Database Developments
HISTORICAL METHODS, April–June 2014, Volume 47, Number 2
Copyright © Taylor & Francis Group, LLC

Routledge
Taylor & Francis Group

iPEHD—The ifo Prussian Economic History Database

SASCHA O. BECKER
Department of Economics and Centre for Competitive Advantage in the Global Economy (CAGE)
University of Warwick

FRANCESCO CINNIRELLA
Ifo Institute
CESifo

Centre for Economic and Policy Research

Center for Competitive Advantage in the Global Economy (CAGE)
University of Warwick

ERIK HORNUNG
Max Planck Institute for Tax Law and Public Finance

LUDGER WOESSMANN
Department of Economics
University of Munich

Ifo Institute

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014

Abstract. This article describes the ifo Prussian Economic History Database (iPEHD), a public use county-level database covering a rich collection of variables for nineteenth-century Prussia. The Royal Prussian Statistical Office collected these data in several censuses over the years 1816–1901. These data provide a unique source for micro-regional empirical research in economic history, enabling analyses of education, religion, fertility, industrialization, and many others. The service of iPEHD is to provide the data in a digitized and structured way.

Keywords: county, database, economic history, nineteenth century, Prussia

Prussia is a fascinating setting to study many of the most fundamental questions in nineteenth-century economic history. Its combination of high micro-regional diversity, uniform institutional setting, and rich sources of surviving data provide researchers with unique opportunities to study the interrelationships between education, religion, fertility, industrialization, economic development, and many other topics.

Starting with the first full-scale population census in 1816, the Royal Prussian Statistical Office collected a large amount of economic and demographic data. Many interesting county-level data have survived in archives. Thanks to the Prussian proverbial orderliness and thoroughness, we have high quality data for the Prussian counties (Kreise) covering nearly the whole range of the nineteenth century. These data provide a unique source for empirical research in economic history, with the particular potential to study historical micro-regional data with modern microeconometric methods.

The ifo Prussian Economic History Database (iPEHD) provides many of these data in a digitized and structured way. Data and documentation, including background information, definitions, and sources of variables, are freely available to any interested researcher at the iPEHD website (www.cesifo-group.de/ipehd). The iPEHD website also makes suggestions on how to merge data from different census waves with varying administrative boundaries into panel datasets. Finally, it provides a collection of thematic maps visualizing the data,

Address correspondence to Francesco Cinnirella, Ifo Institute, Poschingerstr. 5, 81679 Munich, Germany. E-mail: cinnirella@ifo.de

ready-made datasets, and codes to replicate tables from published research.

Throughout, iPEHD covers all Prussian counties, whose number increased over the nineteenth century from 308 in 1816 to 574 in 1901. Drawing from a total of 15 original sources—many of which contain several volumes—iPEHD comprises more than 1,500 variables organized into eight content areas: education, occupation, wages and income tax, industry, agriculture, population, religion, and miscellaneous. In total, iPEHD contains more than half a million data points at the county level. While nowhere near being a complete collection of all available data, iPEHD provides a comprehensive micro-regional database on nineteenth-century economic history in Prussia.

The iPEHD complements several other historical database projects. The North Atlantic Population Project (NAPP), carried out at the Minnesota Population Center (2008) in collaboration with several international partners, includes individual-level census data for the Grand Duchy of Mecklenburg-Schwerin for 1819. To allow comparative analysis, NAPP provides the data in a way that is consistent across countries and compatible with the existing IPUMS series of U.S. census samples.[1] Building on the IPUMS and NAPP projects, the Mosaic project, coordinated by the Max Planck Institute for Demographic Research (MPIDR) in Rostock, started in 2011 with the aim of building a comprehensive and detailed database of historical census micro data for Europe and beyond. So far, harmonized individual-level data are available for regions in France, Germany, Serbia, Spain, and Switzerland.[2]

This article documents the construction of the iPEHD and provides guidance on how to access and use the data. Original sources are documented, and research that has been conducted to date using iPEHD data is briefly described.

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014

A Brief History of the Birth of iPEHD

In 2006, when looking for data to analyze the relationship of religion and literacy with economic outcomes in German history, we stumbled upon the rich county-level data available from the Prussian census of 1871. Soon, we recognized the sheer amount of data that were just sitting around in the statistical annals at German state libraries. The quality of this impressive collection of information, remarkable for the nineteenth century, has generally been regarded as excellent by historians and demographers (cf., e.g., Galloway, Hammel, and Lee 1994).[3] Compared to the selective samples that a lot of historical research is restricted to, the full censuses covering the whole population provide a much more reliable picture of the historical setting.

After the original analysis of the 1871 census data and its subsequent publication (Becker and Woessmann 2009), we explored annals covering rarely examined census data from 1816 to 1821.[4] Although much effort must be undertaken to make these data ready for research and to ensure their comparability, we soon found them to be very promising and equally reliable. A third large data digitization project involved the census of 1849.

TABLE 1. The Scope of iPEHD

Year	No. of variables	No. of county observations	No. of data points
1816	58	308	17,864
1819	5	344	1,720
1821	22	344	7,568
1816–21	24	456	10,944
1829	6	59	354
1849	712	335	238,520
1858	6	342	2,052
1862	4	346	1,384
1864	53	347	18,391
1866a	1	342	342
1866b	11	334	3,674
1871a	25	453	11,325
1871b	14	458	6,412
1878	5	426	2,130
1882a	269	464	124,816
1882b	14	465	6,510
1886a	156	544	84,864
1886b	97	518	50,246
1892	8	550	4,400
1896	15	552	8,280
1901	8	574	4,592
Sum	1,513		606,388

Note: Some of the data points may contain missing information.

