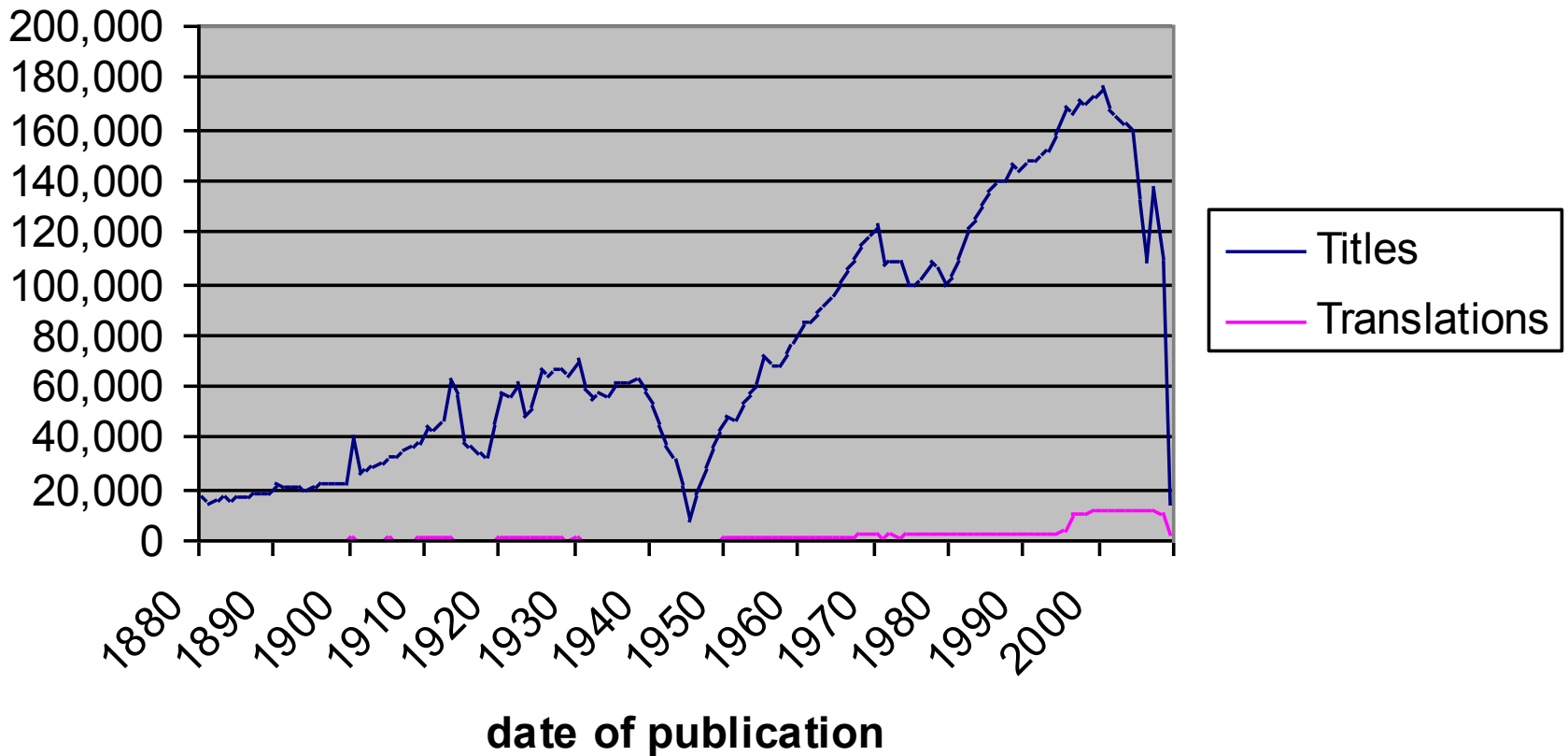
Book Publishing Profiles: Language Data

Latin-language Publishing in Germany
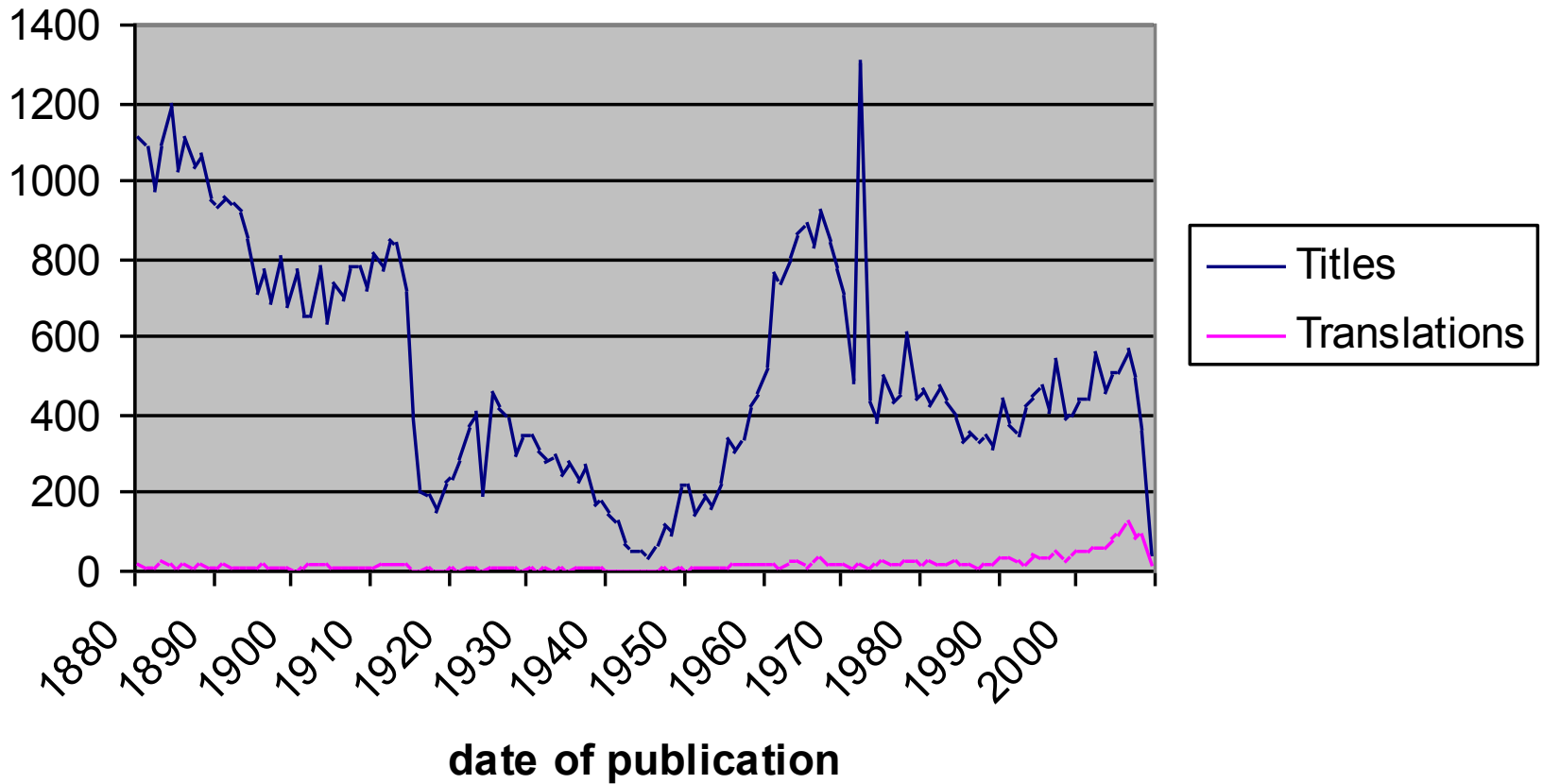
Legend: Titles, Translations

date of publication

Book Publishing Profiles: Language Data

	Publications	Translations	Language	Titles	Translations	Language Distribution
South Africa	299,574	10,063	English	193,504	2,082	65.30%
			Afrikaans	77,449	5,145	26.13%
			Zulu	2,649	487	0.89%
			Xhosa	2,364	553	0.80%
			Sesotho	1,368	298	0.46%
			Sepedi	1,344	177	0.45%
			Setswana	1,290	154	0.44%
			Xitsonga	729	132	0.25%
			Venda	670	115	0.23%
			Ndebele	167	31	0.06%
			Other indigenous languages	96	7	0.03%
			(Dutch)	3,175	82	1.07%
			(German)	1,320	76	0.45%

