

Database of digital media publications on maternal (family) capital in Russia in 2006–2019

Irina E. Kalabikhina¹, Herman A. Klimenko¹, Evgeny P. Banin^{2, 3},
Ekaterina K. Vorobyeva¹, Anna D. Lameeva¹

1 Lomonosov Moscow State University, Moscow, 119991, Russia

2 Bauman Moscow State Technical University, Moscow, 105005, Russia

3 National Research Center «Kurchatov Institute», Moscow, 123182, Russia

Received 1 December ♦ Accepted 1 December 2021 ♦ Published 7 December 2021

Citation: Kalabikhina IE, Klimenko HA, Banin EP, Vorobyeva EK, Lameeva AD (2021) Database of digital media publications on maternal (family) capital in Russia in 2006–2019. *Population and Economics* 5(4): 21–29. <https://doi.org/10.3897/popecon.5.e78723>

Abstract

The database contains data from publications of digital Russian-language media registered in the Russian Federation on the topic of maternity capital published in the period from May 10, 2006 to June 30, 2019. The database includes general data on publications on maternity capital in .csv formats (UTF-8 encoding). Full texts of publications are presented in .xml format.

A specialized request was generated for the aggregator of publications of Russian-language digital mass media public.ru. In total, the database consists of 457,888 publications of 7,665 publishing houses from 1,251 settlements located in 85 regions of Russia. The database includes information about the date and type of publication, publisher, place of publication (municipality), texts about maternity capital, and numbers of unique positive, negative, and neutral words and phrases according to the *RuSentiLex2017* dictionary, as well as full texts of publications.

Keywords

database, digital media, maternal (family) capital, central and municipal media, Russia, sentiment analysis

JEL codes: J10, J13, Z18

Data format and access

The database consists of full-text publications of digital media on the topic of maternity capital. Materials in Russian have been published in federal, regional, and local digital media.

Publication period: May 10, 2006 to June 30, 2019. The database also contains a number of publications dated 2004–2005 and associated with the mention of family capital (less than 100 items).

The database consists of 457,888 publications of 7,665 publishing houses from 1,251 settlements in Russia on the territory of 85 regions. Data format: .csv, .xml (full texts).

Data access: <https://doi.org/10.5281/zenodo.5740417> (Kalabikhina et al. 2021).

The file “Matkap_SMI_17_11_2021.csv” contains processed information from the extended full-text sample by years (contained in the “XML.rar” archive).

Data collection methodology

The authors used the aggregator of Russian-language digital mass media publications *public.ru*. The selection of publications was limited to the time period from May 10, 2006 (when Russian President Vladimir Putin first announced the maternity capital programme in his message to the Federal Assembly, as one of the mechanisms to stimulate fertility and overcome the demographic crisis, see also (Federal Law... 2006)) to June 30, 2019 (until this date, the programme allowed full uploading of media publications without losses during the period of uploading publications from August 1, 2021 to August 15, 2021).

Key words used to select articles on maternity capital were the following: *matcapital*, *maternity capital*, *family capital*, *paternal capital*. The publication had to contain at least two phrases from the request, while the distance between the phrases had to be no more than 4 sentences. This excluded publications in which the topic of maternity capital was mentioned incidentally, indirectly.

Duplicates were removed from the database. Duplicates related to publications that included a full repetition of the text of the publication itself, together with the name of the publishing house and the municipality (location) of the publishing house. Duplications (reprints) of articles in other publishing houses or in other regions were not excluded.

After lemmatization of the text (as well as after reducing the text to lower case, removing unnecessary spaces, numbers and punctuation), the unique positive, negative, and neutral words and phrases (variables) were counted according to the *RuSentiLex2017* dictionary (Loukachevitch and Levchik 2016). Repetitions of tonal words (stances) were not counted.

Database structure and description of variables

Variables in the database are described in Table 1.

The base of full texts (in the archive file “XML.rar”) contains additional information: titles of publications, the surname and name of the author(s), the name of the region of the Russian Federation where the editorial office of the publishing house is located (the dataset contains publications from 85 regions of the Russian Federation), and some additional information.