These 1816, 1849, and 1871 census data became the foundation of iPEHD. In subsequent years, we digitized data from additional censuses to fill in gaps. Although still far from complete, the data provide a rather comprehensive overview of nineteenth-century economic history in Prussia.

iPEHD went online in the summer of 2012. The collection of these data and their provision to the scientific community is part of the project, "Establishment of a leading international center for empirical research on the importance of education for long-term economic development," generously funded by the Leibniz Association under the Pact for Research and Innovation. The project was carried out at the Human Capital and Innovation department at the Ifo Institute-Leibniz Institute for Economic Research at the University of Munich.

Overview and Structure of the Data Contained in iPEHD

iPEHD starts with the population census in 1816, the first full-scale census released by the Royal Prussian Statistical Office, which had been founded in 1805. The 1816 census covers the 308 Prussian counties at the time. Further extensive census data are available in 1849, 1864, 1871, and 1882, but—as indicated in Table 1—many more detailed data were

collected in additional years. As the number of counties grew over time, by 1901, the data cover 574 Prussian counties. In total, iPEHD contains more than 1,500 variables and more than half a million data points, all at the county level.

iPEHD consists of county-level information gathered from these different censuses. The data are currently presented in 76 separate data files, organized by content area, specific topic, and census year. Each data file in iPEHD contains a unique county (Kreis) identifier (discussed in below), the county name, the abbreviated district (Regierungsbezirk) name (rb), and a set of variables of census data. iPEHD stores its data in comma-separated values (csv) format, which is easily accessible from any statistical software. For example, to open the csv data files in Stata, one must type:

insheet using "xxxxxx.csv."

To give an example of a data file, Table 2 shows a brief extract of a few variables for the first few counties (by alphabet) from the data file "ipehd_1819_indu_fac.csv," which contains data on the number of factories in a county in 1819. The variable "fac1819_brick" documents the total number of brick manufactories in a county in 1819, and the variable "mill1819_water" is the total number of water mills.

Areas of Content Covered by iPEHD

The iPEHD data are categorized into the following eight content areas:

- *Education*: This area contains, among others, such data as the number of students, teachers, and schools by school type, literacy, and school finance.
- *Occupation*: This area contains, among others, data on the labor force in agriculture, factories, manufacturing, crafts, and services.
- *Wages and income tax*: This area contains data on daily wages of day laborers, teacher income, and income taxes.
- *Industry*: This area contains data on a huge number of different factories, technologies, and transportation.
- *Agriculture*: This area contains, among others, such data as livestock, crop yields, soil composition, and the distribution of land.
- *Population*: This area contains data on population by age, gender, and marital status, on births and deaths, and on population with disabilities.
- *Religion*: This area contains denomination-specific data on population, literacy, education, occupation, and number of churches.
- *Miscellaneous*: This area contains such data as surface area, buildings, municipalities, and residential areas for each county.

Apart from the data gathered in these eight content areas, the merger file provides information on merger variables necessary to combine data from different census years (see below).

Codebooks

Codebooks provide additional information for each variable contained in iPEHD. There is one codebook for each year so that explanations for each variable can be found

TABLE 2. Extract From an Example Data File

kreiskey1800	County	rb	fac1819_brick	fac1819_lime	fac1819_glass	mill1819_water
277	Achen	AAC	5	10	2	26
33	Adelnau	POS	11	6	0	26
254	Adenau	KOB	0	1	0	71
196	Ahaus	MUN	11	15	0	20
255	Ahrweiler	KOB	0	0	0	51
2	Allenstein	KON	5	0	1	31
219	Altena	ARN	3	13	0	41
257	Altenkirchen	KOB	1	0	0	41
10	Angerburg	GUM	4	26	0	5
53	Angermünde	POT	13	2	0	28
32	Anklam	STE	3	0	0	2
209	Arnsberg	ARN	12	4	0	26
67	Arnswalde	FRA	7	3	3	29
160	Aschersleben	MAG	8	5	0	57
55	(Nieder-)Barnim	POT	8	0	1	30
54	(Ober-)Barnim	POT	18	0	0	36
190	Beckum	MUN	8	3	0	22

Note: Extract from iPEHD data file "ipehd_1819_indu_fac.csv."

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014

in the codebook for the corresponding year. A summary codebook that combines all years is also provided; this summary codebook allows a content search of the whole iPEHD.

The codebooks list the variable name ("variable name"), the name of the data file where it can be found ("ipehd datasets"), an English label ("label"), and the original label in German language ("original label"). The German language label is similar to the table headings found in the original sources. The English label leads with the year and is a shortened (direct) translation of the German label; in cases where a translation is not feasible, the original German term was adopted. In addition, the codebooks indicate the source of each set of variables ("source").

Merging Data From Different Censuses

One of the biggest challenges when analyzing historical data is to ensure comparability over time. A key service of iPEHD is to facilitate the analysis of data from different census years at the county level, holding the administrative boundaries fixed. This section describes the county-level structure in which the data are presented in iPEHD, county identifiers, and the suggested procedure of combining different census years.

Starting after the Congress of Vienna in 1815, Prussia reformed its administrative structure and introduced the county level. At the time, the dimension of a county was meant to follow borders of previously existing administrative units. The maximum distance to the administrative center was meant to be two to three Prussian miles (roughly 15 to 23 km or 9 to 14 miles), such that every inhabitant could travel forth and back within a day. The population size was meant to range between 20,000 in sparsely populated areas and 36,000 in densely populated areas.

Throughout the nineteenth century, various administrative reforms reshaped the county structure of Prussia. As the popu -lation grew over time, it became necessary to divide existing administrative units in order to reduce administrative efforts. Most of these changes were partitions of one county into two or more counties.