Book Publishing Profiles: Translation Data

	Publications	Translations	Original	Translation	Number	
South Africa	299,574	10,063	Afrikaans	English	4,077	40.51%
			English	Afrikaans	1,008	10.02%
			Xhosa	English	453	4.50%
			Zulu	English	357	3.55%
			Afrikaans	Dutch	292	2.90%
			Sesotho	English	219	2.18%
			Afrikaans	German	209	2.08%
			Tswana	English	204	2.03%
			English	German	165	1.64%
			Sepedi	English	109	1.08%
			Venda	English	80	0.79%
			English	Zulu	54	0.54%
			Zulu	Afrikaans	40	0.40%
			English	Xhosa	31	0.31%

Book Publishing Profiles: Translation Data

Translated Works in WorldCat Published in South Africa
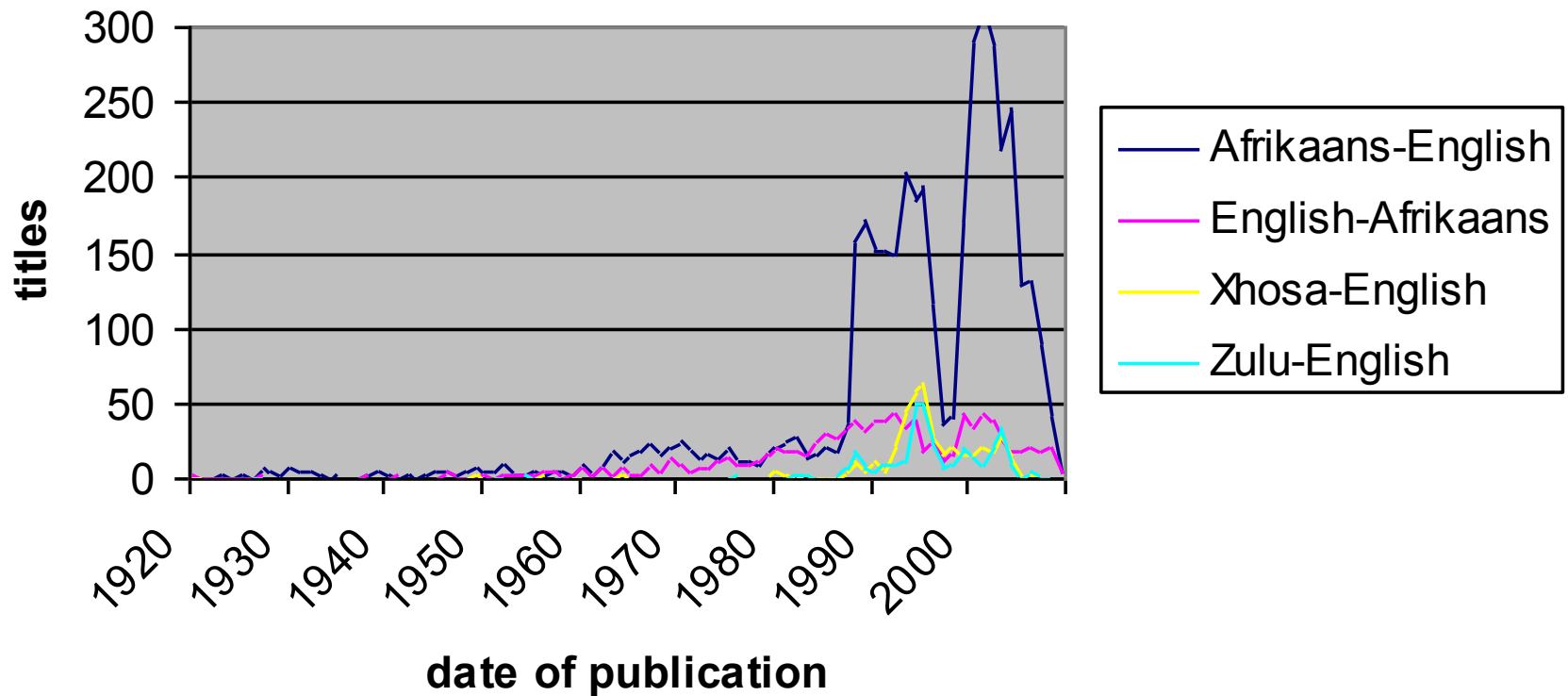
Legend:
- Afrikaans-English
- English-Afrikaans
- Xhosa-English
- Zulu-English