Table 1. Variables in the database of digital media publications on maternal (family) capital in Russia (“Matkap_SMI_17_11_2021.csv”)

Column heading	Description and comments
id	Publication identification number (first numbers are assigned to later publications)
pubData	Date of publication (format “YYYY-MM-DD”)
text	Text from “description” on maternal (family) capital in xml-files after lemmatization (removal of punctuation, lowercase and remove punctuation, spaces, numbers)
source	Name of the electronic edition (publisher)
place	Location (municipality) of the Publishing House. In total, the dataset included publications from 1,251 municipalities
type	Types of publications: bulletin, internet resource, newspaper, magazine, internet publication, news agency, press release, radio programme, TV programme
period	Frequency of publication (daily, monthly, weekly, 2 times in week, weekly, quarterly, bi-weekly, other)
positive	The number of unique positive words and phrases from the <i>RuSentiLex2017</i> dictionary
negative	The number of unique negative words and phrases from the <i>RuSentiLex2017</i> dictionary
neutral	The number of unique neutral words and phrases from the <i>RuSentiLex2017</i> dictionary

Distribution of publications by year, type, and publisher

The distribution of publications in the main dataset by year, type, region, and publisher is shown in the Fig. 1 and Tables 2–4.

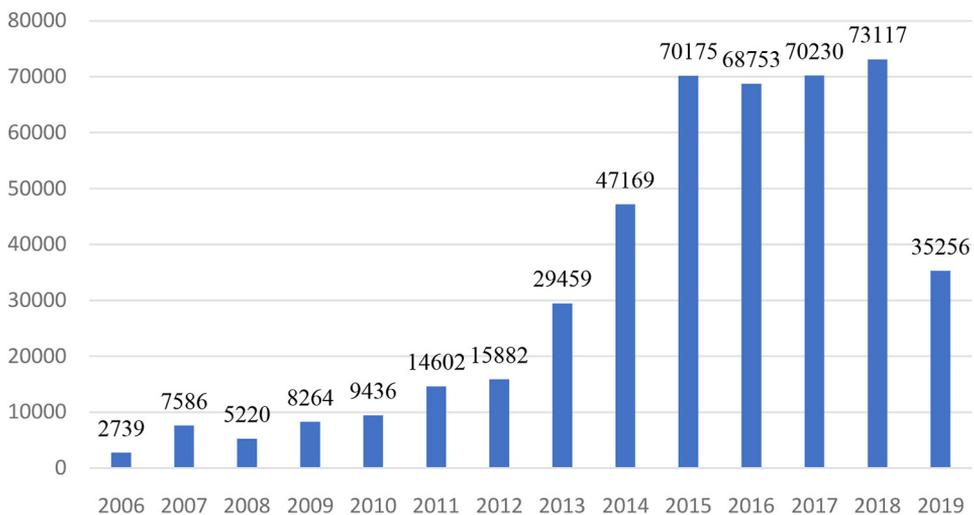


Fig. 1. Number of publications included in the sample, by year. *Note:* The number of publications in 2006 was calculated in the period from May 10, 2006 to December 31, 2006; the number of publications in 2019 was calculated in the period from January 1, 2019 to June 30, 2019.

The largest number of publications was observed in the period from 2015 to 2018. The rapid growth in the number of publications in the period from 2012 to 2015 is associated with an increase in the number of digital media in the Russian Federation, which was caused by the spread of high-speed Internet access across the country.

Table 2. The number and proportion of publications included in the sample, by publication type

Publication type	Number of publications	% of total
Internet resource	323,256	70.60
Newspaper	63,096	13.78
Information agency	31,563	6.89
Internet publication	31,396	6.86
Press release	3,519	0.77
TV programme	2,466	0.54
Magazine	2,245	0.49
Bulletin	323	0.07
Radio programme	24	0.01

The main share of publications on the topic of maternity capital falls on the publication of Internet resources. At the same time, the total share of publications such as press release, TV programme, magazine, bulletin, and radio programme is less than 2% of all publications.

Table 3. Regions with the largest number of media publications on maternity capital (top-15)

Region of the Russian Federation	Number of publications
Moscow	176,627
St. Petersburg	14,082
Sverdlovsk Region	11,843
Chuvash Republic	8,935
Khanty-Mansi Autonomous Okrug	8,673
Republic of Tatarstan	8,663
Rostov Region	7,628
Krasnodar Krai	6,930
Primorsky Krai	6,430
Penza Region	5,957
Chelyabinsk Region	5,870
Moscow Region	5,414
Perm Krai	5,325
Altai Krai	5,235
Krasnoyarsk Krai	5,227

Moscow is leading by a number of publications, followed by St. Petersburg and the Sverdlovsk Region.