Thus, it is usually possible to reconstruct earlier administrative units by aggregating data from later years to the former structure. A drawback of this procedure is that the researcher loses part of the variation provided by having more observations. Still, the procedure appears necessary in order to have intertemporal comparability of the units of observation. The alternative would be to assign the same early data to two or more subsequently parted units, introducing measurement error if observed data were not uniformly distributed in the original area.

A peculiarity of the Prussian county system is the city county. Starting with the introduction of the county level in 1815, the so-called Immediatstädte (immediate towns) became a county themselves. As urbanization advanced, an increasing number of cities were detached from their original county and became a county of their own. Thus, the database often contains a Landkreis (rural county) and a Stadtkreis (city county) with similar names. For example, there are six pairs of Landkreis/Stadtkreis information among the 335 county observations in the 1849 classification and 20 pairs among the 458 county observations in the 1874 classification.

All data in iPEHD reflect the administrative conditions in place at the date of publication of the census. Since censuses often ordered the counties in different ways, identifiers were assigned reflecting the order of each census. Thus, each county in each census has been assigned a continuous number which is unique within a census but not across censuses. The identifiers are named kreiskeyYYYY, where YYYY represents the four-digit year (see below for additional peculiarities of the 1816–21 data).

The year in the identifier denotes the administrative structure of Prussia, which is not necessarily the same as the census year. In some cases, different identifiers (e.g., kreiskey1871 and kreiskey1874) must be assigned to data from the same census year (1871) because the Royal Prussian Statistical Office used different aggregations in different publications of data from the same census.

Researchers may be interested in intertemporal comparisons and in the construction of panel datasets using iPEHD. The iPEHD county identifiers, together with the merge-county file also provided on the iPEHD website, provide a service that facilitates such linkage of comparable units of observation over time. Our suggestion is that in order to obtain a comparable set of observations, researchers should collapse the data to the earliest set of counties in the data. However, the best way to structure and use the data may vary by research project.

To conduct most intertemporal comparisons, our suggestion is to take the following nine-step procedure. To construct cross-sections, the procedure should be followed only until step 3.

1. Choose datasets from the same census year.
2. Merge all datasets from the same census year using the identifier (e.g., kreiskey1882).
3. Save the cross-section.
4. Use the merge-county file provided on the iPEHD website.
5. Drop all duplicate and missing observations from the merge-county file according to the identifier in the cross-section (e.g., kreiskey1882; see below for an example).
6. Merge the merge-county file with the cross-section using the identifier (e.g., kreiskey1882).
7. Aggregate (sum/mean) all variables in the cross-section to the aggregation level of the earliest census in the

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014

analysis using the identifier of the earliest census in the analysis.
8. Repeat steps 1 to 7 for datasets from other census years.
9. Merge the resulting cross-sections using the identifier of the earliest census in the analysis.

Example From the Merger File

In the example shown in Table 3, the eight illustrative counties observed in 1901 were established from six counties in 1874 and five counties in 1849. Between 1849 and 1874, the "Elbing Landkreis" had been divided into "Elbing Stadtkreis" and "Elbing Landkreis." Between 1874 and 1901, the "Danzig Landkreis" had been divided into "Danzig Niederung," "Danzig Höhe," and "Dirschau."

In order to have a comparable set of observations when performing intertemporal comparisons between 1901 and 1849, one must aggregate the observations of "Danzig Niederung," "Danzig Höhe," and "Dirschau" to match "Danzig Landkreis." Thus, one should always aggregate the data to the aggregation level of the earliest census year in the specific analysis (step 7).

However, to perform intertemporal comparisons between, for example, 1874 and 1849, one must drop the duplicate entries of "Danzig Landkreis" from the merger file first (step 5). In addition, one must drop entries from the merger file that have missing observations on the county identifier in the respective year. Such missing observations exist because some territories were annexed by Prussia only after the respective census year.

As one example of how to merge datasets from 1874 and 1849, the following Stata code exemplifies the nine steps of the suggested procedure:

```
insheet using "ipehd_1871_edu_literacy_part2.csv", clear /* Step 1 */
save "ipehd_1871_edu_literacy_part2.dta"
insheet using "ipehd_1871_pop_demo_part2.csv", clear
save "ipehd_1871_pop_demo_part2.dta"
merge 1:1 kreiskey1874 using "ipehd_1871_edu_literacy_part2.dta" /* Step 2 */
drop _merge
save "ipehd_1871_part2.dta" /* Step 3 */
insheet using "ipehd_merge_county.csv", clear /* Step 4 */
duplicates drop kreiskey1874, force /* Step 5 */
drop if kreiskey1874 = .
merge 1:1 kreiskey1874 using "ipehd_1871_part2.dta" /* Step 6 */
collapse (sum) pop* lit* edu*, by (kreiskey1849) /* Step 7 */
drop if kreiskey1849 = .
save "ipehd_1871_part2_2.dta"
insheet using "ipehd_1849_rel_deno.csv", clear /* Step 8 */
save "ipehd_1849_rel_deno.dta"
insheet using "ipehd_merge_county.csv", clear
duplicates drop kreiskey1849, force
drop if kreiskey1849 = .
merge 1:1 kreiskey1849 using "ipehd_1849_rel_deno.dta"
collapse (sum) rel*, by (kreiskey1849)
save "ipehd_1849.dta"
merge 1:1 kreiskey1849 using "ipehd_1871_part2_2.dta" /* Step 9 */
drop _merge
save "ipehd_1849_1871.dta."
```

Peculiarity of the Data From 1816 to 1821

By 1816, Prussia had just commenced the administrative reform that established the county level. In some parts of the country, the reforms had not been finalized even in 1821. Thus, the data from the censuses in 1816 until after 1821 sometimes reflect old administrative units.