y-axis: titles (0, 50, 100, 150, 200, 250, 300)

x-axis: date of publication (1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000)

Book Publishing Profiles: Translation Data

	Publications	Translations	Original	Translation	Number	
Germany	12,843,605	381,141	English	German	149,318	39.18%
			French	German	41,076	10.78%
			Russian	German	14,254	3.74%
			German	English	19,861	5.21%
			Latin	German	12,811	3.36%
			Italian	German	12,103	3.18%
			Swedish	German	12,006	3.15%
			Dutch	German	9,176	2.41%
			Spanish	German	6,998	1.84%
			Greek, Ancient	German	6,178	1.62%
Poland	1,225,446	174,738	English	Polish	72,638	41.57%
			German	Polish	20,626	11.80%
			French	Polish	13,778	7.88%
			Russian	Polish	9,828	5.62%
			Polish	English	8,673	4.96%
			Italian	Polish	4,885	2.80%
			Polish	German	3,085	1.77%
			Latin	Polish	2,687	1.54%
			Spanish	Polish	2,443	1.40%
			Swedish	Polish	1,693	0.97%
			Czech	Polish	1,619	0.93%

Book Publishing Profiles: Translation Data
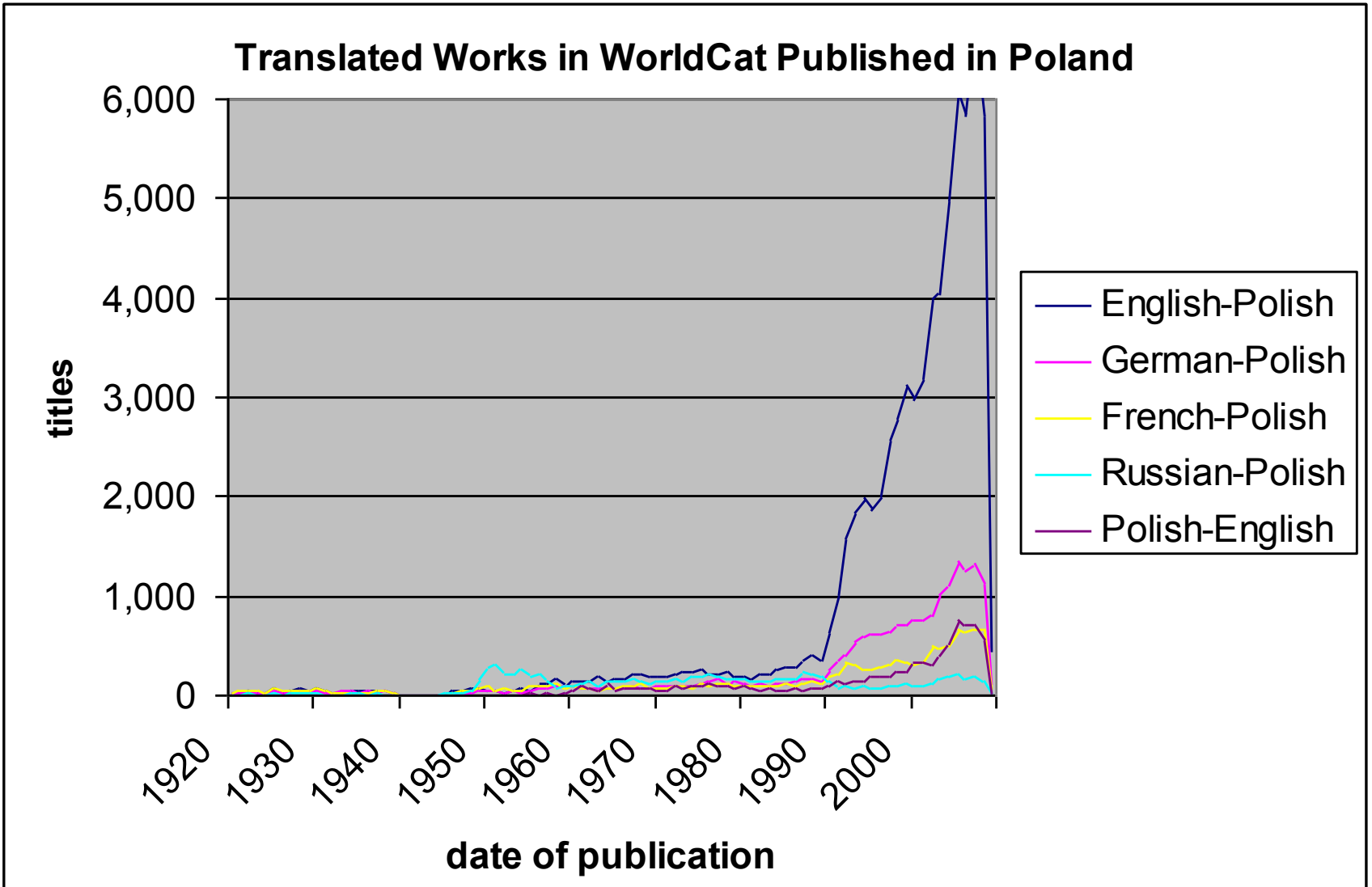
Translated Works in WorldCat Published in Poland

