

NATIONAL LEGAL HERITAGE: LEGAL DATABASE PROVIDING ACCESS TO LEGAL PAST

Jan Kober*, Ján Matejka**, Petr Aubrecht***, David Brůha****¹

Abstract: *This paper focuses on the issue of making historical legal literature in the Czech Republic accessible. For historical and structural reasons, the accessibility of old legal literature in Czech libraries is inadequate. One possible solution to this long-term problem is the systematic digitisation of this old literature. The primary objective would not be preservation digitisation, otherwise the most common type in library digitisation projects, but rather both supplementary digitisation intended to supplement the book and periodicals fund with additional titles, and also 'opening digitisation', meaning digitisation to enable new ways of working with texts and information. The paper also concisely introduces the basic principles and technological, software and licence parameters of the National Legal Heritage project, which is an attempt at the practical application of these conclusions. This pilot project is primarily focused on the interwar period of the First Czechoslovak Republic, which was – in both Czech and Slovak history – an important period of great legal activity and preparations for major reforms.*

Keywords: *Legal history, Law, Database, Legal database, Digitisation, Legal literature, Czechoslovakia, Czech Republic, Slovakia, Interwar period*

1. INTRODUCTION

There is no need to persuade anybody that modern information technologies have brought about a revolution in working with legal documents and in the methods for their analysis. Before, such possibilities were primarily used in contemporary law and contemporary legal literature. Yet the new technological potential for improving access to historical legal literature remains significantly unexploited. The National Legal Heritage project at the Institute of State and Law of the Czech Academy of Sciences (AV ČR), supported by the Czech Republic – the Ministry of Culture², and focusing on making historical legal literature accessible, is an effort to effect gradual change in this field. The goal of the paper you are reading is to introduce also foreign readers to the specific issue of historical legal literature in the Czech Republic. It similarly asks more general questions about the functions and possibilities of such availability, and at the same time familiarises the foreign reader with answers to such questions, selected through the National Legal Heritage project, and therefore with the basic starting points and goals of this project.

* Mgr. Jan Kober, LL.M., Institute of State and Law of the Czech Academy of Sciences, Prague, Czech Republic

** JUDr. Ján Matejka, Ph.D., Institute of State and Law of the Czech Academy of Sciences, Prague, Czech Republic

*** Ing. Petr Aubrecht, Ph.D., Institute of State and Law of the Czech Academy of Sciences, Prague, Czech Republic

**** Bc. David Brůha, Institute of State and Law of the Czech Academy of Sciences, Prague, Czech Republic

¹ This paper has been prepared with support from the NAKI II project No. DG16P02H035 entitled Národní právní dědictví (National Legal Heritage), the provider of which is the Czech Republic – Ministry of Culture. The principal investigator of the research project was JUDr. Ján Matejka, Ph.D., Institute of State and Law of the Czech Academy of Sciences, Prague, Czech Republic.

² The NAKI II project No. DG16P02H035 entitled Národní právní dědictví (National Legal Heritage).

2. HISTORICAL LEGAL LITERATURE IN THE CZECH REPUBLIC

To be able to understand the starting points and context of enabling database access to historical legal literature in the Czech Republic, it is as a rule necessary to provide a concise introduction to the foreign reader, unfamiliar as they are with the situation. The state of Czech (and Slovak) specialised legal libraries is not satisfactory for a number of reasons, while the accessibility of old legal literature in particular is very bad. Explaining the causes of this situation will enable us to better see the requirements that digitisation can fulfil in its local context. As the Bohemian (Czech) state was relatively territorially centralised in the past – in comparison, for example, with the Italian or German lands – in this distant past only two universities with law faculties were established in the Lands of the Bohemian Crown, namely in Prague and in Olomouc. This contrasts with the German and Italian lands, where the establishment of small local universities became a matter of prestige of the local rulers, resulting in a denser network. This situation (only two universities with law faculties in the Bohemian lands) lasted until 1855, when the law faculty in Olomouc – and subsequently the whole university – was abolished.³ This left only a single law faculty, the one in Prague, which was later split (1882) – together with the university – into Czech and German parts which, however, shared the university library. This meant that the Czech development significantly differed from that in other lands, where the territorial particularism led to the establishment of a number of regional universities. Although there were efforts to establish a Czech university in Brno from the turn of the 19th century, this did not take place because of the lack of approval from the government in Vienna. It was only the establishment of the independent state in 1918 that enabled the formation of new universities and two new law faculties.⁴ The further development of the university networks so prevalent in other countries around the world in the second half of the 20th century occurred here with something of a delay, with the re-establishment of the law faculty in Brno (1969) and with the establishment of the law faculties in Olomouc (1991) and in Plzeň (1993). All the newly established faculties had to set up completely new libraries. The library of the Institute of Law, or more precisely the Institute of State and Law of the Czech Academy of Sciences, was formed from other libraries in 1956, while the libraries of the Supreme Court (established after the founding of Czechoslovakia) and the Constitutional Court (established in 1993) are also of note. By contrast, parts of the historical funds of some other libraries have suffered damage, particularly the libraries of the Prague faculty and the Olomouc faculty, both suffering significant flood damage.

³ Cancelled for political reasons as government intervention against democratically-minded students and teachers active in 1848; of the original Francis University of Olomouc, only the Faculty of Theology and the former university library were left as an independent institutions.