Table 4. Number of publications in 20 most common media publishers by number of publications on the topic of maternity capital

Publisher	Number of publications	Publisher	Number of publications
Mngz.ru	5,278	Ttfinance.ru	1,214
Gorodskoyportal.ru/moskva/	4,381	Regions.ru	1,212
IA Regnum	3,095	Publishernews.ru	1,198
Gorodskoyportal.ru/ekaterinburg/	2,570	Chelyabinsk.bezformata.ru	1,187
Socketart.ru	1,970	Podmoskovye.bezformata.ru	1,187
yodda.ru	1,780	Governors.ru	1,130
Rossiyskaya Gazeta	1,724	Ekaterinburg.bezformata.ru	1,126
Gov.cap.ru	1,700	Media-office.ru	1,126
RIA News	1,674	Pskov.bezformata.ru	1,098
Cherkesk.bezformata.ru	1,648	Barnaul.bezformata.ru	1,077

Among the most widespread publishers of publications on the topic of maternity capital, there are both federal, regional and local publishing houses.

Changes in the distribution of publications by type, region, and publisher

Changes in distributions of publications by type, region, and publisher during the period under review is shown in Fig. 2–3 and Tables 5–6.

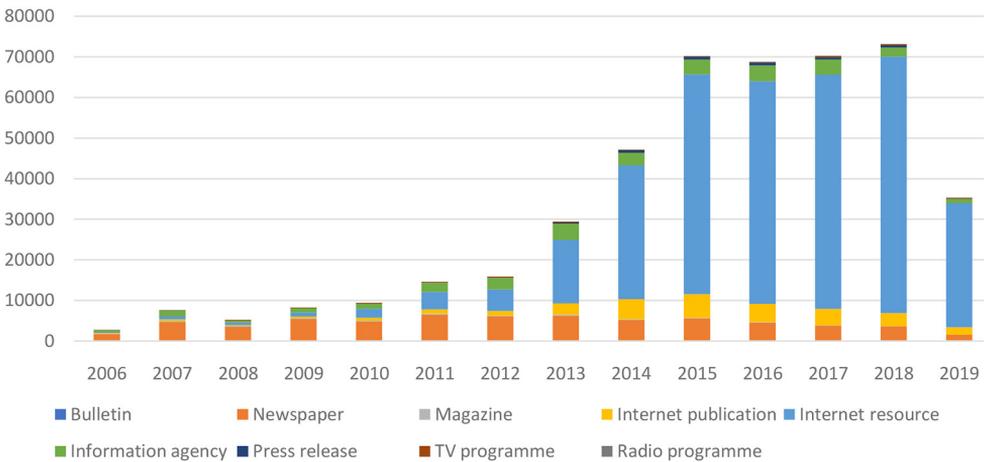


Fig. 2. Change in the distribution of publications by type according to the absolute number of publications for each year. *Note:* The number of publications in 2006 was calculated in the period from May 10, 2006 to December 31, 2006; the number of publications in 2019 was calculated in the period from January 1, 2019 to June 30, 2019.

The main growth in the number of media publications on the topic of maternity capital in 2012–2015 was due to an increase in the number of publications on Internet resources.

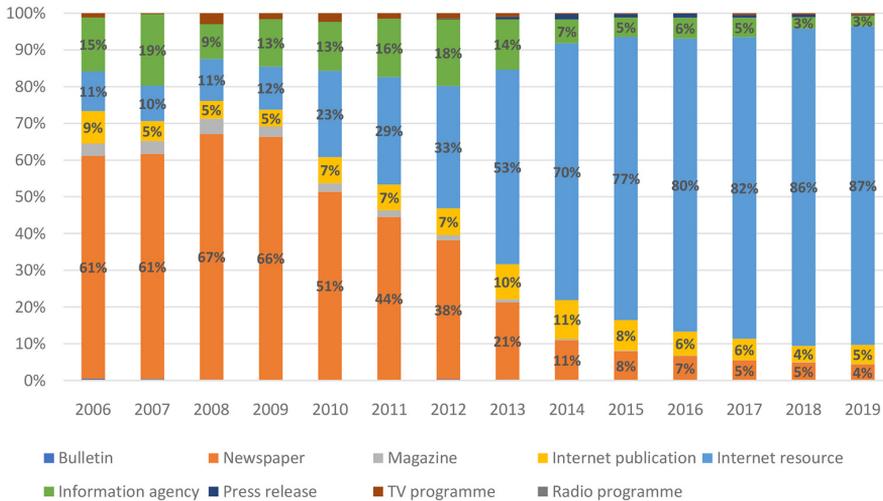


Fig. 3. Change in the distribution of publications by type, relative indicators for each year

