

Aleksandra Fostikov,
Neven Isailović
The Institute of History, Belgrade

**DIGITIZATION OF THE HISTORICAL PERIODICAL
PUBLICATIONS – IČ, BIG, MG – WITH REMARKS
ABOUT SPECIFICITIES AND POSSIBILITIES***

Abstract: This paper focuses on the digitization of the historical Periodical Publications - IČ, BIG and MG and work on it. Special attention is paid to specificities of these periodicals as examples of periodicals in history and humanities. In addition, it includes a few remarks about issues of digitization and presents some possibilities concerning this process. At the end, there is a short addition about markup languages with observations from users (historians) point of view.

Key words: periodicals, digitization, history, internet/intranet, markup language.

Projects concerning digitization of the historical periodical publications, which are carried out within the institutions, have a focus on the first step in digitization in its prime meaning. In order to speed up the digitalization process, which is one of the essential prerequisites for the progress of humanities in this region, the initial efforts in this direction did not fully follow the highest standards – they have mostly included just scanning of the text and its conversion to the files in PDF format, usually raw, without detailed text-search features. As a result, even keywords are not fully searchable. A step forward in the process of digitization was obtained through an agreement, which defines what a scientific text must contain in order to make the digital works searchable, at least on a basic level.¹ In the meantime, almost all periodicals uploaded on the Internet are still being presented as a single big file, which usually contains an entire volume.² Databases like SCIndeks or eLib (or e.g. CEEOL), offer this big file split into more files, where each of them represents one paper with its own necessary additional data, like abstract, volume etc.³

However, for now, just a small portion of that material is available online. Having in mind the problem of the deficiency of digitized material the Institute of History in Belgrade and the Department of History at the University of Belgrade, have begun preparing their own periodicals in electronic form with the goal to upload them in several databases. This process includes three publications for now, two of them issued by the Institute: Istorijski časopis –

* The paper is a result of research on the projects of Ministry of Education, Science and Technological Development of the Republic of Serbia № 177010 and 177029.

¹ Even big projects, like e.g. SCIndeks (<http://scindeks.ceon.rs/Default.aspx?lang=en>) still do not work in full search capacities. Aleksandra Fostikov, Neven Isailović, *Digital Humanities or digital versus humanities*, Review of the National Center for Digitization (in print). Also, see below. The main advantage of the new editions is the fact that they could be put on the Internet right after final preparation of the text for print, before it goes to print, because they are now being prepared in electronic form. As a result, the PDF is usually ready for text searching, and all what it needs is the OCR (it is not a complicated procedure for usual fonts and documents which are already prepared in electronic format).

² Those are usually uploaded by the institutions or anonymously by some academic torrent sites. The exception is the web site dedicated to the publications of the Department of Ethnology and Anthropology of the Faculty of Philosophy at the University of Belgrade <http://www.anthroserbia.org/>.

³ The same situation is generally in all big databases like JSTOR, Science Direct etc.

Historical Review (IČ) and Mešovita građa – Miscellanea (MG), and one by the Department of History: Beogradski istorijski glasnik – Belgrade Historical Review (BIG).⁴

Work in progress: Digitization of the Historical Periodical Publications– IČ, BIG, MG

The Institute of History first began the digitization of its own periodicals only for the singular purpose of having them stored in the electronic edition (especially older numbers) – of course in raw form, big PDFs, like most institutions. Aiming to have those editions not only stored but also fully accessible and searchable over Internet, possibly in open access mode, Institute started cooperation with Serbian Citation Index (<http://scindeks.ceon.rs/>) and CEEOL (Central and Eastern European Online Library) (<http://www.ceeol.com/>). Also, the Institute of History and the Department of History, began cooperation with eLib(rary) of the MISANU (Mathematical Institute of the Serbian Academy of Sciences and Arts) (<http://www.mi.sanu.ac.rs/>). During this work we noticed a few specificities concerning the process of preparing publications which depend on the database (see below), as well as the fact that these are all historical publications.

