

Curating Historical Statistics Data on Baltic Countries in 1897-1939: Providing Data with Rich Metadata

Problem

Historical statistics is, in general, rather **scarce** (data is available for a limited number of countries and for limited periods of time) and **poorly documented** (sources are only generally described). Historical researchers usually have to rely on the data collected in the “Western world” or analyze rather limited time series (ranging for only 30-50 years).

Historical statistics for **Baltic countries** is even more problematic: statistical data were collected and published during the interwar period, however, these data still are **not properly digitized** as to become readily available for research purposes (in machine-readable formats).

The need for **proper data digitalization** is not just a problem of accurate transfer of available historical data tables into machine-readable formats. **Extensive and precise description of data** (each cell, if necessary) is essential. Therefore, we aim to develop a solution for curating historical data sets.

Presently, most of tabular data are stored in **CSV** files with no metadata. When available, metadata is provided either separately (**DDI**, **SDMX** etc.) or integrated into the CSV file (**CVSY**, **CDS/MRT**, **IPAC** etc.). However, historical data are usually assembled from quite diverse available sources and need to be **extensively documented, even on the individual cell level**, which is not easily possible with currently available standards and solutions.

Solution: + Dataverse

<https://dataverse.harvard.edu/dataverse/HistStatsBaltic>

Harvard Dataverse > Historical Statistics of the Baltic States = Baltijos šalių istorinė statistika

This dataverse hosts historical data about the Baltic states collected in the research project "Historical Sociology of Modern Restorations: A Cross-Time Comparative Study of Post-Communist Transformation in the Baltic States" implemented by a team of Vilnius University researchers lead by prof. Zenonas Norkus. This research is funded by the European Social Fund according to the activity "Improvement of researchers' qualification by implementing world-class R&D projects" of Measure No. 09.3.3-LMT-K-712 "Improvement of scientists", other researchers' and students' qualification by implementing practical scientific activities".

The project was started in January, 2018 and the first data sets were published in June, 2018. No less than 200 new datasets of historical statistics about the Baltic states will be published by the end of the project in January, 2022. The aim of the project is the modernising and modern societies. The fruitfulness of new theory is: Baltic countries during post-communist and interwar periods. The similarity hypotheses which are tested by comparisons with other cases of modern France to the restoration of the Jewish state which is most radical case as because they comprised here the restoration of independent states. The assessment of its outcomes in new way: comparing the relative magnitude method is the scarcity of cross-nationally and cross-time comparable data

Thus, one of the objectives of the project is to collect and publish for free Baltic countries that would allow to compare the progress of these countries

See also [Lithuanian Data Archive for Social Sciences and Humanities \(LIT\)](#)
Visit [Shiny application](#) for data browsing and visualization.

Search this dataverse...

1 to 3 of 3 Results

Number of Primary Schools in Lithuania, 1919-1939 = Pradinių mokyklų skaičius Lietuvoje, 1919-1939 m.
Apr 6, 2020
Norkus, Zenonas
Schools in Latvia Dataverse, V3
This dataset contains data published implementing prc

Description
This dataset contains data on numbers of primary schools in Lithuania in 1919-1939.
Dataset "Numbers of Primary Schools in Lithuania, 1919-1939" was published implementing project "Historical Sociology of Modern Restorations: A Cross-Time Comparative Study of Post-Communist Transformation in the Baltic States" from 2018 to 2022. Project leader is prof. Zenonas Norkus. Project is funded by the European Social Fund according to the activity "Improvement of researchers' qualification by implementing world-class R&D projects of Measure No. 09.3.3-LMT-K-712".
Visit [Shiny application](#) for data browsing and visualization.
(2019-10-17)

Subject
Arts and Humanities; Social Sciences

Keyword
PRIMARY SCHOOLS = PRADINĖS MOKYKLOS, LITHUANIA = LIETUVA

Related Publication
Norkus, Zenonas et al. Numbers of Primary Schools in Lithuania, 1919-1939 = Vidurinių mokyklų skaičius Lietuvoje, 1919-1939 m.: LIDA [distributor], 2019. (Vilnius : Vilnius University, 2019). (HISTEDU: Education). url: [LIDA_HISTEDU_0132](#)

Notes
Sample dataset, only four county data included.