The increase in publications of the *Internet resource* type was mainly due to a decrease in the share of newspapers, which went down from 60.68% in 2006 (while the maximum was observed in 2008 — 66.78%) to 4.37% in 2019. Also, during the period under review, the share of publications by news agencies decreased: from 19.5% in 2007 to 2.9% in 2019.

Moscow shows the highest number of publications on the topic of maternity capital over the whole period (Table 5). The share of other regions often demonstrates strong volatility (for example, in the Khanty-Mansi Autonomous Okrug, where the share of publications in the total number of publications decreased from 4.78% in 2017 to 0.77% in 2019), but the share of each region does not exceed 5% annually.

Many publishers demonstrate zero indicators for the year (see Table 6) due to the absence of this publisher on the Internet in a given year. There has also been a major increase in the emergence of publishers on the Internet from 2012 to 2015.

Possible database applications

The data are suitable for analyzing the publication activity of individual Russian publishers (variable “Publishing House/Information Agency”) or publishers of territorial administrative units (variable “Location (municipality) of the Publishing House” for municipalities and “Region of the Publishing House” for regions of the Russian Federation) on the topic of maternity capital. The publication activity of the regions enables analyzing the reaction of the media to events in the field of maternity capital taking place at both the federal and regional or local levels — for example, the reaction of the media to key dates related to the consideration, discussion, adoption, and publication of a draft law at the federal level (Kalabikhina et al. in press). Measuring publication activity enables assessing the mood in the region regarding the demographic event taking place at any level.

Table 5. Distribution of publications by region (share of top-15 regions with the largest total number of media publications), 2006–2019, % in column

Region	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Moscow	45.02	30.62	31.21	26.33	27.65	24.85	26.63	40.12	43.49	40.78	42.33	40.12	36.54	39.28
St. Petersburg	4.82	4.67	2.32	2.69	3.25	3.16	3.58	3.99	3.31	2.25	1.90	3.29	3.64	3.74
Sverdlovsk Region	2.15	2.37	1.30	1.43	2.30	2.64	2.13	2.37	3.00	3.32	2.97	2.25	2.18	2.31
Chuvash Republic	0.40	1.11	0.82	0.83	0.35	0.56	0.68	1.42	3.18	2.72	2.09	2.20	1.74	1.19
Khanty-Mansi Autonomous Okrug	0.99	0.80	1.07	0.93	1.02	0.98	0.93	0.94	1.80	0.58	2.40	4.78	1.72	0.77
Republic of Tatarstan	0.51	2.89	1.13	1.04	1.58	1.49	1.24	1.86	1.88	1.82	1.89	2.32	1.90	1.96
Rostov Region	1.61	2.56	2.61	2.88	2.79	3.24	3.39	1.84	1.46	1.47	1.38	1.46	1.43	1.30
Krasnodar Krai	1.06	1.60	2.80	2.86	2.62	3.05	2.86	2.21	2.18	1.43	1.08	1.12	0.90	1.08
Primorsky Krai	1.83	2.68	1.95	1.67	1.59	3.17	3.60	1.64	0.99	0.99	1.28	1.23	1.14	1.50
Penza Region	0.47	0.26	0.25	0.48	0.39	0.15	0.12	0.57	1.54	1.97	1.46	1.15	1.69	1.31
Chelyabinsk Region	0.51	1.40	1.86	3.34	2.16	1.34	1.53	1.60	1.17	1.28	0.98	1.13	1.11	1.50
Moscow Region	0.51	1.04	0.98	0.65	0.48	0.95	1.11	0.56	0.91	1.21	1.32	0.76	1.80	1.86
Perm Krai	1.83	1.67	2.28	1.56	1.38	1.34	1.65	1.32	1.48	1.26	1.29	0.94	0.75	0.69
Altai Krai	1.06	1.53	0.69	1.16	1.48	1.76	1.34	1.09	0.88	0.81	1.00	1.37	1.27	1.32
Krasnoyarsk Krai	1.10	1.16	1.02	1.32	1.95	3.42	1.64	1.14	1.10	1.11	1.04	0.92	0.95	0.87