Unfortunately, due to the reform, these old units were subsequently aggregated and then newly divided in order to establish new counties. This makes it impossible to accurately match the data of a few administrative units from the early censuses to counties in subsequent censuses. The kreiskey1800 is thus coded so as to aggregate the data to a higher level. The kreiskey1800 can be used to link the 1816–21 data to later periods.

However, iPEHD also provides a unique identifier that allows merging data from the same census for these cross

TABLE 3. Example From the Merge File

Kreiskey 1901	County1901	Kreiskey 1874	County1874	Kreiskey 1849	County1849
38	ELBING STADTKREIS	38	ELBING STADTKREIS	37	ELBING LANDKREIS
39	ELBING LANDKREIS	39	ELBING LANDKREIS	37	ELBING LANDKREIS
40	MARIENBURG IN PREUSSEN	40	MARIENBURG IN PREUSSEN	38	MARIENBURG IN PREUSSEN
41	DANZIG STADTKREIS	41	DANZIG STADTKREIS	39	DANZIG STADTKREIS
42	DANZIG NIEDERUNG	42	DANZIG LANDKREIS	40	DANZIG LANDKREIS
43	DANZIG HOHE	42	DANZIG LANDKREIS	40	DANZIG LANDKREIS
44	DIRSCHAU	42	DANZIG LANDKREIS	40	DANZIG LANDKREIS
45	PREUSSISCH STARGARD	43	PREUSSISCH STARGARD	41	PREUSSISCH STARGARD

Note. Extract from the iPEHD merge file "ipehd_merge_county.csv."

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014

-sections. These identifiers are named "id1816" and "id1819." In order to merge data from 1816 to other data from 1816, researchers should use id1816. In order to merge data from 1819 or 1821 to other data from 1819 or 1821, researchers should use id1819.

In order to merge data from 1816, 1819, or 1821 to data from subsequent censuses, one should take the following steps:

1. Choose datasets from 1816, 1819, or 1821.
2. Merge all datasets from the same census using the identifier (idYYYY).
3. Aggregate (sum/mean) all cross-sections using the identifier "kreiskey1800."
4. Merge the cross-section with aggregated data from subsequent censuses using the identifier "kreiskey1800."

Original Sources Underlying the iPEHD Data

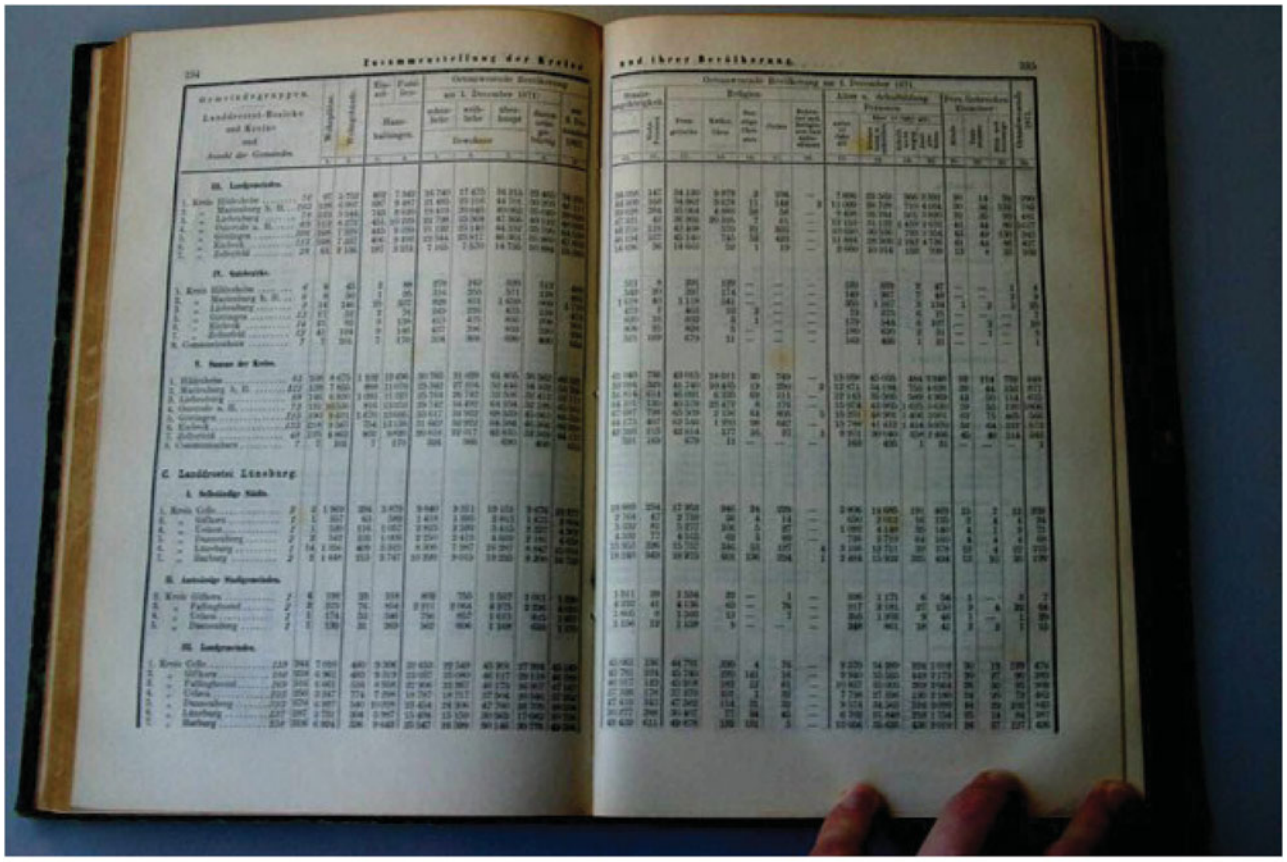
The iPEHD data have been digitized from different sources originally published by the Royal Prussian Statistical Bureau or its employees. These original historical volumes should be consulted for detailed information regarding the exact attributes of the data.[5] Figure 1 shows two example pages from such source volumes. In digitizing the data, the original data were neither modified nor aggregated. Rather, iPEHD accurately copies table headings and data entries as in the original tables. iPEHD only corrects obvious typos in the printed original as in the case where the column and/or row entries did not add up to the printed sums.