A service of the Centre for Evaluation in Education and Science (CEON/CEES) (<http://scindeks.ceon.rs/>), is a part of the integrated system of scientific bibliographic information of Serbia. Organized in cooperation with the National Library of Serbia, it is financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia. This service has several functions: to monitor the country's scientific production in terms of serial publications, make citation indexes, do statistical reports on leading journals and authors of scholarly papers published in these journals, and, for some of the leading serial publications, provide free access to complete text articles and reviews. In cooperation with the main NGO, the one responsible for the implementation of this program, SCIndeks is making a list of information concerning periodicals and gathering the vital data needed for making bibliometric reports, similarly to the network Web of Science (WoS). In this kind of cooperation, process of digitization was the easiest – the raw PDF was given away on CD to the National Library of Serbia, and SCIndeks did all of the following steps in preparation.

For researchers in the field of humanities it is important that this service provides individual and journals' citation indexes, open access to the articles and reviews in some of the journals, access to the bibliometric reports for each of the included national journals, searchability of bibliographic data, titles, keywords, abstracts, and in some cases, fully searchable articles. SCIndeks e-Ur service could also be very important (this service is based on the paid subscription). It allows new possibilities, like *online-first* publishing, automatic transfer of metadata to SCIndeks and other databases that support international protocols for data transfer or *CrossCheck* (control for originality of submitted papers, to prevent plagiarism and duplication of publications).⁵

CEEOL is a library of electronic versions of scientific serial publications, recently enriched by an e-library devoted to monographic publications (<http://www.dibido.eu/>). It was launched in order to promote the editions in the scholarly field of humanities in Central and Eastern Europe. Although this is not a free service but a portal based on the principle of

⁴ In the team, which is working on uploading part of digitization for now, are Slavica Merenik, Neven Isailović, Aleksandra Fostikov, Slobodan Simić and Jelena Mrgić (preparing electronic editions for print and uploading them for the Institute and for the Department).

⁵ http://ceon.rs/index.php?option=com_content&task=view&id=246&Itemid=73

subscription, it is very important because it includes many generally unavailable regional and local journals, and the institutional subscription price is significantly lower than the cost of other databases and e-libraries which are considered more significant at the world level.

For researchers in the field of humanities it is important that this base provides the following services: complete texts of articles, reviews and other material published in journals, and, to those who signed a cooperation agreement with the CEEOL, also a full searchability of articles and access to rare editions. There are numerous advantages for journal publishers in the field of humanities too. Each magazine can be published free-of-charge in the CEEOL e-library, and there is a possibility of earning money on the digitized articles. In the process of digitizing, data are being entered into the MDB databases, which are packed, along with the PDF versions of the OCR-ed articles, in the CAB archives. A user friendly interface is included, so the MDB databases are not filled directly, but via an EXE application that contains various fields (e.g. name of the article, author, address, name of the journal, volume of the journal, language, number of pages and their range, scholarly discipline of the article, keywords, abstract/summary etc.). The publisher receives an annual report containing the specific information of the downloaded articles such as: which articles were downloaded, how many times and how much money they earned. With this information available, we can get a notion of "popularity" of certain research topics and researchers who were working on them. In addition, we can check out the hypothesis of the importance of foreign language and "modernity" of topics for journals' rank on national and international lists. We can also determine if there is a tendency of growth or decline concerning the general interest in the journal, authors or topics.

Cooperation with eLibrary or electronic base of publications of Mathematical Institute of the Serbian Academy of Sciences and Arts (eLib MISANU) for now is, maybe, the most difficult one (the friendly interface is in the preparation process at the moment and should be put to use soon). After a detailed process of cutting of the raw PDF and preparing the text files (.cap) with all the data in LaTeX processor, it will go to verification and then through to upload.⁶ Nevertheless, this fact allowed us to learn LaTeX typesetting, and its advantages.⁷

