

The digitized Newspaper Collection as National Patrimony of the Russian Federation

A.A. Dzhigo

Ph.D, Head, Research Department of Library Science
Russian State Library
Moscow, Russian Federation



Copyright © 2013 by **A.A. Dzhigo**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

Abstract:

Current issues of digitization newspapers' collections are in the focus of this topic. The sharing of responsibilities for the digitization of newspapers, the selection criteria and the digitization priorities are discussed. Particularly the definition of the scanning method (in the library or outsourced) and the solutions for the best representation of newspapers either as artifacts or as information set are of a great importance. There was provided a detailed analysis of the activities of the federal and regional libraries, newspapers publishers and commercial information centers devoted to the digitization of the national printed newspapers. The emphasis was made to the problems of practical protection of copyright and intellectual property in digitization of the periodicals and in the usage of these materials by the institutions including libraries.

The newspaper collection of the libraries of the Russian Federation represent the history of the development of human thought, progress, culture, and education. The history of the establishment and development of newspapers is filled with profound dramatic events. The social role was defined and continues to be defined by those who own the material basis of the press and the means of its distribution. This is confirmed by the entire history of the development of the Russian press and, in particular, the Soviet period. After the October Revolution, newspapers were transformed into weapons of propaganda, agitation, ideology, the political and economic education of the people. The table below shows the steady growth of the number of newspaper titles from the beginning of the 18th century to the present day¹. It is very important to note that the newspaper collection includes newspapers published in the 59 languages of the

¹ <http://www.bookchamber.ru>

peoples of the Russian Federation. One of the particular features of the newspaper collection is the large number of large circulation newspapers published by various enterprises, organizations, and educational institutions. On the whole it can be suggested that the total collection of print newspapers of the libraries of the Russian Federation for the 20th century alone amounts to more than 8 million annual cumulative sets, comprising 24 million inventory units.

Table 1.

Numerical indicators of newspaper publishing in Russia

No.	Historical Period	Number of titles
1.	From the beginning of the 18th C.- to the beginning of the 19th C	32 titles
2.	19th - to the beginning of the 20th C. (1917)	approx. 380 titles
3.	October 1917 - to 1991	approx. 3,500 titles
4.	From 1992 to the present	approx. 11,000 titles

The historical significance of newspapers and the large paper formats clearly show the necessity of digitizing the Russian newspaper collection. The main problems facing Russian specialists in the digitization of newspapers are the following:

- distribution of responsibility for digitizing newspapers;
- criteria for selection and establishing priorities for digitization;
- search for best solutions (best practices) (e.g., scanning in-house or outsourcing);
- forms of representing newspapers (as artifacts or only information blocks).

At the present time the following bodies are involved in the transfer of print newspapers into electronic formats:

- on the Federal level - the national libraries (Russian State Library, Russian National Library, the B. Eltsin Presidential Library);
- on the regional level - the libraries of the subjects of the Russian Federation (national libraries of republics of the Russian Federation, regional and oblast levels and other libraries);
- individual commercial information centers.

The main digitized collections are the newspapers published from the 19th to the beginning of the 20th centuries. This is linked to issues of copyright and intellectual property protection. According to the Civil Code of the Russian Federation, intellectual property is protected for 70 years². Publishers of newspapers own these rights exclusively: only they, on their own terms, in contractually agreed upon conditions for the abrogation of these exclusive rights or licensing agreements, can allow other legal bodies, including libraries, the use of the results of intellectual work.

² Civil code of the R.F. Chapter 69, pp.1225-1229. Moscow: Status, 2013. 685p.

Table 2

Quantitative indicators of digitized newspapers

No	Organization involved in digitizing newspapers	Number of titles of digitized newspapers	Time span
1.	National libraries of the RF (RSL, RNL, EPL)	approx. 220	18-early 20 th
2.	Libraries of subjects of RF (republics of the RF, other local)	366	18C-1960s
3.	Commercial organizations, newspaper publishers	1123	From establishment with of newspapers to present

At the present time the Russian State Library along with the Russian National Library has prepared the electronic versions of the first all-Russian newspapers of the 18th - 19th centuries. The newspaper "Vedomosti" for 1703 to 1727 presents the greatest interest to readers. No single library in the country has a complete set of this newspaper. The electronic versions of the newspapers "Kino" (1925-1940) and "Soviet Sport" (1922-1945) found in the Russian State Library are in high reader demand.

The priority plan for the digitization of newspaper sets from the Russian State Library and the Russian National Library include print newspapers from the 19th century to the beginning of the 20th century which enjoy a sustained mass demand (Northern Bee, The Petersburg Sheet, New Time, and so on). In addition, included in the listing for digitization are newspapers of the 19th century which are represented in the collections of both libraries in comprehensive runs. Considering unusual reader demands, newspapers of the period of the revolution and civil war in Russia, especially the collection of "non-soviet newspapers" are also included in the plans for digitization. In toto, transfer to electronic formats of "guberniia" news of the 19th and early 20th centuries, accompanied by detailed bibliographic records, is also being planned. But for work with this given block, regional library newspapers services must be called upon.