The following list documents all the volumes used as sources for iPEHD. There are a total of 15 original sources, many of which consist of several volumes:

- 1816–21: Mützell, Alexander A. (1821–5). Neues Topographisch-statistisch-geographisches Wörterbuch des Preussischen Staats, Vol. 1–6. Halle: Karl August Kümmel.
- 1829: Preussisches Statistisches Landesamt (1829). Beiträge zur Statistik der Königlichen Preussischen Rheinlande, aus amtlichen Nachrichten zusammengestellt. Aachen: J. A. Mayer.
- 1849: Statistisches Bureau zu Berlin (1851–5). Tabellen und amtliche Nachrichten über den Preussischen Staat für das Jahr 1849, Vol. 1–6b. Berlin: Statistisches Bureau zu Berlin.
- 1858: Meitzen, August (1868). Der Boden und die landwirthschaftlichen Verhältnisse des Preussischen Staates, Vol. 1–4. Berlin: Verlag von Paul Parey.
- 1862: Königlich Preussisches Statistisches Bureau (1863). Die Eisen-, Stein- und Wasserstrassen des preussischen Staates im Jahre 1862, in Zeitschrift des Königlich Preussischen Statistischen Bureaus, Vol. 3, 206–14. Berlin: Verlag der Königlichen Geheimen Ober-Hofbuchdruckerei.
- 1864: Königliches Statistisches Bureau in Berlin (1867). Die Ergebnisse der Volkszählung und Volksbeschreibung, der Gebäude und Viehzählung, nach den Aufnahmen vom 3. December 1864, resp. Anfang 1865 und die Statistik der Bewegung der Bevölkerung in den Jahren 1862, 1863 und 1864. Preussische Statistik Vol. 10. Berlin: Verlag von Ernst Kuehn.
- 1866: Meitzen, August (1868). Der Boden und die landwirthschaftlichen Verhältnisse des Preussischen Staates, Vol. 1–4. Berlin: Verlag von Paul Parey.
- 1871: Königliches Statistisches Bureau (1873–4). Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: Nach den Urmaterialien der allgemeinen Volkszählung vom 1.December 1871, Vol. 1–11. Berlin: Verlag des Königlichen Statistischen Bureaus.
- Königliches Statistisches Bureau in Berlin (1875). Die Ergebnisse der Volkszählung und Volksbeschreibung im Preussischen Staate vom 1. December 1871. Preussische Statistik Vol. 30. Berlin: Verlag des Königlichen Statistischen Bureaus.
- 1878: Herrfurth, Ludwig and Conrad Studt (1880). Finanzstatistik der Kreise des preussischen Staates für das Jahr 1877/78. Zeitschrift des Preussischen Statistischen Landesamtes, Ergänzungshefte, Vol. 7. Berlin: Verlag des Königlichen Statistischen Bureaus.
- 1882: Königliches Statistisches Bureau in Berlin (1884/85). Die Ergebnisse der Berufsstatistik vom 5. Juni 1882 im preussischen Staat. Preussische Statistik Vol. 76 a–c. Berlin: Verlag des Königlichen Statistischen Bureaus.
- 1886 (Education): Königliches Statistisches Bureau in Berlin (1889). Das gesammte Volksschulwesen im preußischen Staate im Jahre 1886. Preussische Statistik Vol. 101. Berlin: Verlag des Königlichen Statistischen Bureaus.
- 1886 (Agriculture): Königliches Statistisches Bureau in Berlin (1887). Die Ergebnisse der Ermittelung des Ernteertrags im preussischen Staate für das Jahr 1886. Preussische Statistik Vol. 92. Berlin: Verlag des Königlichen Statistischen Bureaus.
- 1892: Neuhaus, Georg (1904). Die ortsüblichen Tagelöhne gewöhnlicher Tagearbeiter in Preußen 1892 und 1901, in Zeitschrift des Königlich Preussischen Statistischen Bureaus, Vol. 44, 310–46. Berlin: Verlag des Königlichen Statistischen Bureaus.
- 1896: Königliches Statistisches Bureau in Berlin (1897). Die Ergebnisse der Ermittelung des Ernteertrags im preussischen Staate für das Jahr 1896. Preussische Statistik Vol. 147. Berlin: Verlag des Königlichen Statistischen Bureaus.
- 1901: Neuhaus, Georg (1904). Die ortsüblichen Tagelöhne gewöhnlicher Tagearbeiter in Preußen 1892 und 1901, in Zeitschrift des Königlich Preussischen

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014



(2.)