Legend:
- English-Polish
- German-Polish
- French-Polish
- Russian-Polish
- Polish-English

y-axis: titles (0; 1,000; 2,000; 3,000; 4,000; 5,000; 6,000)

x-axis: date of publication (1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000)

Book Publishing Profiles: Translation Data

	Publications	Translations	Original	Translation	Number	
Bolivia	58,838	780	English	Spanish	211	27.05%
			French	Spanish	61	7.82%
			Spanish	English	58	7.44%
			German	Spanish	50	6.41%
			Aymara	Spanish	32	4.10%
Chile	265,948	7,351	English	Spanish	2,571	34.97%
			French	Spanish	1,365	18.57%
			German	Spanish	575	7.82%
			Spanish	English	442	6.01%
			Italian	Spanish	410	5.58%
Thailand	181,003	5,700	English	Thai	2,322	40.74%
			Thai	English	983	17.25%
			Chinese	Thai	237	4.16%
			Pali	Thai	178	3.12%
			French	Thai	165	2.89%
			French	English	154	2.70%

Translated Works in WorldCat Published in Thailand

titles

120
100
80
60
40
20
0

Thai-English
English-Thai
Thai-Chinese
Thai-French

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000

date of publication

Book Publishing Profiles: Translation Data

Translated Works In Worldcat Published in Italy

Legend:
- English-Italian
- French-Italian
- German-Italian
- Latin-Italian
- Italian-English
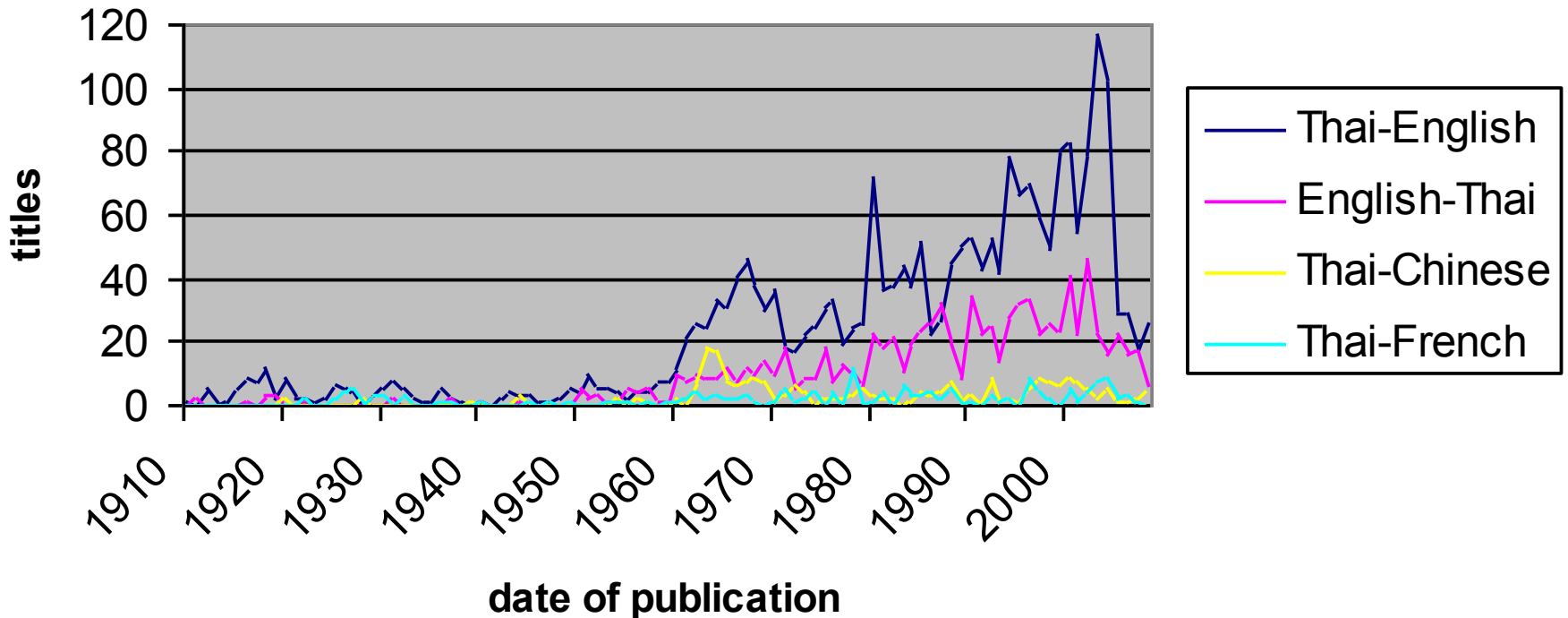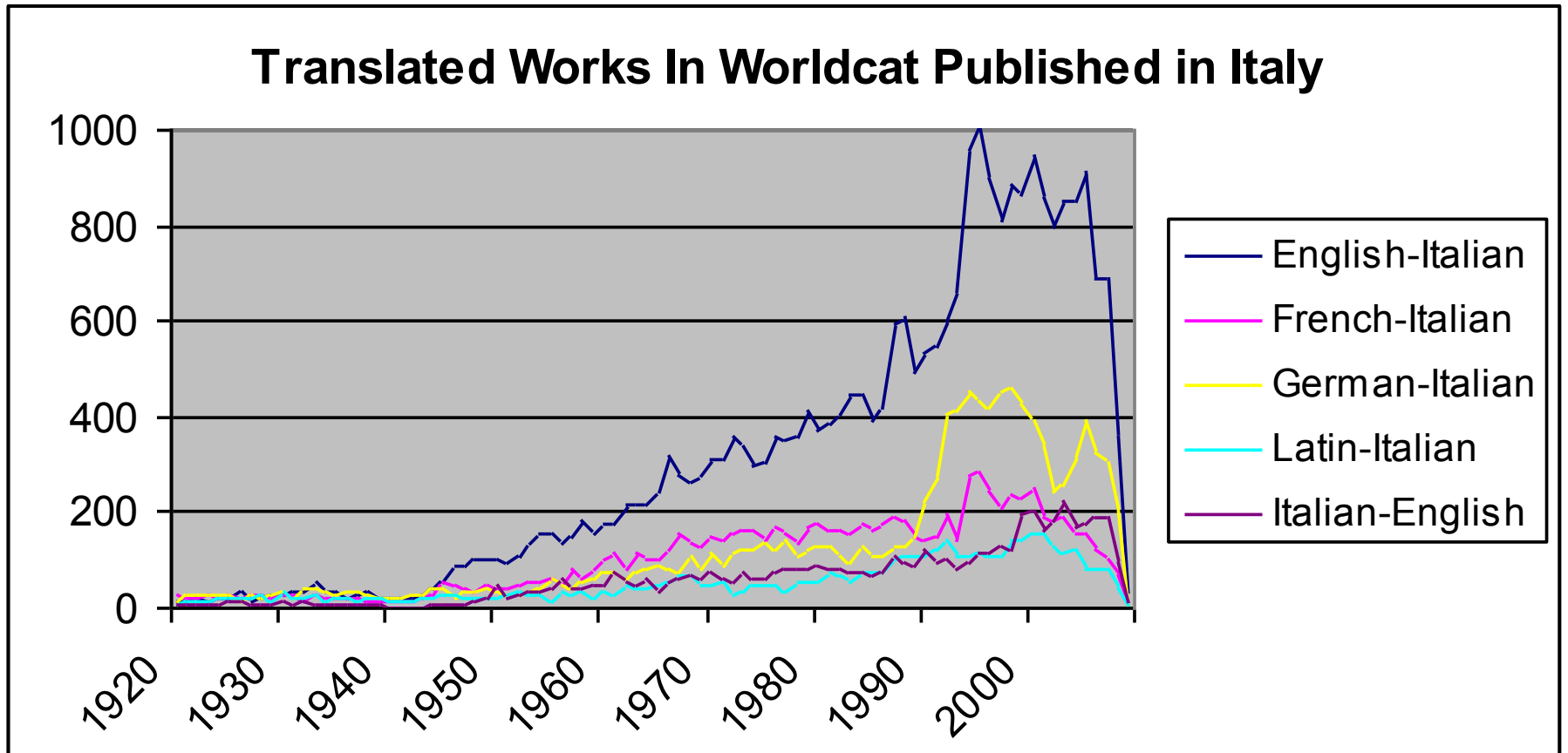
Next Steps