For researchers in the field of humanities it is important that this base provides open access and as soon as the new interface becomes available, it will be possible for users to independently update the pages of their publications for free. In this way, the digitization of historical publications could be intensified.⁸ At the same time, those historical periodicals will be available through eLib on the *Сербиа-Форум* (<http://serbia-forum.mi.sanu.ac.rs/>)⁹ In addition, this cooperation will also include the cooperation with NCD (<http://www.ncd.org.rs/>) and especially with their project named Internet Library of Serbia for Arts and Humanities – IbiS (http://ibis.mi.sanu.ac.rs/index_e.html).¹⁰

⁶ <http://elib.mi.sanu.ac.rs/pages/about.php>, Tatjana Jakšić, Aleksandar Pejović: *Baza elektronskih časopisa*, http://www.ncd.org.rs/ncd_sr/Events/NCDbeograd2012/KnjigaApstrakata2012.pdf, pp.12; http://poincare.matf.bg.ac.rs/~ncd/Beograd_2012_Prezentacije/TJaksic_Elib_NCD-konferencija.pdf (20.03.2013).

⁷ See the next section: *About Specificities and Possibilities in the Digitization Process of the Historical Periodical Publications*.

⁸ There is an idea to include some other historical periodicals like Stari srpski arhiv (SSA) or Braničevski glasnik in the cooperation with MISANU's eLibrary.

⁹ <http://serbia-forum.mi.sanu.ac.rs/>

¹⁰ This project has been launched by The Association of Researchers of the Institutes for Humanities and NCD. It should be known that, in the future, IbiS could represent the framework for e-library of digital humanities in Serbia. As it says on its homepage: *The goal of the IbiS is to present the Serbian scientific and cultural legacy as well as present-day scientific production in the fields of humanities and related disciplines to a broad audience -*

About Specificities and Possibilities in the Digitization Process of the Historical Periodical Publications

Working on digitization of the above-mentioned periodicals, especially on the preparation of the cap. for eLib, highlighted some specificities in two ways¹¹: on the one hand there are specificities concerning external elements, and on the other there is a fact that some articles could be described as primary sources (for earlier periods of history) or as critical editions of written sources.

In reviewing the external elements, there are a few details such as the fact that earlier editions usually do not have any standard classification like keywords, abstract etc. For this group of journals, after making the raw PDF of an article, the next step in digitization should be done by librarians or researchers, who must write in the missing lines. In some cases, we use summaries instead of abstracts of these articles for these purposes. Also, some elements, like the UDC (especially in newest editions), could be more complex and longer for history articles and of course it needs more space and components than, for an example, in mathematics.¹² Then there could be a problem with keywords, which in history or other humanities may not be a good way of categorization because it is usually very extensive in description and thus not specific enough to be used for the research of the data. For example, in one of the articles, the keywords like *Serbia, 14th Century*, etc... are so extensively descriptive leaving only a personal name indicative for future researching by keywords. Therefore, it is necessary to make full (re)searchability of texts as soon as possible for the purpose of finding specific historical fact, or making a solid base for future researching and, eventually, one day, making a change in our view of history.

The question in which way some documents could be prepared for the database, is also interesting. In the case of eLib, we worked in LaTeX instead in MS Word, and, it must be said, that this tool is remarkable. Beside other great benefits, it is also very useful for Multilingual Typesetting.¹³ Moreover, as LaTeX is remarkable document preparation system, it goes also for TEI when it comes to text encoding in humanities. In addition, it is the exactly existence of TEI that raises the question of how it is possible to digitize some specific type of historical periodicals. Because, among others, there is one group of the older historical periodicals whose articles could represent source of its time (like *Glasnik Društva srpske slovesnosti*)¹⁴ and another group of publications, which represent critical editions of written sources from earlier periods of history, like MG or *Stari srpski arhiv* (SSA). For now, both of the mentioned types are digitized like any other historical periodicals, but, it is important to mention that in the future some of those types like MG or SSA could also be a solid base for

to national and international professionals and scientists. Digitization of national scientific and cultural legacy and current scientific production is highly important not only for facilitating the access to these assets and for the international scientific exchange, but also for preservation of the cultural identity, education of the local public and thus ultimately contributing to the diversity of the European cultural map.