Kreise, bzw. grössere Städte.	Im Jahre 1861 und früher Geborne, welche nicht lesen und schreiben können (Analphabeten): evangelische: männl.	evangelische: weibl.	katholische: männl.	katholische: weibl.	israelitische: männl.	israelitische: weibl.	übrige: männl.	übrige: weibl.	Analphabeten überhaupt: männl.	Analphabeten überhaupt: weibl.	Die Angabe über Schulbildung fehlt für Personen: männl.	Die Angabe über Schulbildung fehlt für Personen: weibl.	Kinder, geboren in den Jahren 1871—1862: männl.	Kinder, geboren in den Jahren 1871—1862: weibl.
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Regierungsbezirk Königsberg.														
1. Memel	4 500	7 388	539	451	81	148	22	36	5 142	8 023	365	540	6 564	6 553
2. Fischhausen	3 109	3 699	31	8	7	8	5	11	3 152	3 726	215	276	5 861	5 654
3. Königsberg (Stadtkreis)	2 659	5 757	149	178	92	186	6	36	2 906	6 157	32	138	10 524	10 303
4. „ (Landkreis)	3 889	4 658	20	11	13	10	3	2	3 925	4 681	385	561	6 299	6 187
5. Labiau	4 745	7 006	32	12	18	26	10	17	4 805	7 061	137	299	6 402	6 487
6. Wehlau	3 473	4 761	26	12	13	14	1	3	3 513	4 790	333	497	5 764	5 790
7. Gerdauen	2 348	3 368	6	4	13	10	—	—	2 367	3 382	365	414	4 935	4 897
8. Rastenburg	2 499	4 012	177	214	9	4	2	3	2 687	4 233	101	322	5 181	5 317
9. Friedland	2 390	3 495	13	14	8	5	9	20	2 420	3 534	686	832	5 405	5 434
10. Preuss. Eylau	3 504	4 667	42	62	4	7	22	47	3 572	4 783	364	510	7 145	7 239
11. Heiligenbeil	3 100	4 349	45	61	4	7	20	56	3 169	4 473	233	397	5 888	5 934
12. Braunsberg	191	312	2 150	3 853	26	13	1	1	2 368	4 179	41	71	6 393	6 287
13. Heilsberg	139	209	2 321	4 230	9	12	—	1	2 469	4 452	435	555	6 599	6 688
14. Rössel	428	521	4 065	6 117	20	32	—	—	4 513	6 670	332	391	6 319	6 354
15. Allenstein	370	234	6 641	8 794	12	18	3	—	7 026	9 046	419	550	7 499	7 442
16. Ortelsburg	5 868	8 234	928	799	32	34	59	104	6 887	9 171	348	465	9 139	9 080
17. Neidenburg	4 521	6 274	2 216	1 632	26	35	19	1	6 782	7 942	313	399	7 323	7 381
18. Osterode a./Drewenz	7 843	10 093	1 229	1 276	30	37	8	15	9 110	11 421	197	315	8 875	8 955
19. Mohrungen	3 974	5 827	230	240	17	30	7	8	4 228	6 105	233	410	7 370	7 359
20. Preussisch Holland	1 612	2 285	121	144	5	10	3	2	1 741	2 441	135	159	5 922	5 651
Summe I.	61 162	87 149	20 981	28 112	439	646	200	363	82 782	116 270	5 669	8 101	135 407	134 992

FIGURE 1. Example pages from source volumes. ***Note:*** **The top picture is from Königliches Statistisches Bureau (1873), VIII: 234–5. The bottom picture is from Königliches Statistisches Bureau in Berlin (1875), 117 (color figure available online).**

Statistischen Bureaus, Vol. 44, 310–46. Berlin: Verlag des Königlichen Statistischen Bureaus.

Existing Research Using the iPEHD Data

By now, a lot of research in economic history has used data from the iPEHD. This research is briefly described in this section.[6] For those articles already published in academic journals, the iPEHD website provides ready-made datasets and codes in Stata to replicate the tables published in the articles. In addition, many more projects are currently under way and will be added to the website as publications become available.

A first set of articles using iPEHD data focused on the relationship between religion, human capital, and economic outcomes. Becker and Woessmann (2009) developed a human capital theory of Protestant economic history.[7] They used data from several censuses (Population 1871, Occupation 1882, Education 1886) and additional sources (including the Income Tax Statistics 1877) to show that the higher economic prosperity of Protestant relative to Catholic counties can be accounted for by Protestants' higher literacy, which was presumably spurred by instruction in reading the Bible. This finding suggests that explanations in the tradition of Max Weber that are based purely on differential work ethics may have limited explanatory power. Using data from the first Prussian census in 1816, Sascha Becker and Ludger Woessmann (2010) showed that Protestantism was related to higher schooling already in 1816, before the Industrial Revolution, ruling out that Protestant education resulted from industrialization. Also using 1816 data, among others, Becker and Woessmann (2008) showed that a larger share of Protestants in a county's population decreased the gender gap in basic education.

Using data from 1816–21 and 1869–71, Becker and Woessmann (2011b) found a substantial positive effect of Protestantism on suicide. Combining income data with data on Protestant church attendance in Prussian counties for six waves from 1886–1911, Becker and Woessmann (2013) found that in contrast to a negative cross-sectional association, panel analyses did not confirm a significant relationship between income and church attendance.

A second set of articles investigates the role of education during the phases of the Industrial Revolution. Sascha Becker, Erik Hornung, and Ludger Woessmann (2011) combined school-enrollment and factory-employment data from 1816, 1849, and 1882 to show that in contrast to the previous view based on British evidence, basic education was significantly associated with industrialization outside the textile sector in both phases of the Industrial Revolution.[8] Combining data from several censuses that effectively span the entire nineteenth century (1816, 1849, 1864, 1886, and 1896), as well as data from a 1866 classification of soil composition, Francesco Cinnirella and Erik Hornung (2011) found that landownership concentration, a proxy for the institution of serf labor, had a negative impact on school enrollment which diminished during the second half of the century.