- **Final refinements to data extraction**

- **Integrate OCLC Audience Level**

- **Complete extraction of "foreign" holdings data**

- **Check results against *Index Translationum***

- **Go global!**
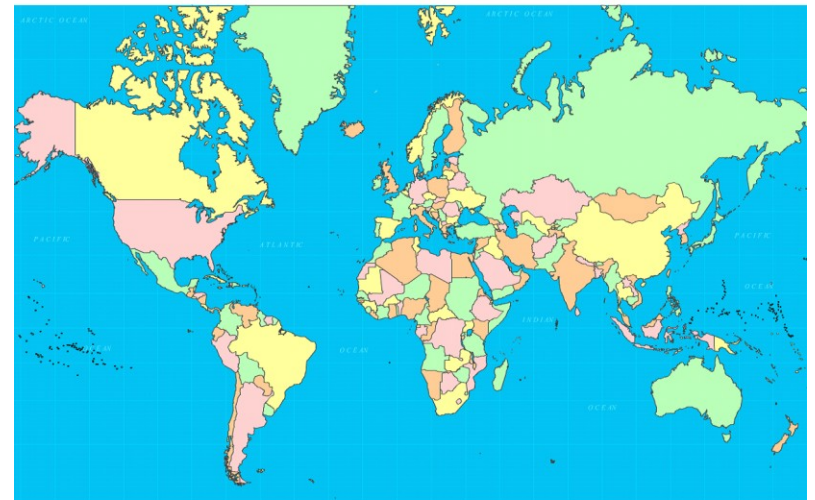
Questions?

Timothy J. Dickey, Ph.D.

dickeyt@oclc.org

www.oclc.org/research/projects/globalbooks/default.htm

Special thanks to Jeremy Browning, Lynn Silipigni Connaway, Michael Heaney, and Karen Smith-Yoshimura

OCLC™ The world's libraries.
Connected.