Files **Metadata** Terms Versions

1 File

edu_schools_primary_LIT.tab
Tab-Separated Values - 20.7 KB - Apr 6, 2020 - 3 Downloads
MD5: 4a4713b116108cc2c90e10f32b9c20fd
Tabular data (tab delimited) with metadata before the header (first 6 lines identified with #). Source and notes columns included beside data columns.

Solution: Extended tabular (CSV) data + Shiny

- Dataverse** provides user-friendly solution for storing and publishing data with a possibility to remotely download data. Different types of data can be uploaded and stored on the Dataverse.
- Extended tabular format was developed** (still needs to be developed further), which contains basic dataset-level metadata before the header (column names), and extended cell-level metadata (data source and notes) in the columns beside the data columns (see below). This format can be **easily digested by R/Python**.

	A	B	C	D	E	F	G	H	I	J	
1	# title_en: Number of Primary Schools in Estonia, 1921-1939										
2	# title_it: Pradinių mokyklų skaičius Estijoje, 1921-1939 m.										
3	# abstract_en: <p>This dataset contains data on numbers of primary schools in Estonia in 1921-1939.</p><p>Dataset "Nu										
4	# abstract_it: <p>Šiame duomenų rinkinyje pateikiamas pradinių mokyklų skaičius Estijoje 1921-1939 metais.</p><p>Duc										
5	# notes_en: <p>Extended description is available <a href="http://www.lidata.eu/en/index_search_results_data.php?pid=LID										
6	# notes_it: <p>Spłęstinis aprašas pateiktas <a href="http://www.lidata.eu/index_search_results_data.php?pid=LIDA:hista										
7	Year = Metal Viru Count Sources = Pa Järva Cour Sources = Notes = Pa Harju Cour Sources = Notes = Pa L										
8	1919/20	ND			ND			ND		NE	
9	1920/21	ND			ND			ND		NE	
10	1921/22		233	Eesti Demograafia, Vihk		88	Eesti Demograafia, Vihk		158	Eesti Demograafia, Vihk	
11	1922/23		223	Estonie de Different da		76	Estonie de Different da		164	Estonie de Different da	
12	1923/24		224	Estonie de 1920-1930 P		70	Estonie de 1920-1930 P		164	Estonie de 1920-1930 P	
13	1924/25										
14	1925/26										
15	1926/27										
16	1927/28										
17	1928/29										
18	1929/30										
19	1930/31										
20	1931/32										
21	1932/33										
22	1933/34										
23	1934/35										
24	1935/36										
25	1936/37										
26	1937/38										
27	1938/39										
28											

Historical Statistics of the Baltic States = Baltijos šalių istorinė statistika

Choose a dataset = Pasirinkite duomenų rinkinį

Number of Primary Schools in Lithuania, 1919-1939 = Pradinių mokyklų skaičius Lietuvoje, 1919-1939 m.

Table = Lentelė Plot = Grafikas Descriptions = Aprašai

Table = Lentelė

Table with sources = Lentelė su šaltiniais

Table with notes = Lentelė su pastabomis

Year	Count	Source	Notes
1919/20	ND		NA
1920/21	73		NA
1921/22	76		NA
1922/23	96		NA
1923/24	86		NA
1924/25	92		147
1925/26	106		147
1926/27	107	Primary schools in 1926.	150 Primary schools in 1926.

Important territorial changes:
https://lt.wikipedia.org/wiki/Bir%C5%BE%C5%B3_apskritis,
https://lt.wikipedia.org/wiki/%C5%A0ven%C4%8Dioni%C5%B3_apskritis,
https://lt.wikipedia.org/wiki/Pasvalio_apskritis.

Authors

Vaidas Morkevičius

Kaunas University of Technology, Lithuania

Supported by the funding under the project “RDA Europe Node - Lithuania”



Giedrius Žvaliauskas

Vilnius University, Lithuania

Supported by the funding from the European Social Fund (project number 09.3.3-LMT-K-712-01-0006) under grant agreement with the Research Council of Lithuania (LMT)