Finally, the iPEHD data have been used to investigate the relationship between fertility and education. Becker, Cinnirella, and Woessmann (2010) used data from the 1849 census and other sources to show that a tradeoff between child quantity and quality existed already in the mid-nineteenth century and suggest that causation between fertility and education runs both ways. Using data from the 1816 census, Becker, Cinnirella, and Woessmann (2012) found a significant negative effect of education on fertility (evidence for a child quantity-quality tradeoff) already several decades before the demographic transition and show that it is robust to accounting for spatial autocorrelation. Combining data from three censuses (1816, 1849, and 1867), Becker, Cinnirella, and Woessmann (2013) turned to the education level of the parental generation and found a negative association of women's education with fertility, despite controlling for several demand and supply factors.

Additional Features of the iPEHD Website

The iPEHD website contains a number of additional features. For example, it provides a collection of thematic maps, produced using geographic information system (GIS) software, which visualize the geographical distribution of several interesting variables across the Prussian territory (see Figure 2 for an example). While the underlying historical boundary files were primarily devised for illustrative purposes, the Mosaic project recently provided on its website a set of GIS files of annual historical administrative boundaries for the German Union (1815–70) and the German Empire (1871–1945) produced by the Population History GIS Collection project at the MPIDR (2011).[9]

iPEHD is not the only project dealing with historical Prussian data at the county level. Other projects provide such services as maps, information on territorial changes, additional data, and other material on Prussian counties. Links to websites of several of these projects, whose work are highly appreciated and can be viewed as complementary to iPEHD, are provided on the iPEHD website.

Finally, the iPEHD website contains a section on frequently asked questions, providing answers to standard problems encountered by iPEHD users.

Conclusions

The data contained in iPEHD, originally collected by the Royal Prussian Statistical Office, are an impressive collection of information whose quality, already in the nineteenth century, is generally regarded as excellent. Now digitized from censuses located in archives, these county-level data provide information on education, occupation, income and tax measures, industry, agriculture, demographics, religion, and

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014

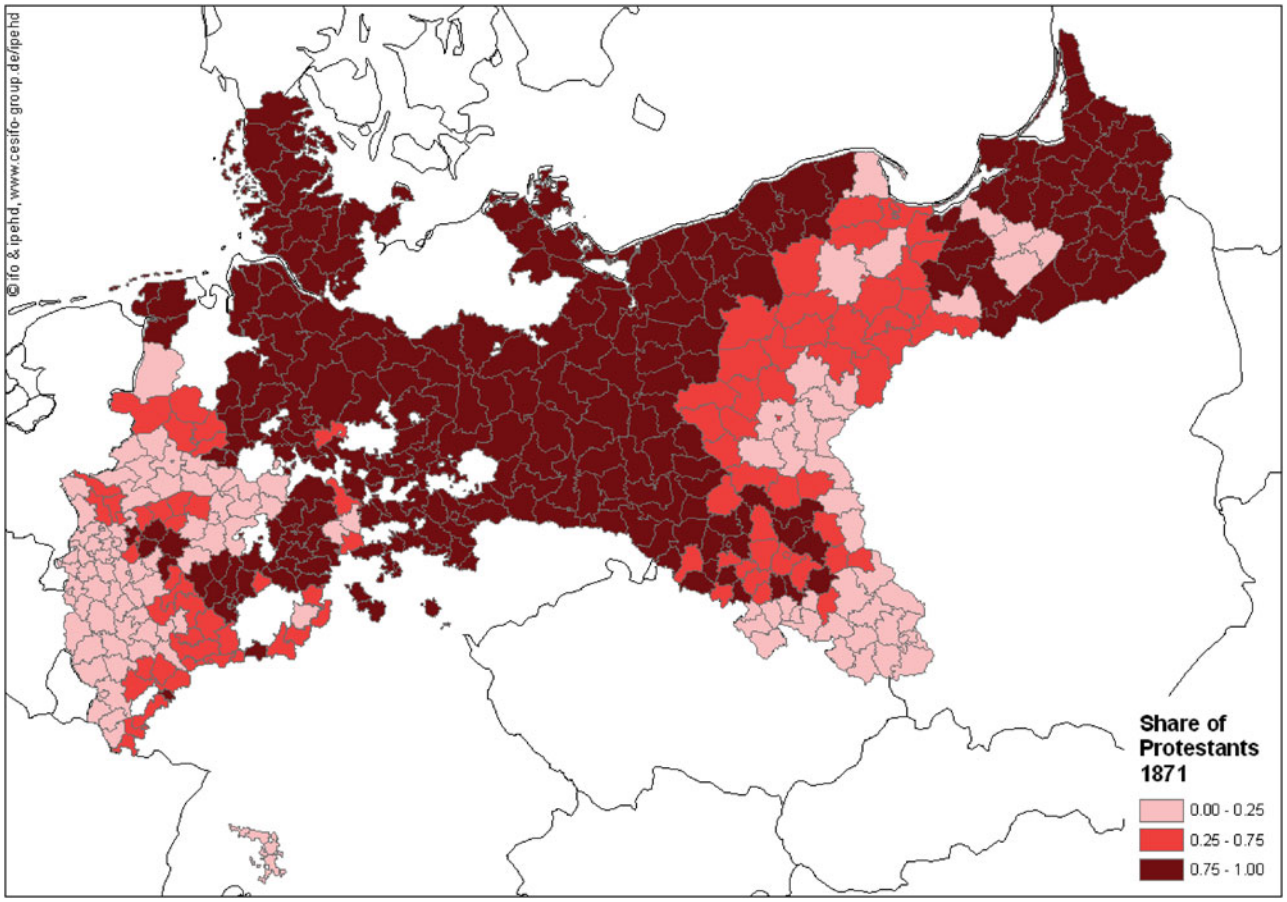


FIGURE 2. Protestantism in nineteenth-century Prussia. ***Note:*** **County-level depiction based on the 1871 Prussian Population Census. For details, see Becker and Woessmann (2009) (color figure available online).**

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014

other topics. This database should facilitate future quantitative research on the economic history of nineteenth-century Prussia.