¹¹ The usual problems with full OCR of texts will not be included here since some of them must be retyped almost in full for making full text search available.

¹² Moreover, of course that most of the databases are prepared by researchers from natural sciences. This problem is also obvious in the interface for the reviewing of the projects.

¹³ Johannes Braams, *Babel, a multilingual package for use with Latex's standard document classes*, Printed April 1, 2005; Apostolos Syropoulos, Antonis Tsolomitis, Nick Sofroniou, *Digital typography using Latex* (With 68 Illustrations), 2003 Springer-Verlag New York, Inc., ISBN 0-387-95217-9, pp. 301 i dalje. On multilingual typesetting also see: StevePeter, *TEX and Linguistics*, TUGboat, Volume 25 (2004), No. 1— Proceedings of the Practical TEX 2004 Conference, 58–62. <http://www.tug.org/TUGboat/tb25-1/peter.pdf> (25.03. 2013.)

¹⁴ <http://www.sanu.ac.rs/English/Clanstvo/Istorija.aspx?arg=1> (25.03. 2013.)

creating new digital repository of primary sources for Serbian and history of the Balkans, encoded with TEI.

ДИГИТАЛИЗАЦИЈА ИСТОРИЈСКИХ ЧАСОПИСА - ИЧ, МГ, БИГ-СА ОСВРТОМ НА СПЕЦИФИЧНОСТИ И МОГУЋНОСТИ

Досадашња дигитализација историјске периодике углавном се сводила на очување периодике, а посебно старије грађе, у електронском облику. За ову сврху прављени су углавном сирови и велики PDF-ови, који су у том облику и били постављани на интернет ради какве такве доступности и претраживости. Како би се дигитализација убрзала, а самим тим и омогућио приступ и претраживање периодике путем интернета, по могућству у *open access mode*, поједине институције ступиле су у сарадњу са виртуелним базама (библиотекама). Између осталих, дигитализацији су приступили и Историјски институт Београд (ИИБ), са часописима Историјски часопис (ИЧ) и Мешовита грађа (МГ), као и Одељење за историју Филозофског факултета, са својим гласилом Београдски историјски гласник (БИГ). Током овог процеса ИИБ је започео сарадњу са Српским цитатним индексом (<http://scindeks.ceon.rs/>) и Онлајн Библиотеком за централну и источну Европу (CEEOL - Central and Eastern European Online Library) (<http://www.ceeol.com/>), као и са Електронском библиотеком Математичког института САНУ (МИСАНУ) (<http://www.mi.sanu.ac.rs/>), са којом је сарадњу отпочело и Одељење за историју.

Како свака од поменутих база (електронских библиотека) има себи својствене карактеристике и могућности, сарадња са истим омогућила нам је да сагледамо дигитализацију историјске периодике из више углова. Иако у првом тренутку можда најкомпликованија, услед тренутно непостојећег лако приступног формулара (који ће ускоро стартовати), сарадња са eLib МИСАНУ омогућила нам је да увидимо вишејезично сетовање у LaTeX-у, који се испоставио изванредним у припреми самог текстуалног фајла за историјску периодичку, исто онолико колико је ТЕИ стандард погодан за опис докумената.

Додатно увиђање различитих могућности које пружа ТЕИ стандард за опис довело нас је и до питања на који начин би било правилно дигитализовати поједине историјске часописе попут Старог српског архива (ССА) који заправо издаје критички приређену грађу или неких старијих часописа попут Гласника Друштва српске словесности, чији одређени радови заправо представљају извор за своје време. Сви часописи овог типа за сада се дигитализују на уобичајни начин, иако би могли да представљају основу будуће библиотеке примарних извора за историју Србије.

На крају, неопходно је да додамо да је у будућности нужно да текстови у целини буду претраживи, како би се заиста омогућио приступ и претраживање сваке појединости у оквиру базе, те да би то требало да буде даљи правац дигитализације. Кључне речи у историјској науци често нису тако прецизне и прешироког су спектра, а претраживање путем њих не даје очекиване резултате.

aleks.fostikov@gmail.com

neven.isailovic@gmail.com