Table 6. Change in the distribution of publications by publisher (20 publishers with the largest total number of publications) by the absolute number of publications for each year

Publishers	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Mngz.ru	0	0	0	0	0	0	26	53	620	0	1,091	2,899	589	0
Gorodskoyportal.ru/moskva/	0	0	0	0	0	0	0	0	353	823	1,469	1,624	80	32
IA Regnum	70	444	182	406	198	268	328	374	137	221	199	149	61	58
Gorodskoyportal.ru/ekaterinburg/	0	0	0	0	0	0	0	0	343	824	814	163	289	137
Socket.ru	0	0	0	0	0	0	94	390	630	620	234	2	0	0
yodda.ru	0	0	0	0	0	0	0	0	0	0	968	715	97	0
Rossiyskaya Gazeta	87	140	109	124	120	142	121	144	129	152	142	126	121	67
Gov.cap.ru	0	0	0	0	0	0	0	164	561	582	165	162	66	0
RIA News	0	0	0	0	0	212	318	364	156	79	89	212	149	95
Cherkesk.bezformata.ru	0	0	0	0	0	0	0	0	0	101	346	411	539	251
Ttfinance.ru	0	0	0	0	0	0	0	5	25	49	403	482	209	41
Regions.ru	55	161	72	112	81	94	103	131	136	85	55	66	52	9
Publishernews.ru	0	0	0	2	1	3	7	85	183	175	251	195	251	45
Chelyabinsk.bezformata.ru	0	0	0	0	0	0	0	0	0	88	140	253	413	293
Podmoskovye.bezformata.ru	0	0	0	0	0	0	0	0	0	45	106	217	567	252
Governors.ru	0	0	0	0	0	0	0	330	307	317	176	0	0	0
Ekaterinburg.bezformata.ru	0	0	0	0	0	0	0	0	0	61	127	258	445	235
Media-office.ru	0	0	0	0	0	0	0	0	280	273	242	194	108	29
Pskov.bezformata.ru	0	0	0	0	0	0	0	0	0	79	212	264	404	139
Barnaul.bezformata.ru	0	0	0	0	0	0	0	0	0	60	127	263	380	247

Sentiment analysis is performed for the elements of the entire database, which allows to study the sentiment of various groups of publications.

In the next versions, we plan to expand the database of media publications for other queries and topics related to fertility policy, namely, family benefits and maternity leave.

Reference list

- Loukachevitch N, Levchik A (2016) Creating a General Russian Sentiment Lexicon. In: N.Calzolari, K.Choukri et al. (eds.) Proceedings of the Tenth International Language Resources and Evaluation Conference (LREC-2016), Portorož (Slovenia), May 2016. URL: <https://aclanthology.org/L16-1186.pdf>
- Kalabikhina IE, Klimenko GA, Banin EP, Vorobieva EK, Lameeva AD (2021) Database of digital media publications on maternal (family) capital in Russia in 2006-2019 [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5740417>

Other data sources

Federal Law (2006) “On additional measures of state support for families with children” of December 29, 2006 No. 256-FL (with amendments and annexes). URL: <https://base.garant.ru/12151286/> (accessed November 15, 2021)

Information about the authors

- Irina Evgenievna Kalabikhina, Doctor of Sciences (Economics), Professor, Head of the Department of Population, Faculty of Economics, Scientific and Educational School “Brain, Cognitive Systems, Artificial Intelligence”, Lomonosov Moscow State University, kalabikhina@econ.msu.ru
- Evgeny Petrovich Banin, Candidate of Engineering Sciences, Research Engineer, Research Center “Kurchatov Institute”, Bauman Moscow State Technical University, evg.banin@gmail.com
- German Andreevich Klimenko, postgraduate student, engineer at the Research Laboratory of Population Economics and Demography, Faculty of Economics, Lomonosov Moscow State University, german89000@mail.ru
- Ekaterina Kirillovna Vorobieva, graduate student, Faculty of Economics, Lomonosov Moscow State University, ekaterina.vrb@mail.ru
- Anna Dmitrievna Lameeva, graduate student, Faculty of Economics, Lomonosov Moscow State University, lameevaanna@mail.ru