Although iPEHD provides the service of supplying the historical data in a digitized and structured way and suggests ways on how to merge the data from different sources, researchers must think carefully how to use the data in the context of their specific research projects. For instance, building panel datasets from the different census waves with varying administrative boundaries is a complex task that requires knowledge of the structure of the original data and careful planning. Researchers planning to use the raw data contained in iPEHD should make sure to be well acquainted with the data structure and specifics as described in this documentation.

We hope that iPEHD provides a major service to researchers interested in Prussian economic history. Researchers who use data from iPEHD are kindly requested to cite this article as a source and send an electronic copy of their work to iPEHD@ifo.de.

Acknowledgments

Over the years, a large number of research assistants have contributed to the digitization work underlying iPEHD. We are grateful for their contributions, especially to Christian Steibl, as well as to Rajesh Bhateja. We also thank the editor and an anonymous referee for their helpful comments.

Funding

Financial support by the Pact for Research and Innovation of the Leibniz Association is gratefully acknowledged.

NOTES

1. NAPP contains individual-level data of complete censuses for Canada (1881), Great Britain (1881), Norway (1801, 1865, 1900), Sweden (1900), the United States (1880), and Iceland (1801, 1901). See www.nappdata.org.
2. Additional non-harmonized datasets are available for samples in Germany, Poland-Lithuania, Austria, Italy, Switzerland, and Albania. See www.censusmosaic.org.

3. After we had digitized the data used in Sascha Becker and Ludger Woessmann (2009), the data from that project became available online at www.patrickgalloway.com.

4. We are grateful to Davide Cantoni for pointing us to these data sources at the time.

5. Most of the original historical volumes have been consulted at the Bavarian State Library in Munich, whose cooperation is gratefully acknowledged. Since then, some original statistical sources have been made available online through the Google Books Library Project, including the data sources of 1816 and 1849 (see below).

6. Becker and Woessmann (2011a) provided a non-technical survey that summarized some of the research conducted using the iPEHD data on the effects of the Protestant Reformation on human capital.

7. Work on this article was started in 2006, and the first working-paper version was released in 2007.

8. Going back even further in time, Erik Hornung (2013) combined Huguenot immigration lists from 1700 with firm-level input and output data from 1802 to show substantial long-term effects of the immigration of skilled Huguenots on the productivity of Prussian textile manufactories.

9. Their Prussian geocodes are based on the identifiers used by Patrick Galloway's collection of Prussian data and not harmonized with the iPEHD county identifiers, thus requiring case-by-case adjustments.

REFERENCES

Becker, S. O., F. Cinnirella, and L. Woessmann. 2010. The trade-off between fertility and education: Evidence from before the demographic transition. *Journal of Economic Growth* 15(3):177–204.

———. 2012. The effect of investment in children's education on fertility in 1816 Prussia. *Cliometrica* 6(1):29–44.

———. 2013. Does women's education affect fertility? Evidence from pre-demographic transition Prussia. *European Review of Economic History* 17(1):24–44.

Becker, S. O., E. Hornung, and L. Woessmann. 2011. Education and catch-up in the Industrial Revolution. *American Economic Journal: Macroeconomics* 3(3):92–126.

Becker, S. O., and L. Woessmann. 2008. Luther and the girls: Religious denomination and the female education gap in nineteenth-century Prussia. *Scandinavian Journal of Economics* 110(4):777–805.

———. 2009. Was Weber wrong? A human capital theory of Protestant economic history. *Quarterly Journal of Economics* 124(2):531–96.

———. 2010. The effect of Protestantism on education before the industrialization: Evidence from 1816 Prussia. *Economics Letters* 107(2):224–8.

———. 2011a. The effects of the Protestant Reformation on human capital. In *The Oxford Handbook of the Economics of Religion*, edited by R. M. McCleary, 93–110. Oxford: Oxford University Press.

———. 2011b. *Knocking on heaven's door? Protestantism and suicide*. CESifo Working Paper 3499. Munich: CESifo.

———. 2013. Not the opium of the people: Income and secularization in a panel of Prussian counties. *American Economic Review, Papers and Proceedings* 103(3):539–44.

Cinnirella, F., and E. Hornung. 2011. *Landownership concentration and the expansion of education*. CESifo Working Paper 3603. Munich: CESifo.

Galloway, P. R., E. A. Hammel, and R. D. Lee. 1994. Fertility decline in Prussia, 1875–1910: A pooled cross-section time series analysis. *Population Studies* 48(1):135–58.

Hornung, E. 2013. Immigration and the diffusion of technology: The Huguenot diaspora in Prussia. *American Economic Review* 104(1):84–122.

Minnesota Population Center. 2008. *North Atlantic Population Project: Complete Count Microdata*. Version 2.0 [Machine-readable database]. Minneapolis: Author.

MPIDR (Max Planck Institute for Demographic Research). 2011. *MPIDR Population History GIS Collection* (partly based on Hubatsch and Klein 1975 ff.). Rostock: MPIDR and Chair for Geodesy and Geoinformatics, University of Rostock.

Downloaded by [MPI Psychiatry] at 00:23 05 June 2014