The Presidential Library named for Boris Eltsin is digitizing the newspapers of the early years of the Soviet government, newspapers of the war period and periodicals of the "White" emigration. The digitization of more than 200 titles is being planned. Among them -- "Will of the people", "Land and Will", "Peasant and worker"; among those titles published outside the border of the Soviet Union by the representatives of the first wave of emigration, -- the newspapers "New Russian Word" (New York), "New World (Berlin), "Russian Gazette" (Paris), The Shanghai Dawn (Shanghai) and others. The mass media information sources of that period are very much in demand by readers. They reflect the social mood of the country and can be useful in research on historical journalism.

Of the 82 libraries of the subjects of the Russian Federation (national libraries of republics that are part of the Russian Federation, regional, "oblast" and other libraries), only 38 are involved in the digitization of newspaper collections. In the period from 2005 to 2013, 366 regional newspaper titles have been digitized. Territorially, these cover local republic newspapers in Russian and the national language (excluding Russian), regional, oblast, city and large circulation (factory, education institutions and others) newspapers. The chronological framework is wider than the electronic newspaper

versions of the national libraries of Russia. These are not only the newspapers of the 18th - to the early 20th centuries. The electronic newspaper collections of the regional libraries include newspapers of the entire 20th century. The digitization of newspapers on the regional level is accomplished by various methods:

- individual efforts of the libraries;
- fulfillment of orders for digitization of “kraevedcheskie (regional studies) newspapers from large federal libraries;
- cooperation of the library and the editors of individual regional newspapers.

The greatest interest in the above-named methods is in the cooperative work of the library and the editors of the newspapers. This is accomplished by a licensing agreement covering access, as a rule, to certain newspapers in the informational-telecommunications network. There are two forms of usage. Practically the entire electronic newspaper collection up to the 1950's is housed on a library server on a charge-free basis. The use of the electronic versions of the newspaper from 1951 to the present is covered by a licensing agreement for a given period on a fee-for-service basis. For example, the Belgorod State University Research library site hosts local newspapers (approximately 20 titles of oblast, city, and regional newspapers) in its “Newspapers of the Oblast” data base³. Licensing agreements are made with the editors of the newspapers, according to which they (the editors) extend rights to the library for the use of the electronic versions of its newspapers.

In our national practice, as a rule, OCR (Optical character recognition) is used to create the digital image. The quality of the scanned text is by and large dependent on the condition of the individual issue of the newspaper. The degree of exactness depends on the OCR program used and is usually within the limits of 68% (without correction) to 99.8% (manual correction). The quality of machine readable texts of newspapers improves if some text blocks are identified before recognition. Here there are 3 successive stages: zoning, recognition, and segmentation. With zoning the page is analyzed for the purpose of identifying all the elements, such as horizontal and vertical lines, text blocks, illustrations. Each of these elements is defined and assigned group characteristics. Then, recognition of text blocks takes place. During this operation the position of each word and symbol is recorded in special “square” coordinates. And in the final stage -- segmentation -- the results of the analysis of topology and recognition are joined in order to distinguish various objects of the page, article, paragraphs of the articles, etc.

In the sector of information services offered to libraries in the sphere of newspaper resources, commercial organizations occupy their own definite place. However, one can not observe purposeful and large scale work in the digitization of newspapers. As a rule, user demand for digital formats are for individual issues of a newspaper, annual cumulation, individual articles are scanned, and very rarely is the digitization of the newspaper for the entire period of its existence requested by the editors of the

³ <http://bgunb/Corporation/Info.aspx?r i=5>

newspapers. The exception is the activity of international firms such as “East View” (its subsidiary is the Russian organization EVIS) and Integrum World Wide.

In the past 5 years through the efforts of East View, electronic archives of the largest Russian newspapers have been created: “Pravda (1912-2012), “Izvestiia” (1917-2012), “Argumenty i fakty” (Arguments and facts) (1983-2012), “Literaturnaia gazeta (Literary gazette) (1929-2012). The data bases of the above-named newspapers contain more than 350,000 newspaper pages, accessible online, from any browser and any operating system (Windows, MacOS, Linus). The modern technology of “digital microfilm” has been used in the digitization of the archives of these newspapers. All the graphic images have been scanned from existing microfilms, after which the images were optimized and the text recognized. For more recent issues, existing electronic versions of the newspaper issues were used. Along with the graphic representation of the newspaper lines in pdf format, the electronic version also contains the textual layer. This allows for not only the viewing of the issues of the newspaper, but also allows for key-word searches with highlighting of the results of the search.

The Integrum World Wide electronic archive contains about 270 full text versions of central and about 700 full text versions of regional newspapers. A search for the desired information or individual newspaper article is achieved by key-word search, through the use of a paradigm search for Russian, English, German, and Ukrainian language words, by article headlines, by authorized sources, as well as by geographical and time parameters.

East View and Integrum World Wide offer their services on a commercial basis. The licensing agreement is the principal method for Russian library acquisitions. Its product is the provision of access to networked newspaper resources. However, the license never assumes that the library has the right to store these resources, preserve them, or provide long-term access to them. Access to such electronic newspapers in libraries usually is provided for a time period determined in the license. The commercial entity stipulates the conditions of access to the server, but does not transfer the rights for permanent use to the library.