⁴ The law faculties of Masaryk University in Brno (1919) and Comenius University in Bratislava (1921). These faculties, however, began in modest conditions and their seminar libraries were only created with enormous difficulty and with great efforts by the teachers at the time. In addition, the Brno faculty was abolished for political reasons between 1950 and 1969.

3. THE SOLUTION METHOD AND ITS IMPORTANCE

The situation as described can of course be addressed in various ways. One traditional and proven solution during the establishment of new specialised legal libraries is the planned construction of a book and journal fund through the systematic acquisition of historical and contemporary legal literature, managed by prominent experts with targeted financing. One example of such an approach was the excellent work by the Czech lawyer Jan Štěpán (1914–2002) during the establishment of specialised foreign legal libraries in the 1970s and 1980s. This is a very good solution, yet expensive and slow. Another possible solution to this long-term problem, and one that has only become fully technically feasible in recent decades, is the systematic digitisation of this old literature. This approach also offers many advantages: the potential ubiquity of a digital work means it does not serve merely a single institution or restricted circle (institutionally or in terms of space and time) of users, but – with the application of the right concept – basically anybody anywhere in the world, and theoretically also a large number of concurrent users. This approach, however, differs in many ways from the regular forms of digitisation performed primarily by national state libraries, as the primary focus of such digitisation is not preservation, as is most common in library practice. Libraries primarily perform digitisation to preserve printed matter at risk of destruction or damage (e.g. old documents on poor-quality newsprint or other fragile or low-quality paper), and potentially to protect rare documents (old prints) or frequently used manuscripts (old manuscripts; in the archives also registers of births and deaths, books of burgher rights, etc.). In our case, however, these are not books either in danger damage or especially rare. Hence, we speak not of preservation digitisation but of supplementary digitisation – an effort to supplement the book and journal fund with additional accessible titles. We can also speak of opening digitisation, meaning enabling new possibilities for working with texts and information. Here the text is not available only through the traditional use and search methods such as linear reading, searching in metadata in the form of a traditional material or personal register or in a list of the content of the book or journal volume, in collections of legal decisions, etc. Users now have much wider options. Digitisation, corrected conversion using OCR, supplemented with metadata and the integration of texts into a database enables incomparably wider options for searching, as well as text and information analysis.

4. DATABASE CONTENT

When preparing the concept of the National Legal Heritage project, we decided not to limit it for example to only a single specific type of legal literature, but to attempt to provide a wide cross section of the literary output of the time. The database thus provides all types of historical legal literature, from monographs, through journals and collections of legal decisions, to *libri amicorum* and other collections. From the temporal perspective, we focused primarily on the interwar period in Czechoslovakia and within this period primarily on legal literature in the Czech language. This is a period of extraordinary importance from the perspective of the development of Czech (and Czechoslovak) law, a period incomparable to any other. The renewed independence produced unusually high levels of legal activity connected with the establishment of the new life of the state, with the



Image 1. Leopold Heyrovský (1852–1924), Professor of Roman Law at the Faculty of Law of Charles University. Photograph from the newly accessible Liber amicorum for Heyrovský (1924).

preparation of a series of reforms as well as unification works. From this period, still resting on Cisleithanian and Transleithanian legal bases, we returned, to a lesser extent, also to the Czech literary production of the immediately preceding period (some journals, *Všeobecný slovník právní* [General Legal Encyclopaedia], etc.). Of the journals, it included on the one hand important scientific journals (*Sborník věd právních a státních* [Collection of Legal and State Sciences]), and on the other hand key journals focused on legal practice (*Právník* [Lawyer], *České právo* [Czech Law], *Všehrd* [Všehrd Journal], *Právní praxe* [Legal Practice], *Zprávy Právnícké jednoty moravské* [Reports of the Moravian Lawyers' Association] and others). Considering the large number of titles, the digitisation process is still far from exhausting the period output. Offprints and overprints (often separately paginated) make up a special group on the border between journals and independent works. Collections of legal decisions are another interesting area of digitisation. The database provides interwar volumes of three key collections: the collection of the Supreme Administrative Court, the administrative order (Bohuslav administrative decisions), the collection of the Supreme Court, in its civil law edition and criminal law edition (*Vážný* decisions). Judicial decisions can also be found in a number of journals oriented on practice, in particular the *Právník* journal. The included monographs are primarily specialised legal works, but only a selection as the total output was very extensive. Attention might for example be drawn to selected works by Jan Krčmář, Rudolf Rauscher, Miloslav Stieber and others. *Libri amicorum* and other collections also make up a significant part of the content. It would be remiss not to mention also dictionaries, encyclopaedias, and similar

handbooks. The database provides the first extensive encyclopaedic legal handbook, the *Všeobecný slovník právní* [General Legal Encyclopaedia] dating from the end of the 19th century, the value of which is, however, more documentary than scientific. On the other hand, the later *Slovník veřejného práva československého* [Dictionary of Czechoslovak Public Law] demonstrated a high expert standard in many of its entries. At that time, these were seen as a prestigious publication platform and often the study of them was therefore both interesting and valuable. The digitisation project included a selection of important *libri amicorum* of the time, for example the *Libri amicorum* for Antonín Randa or Leopold Heyrovský, two world-famous professors at the university in Prague, or the *Liber amicorum* for August Miříčka. Of the other collections, we might for example mention the representative jubilee collection *Naše právo a stát* [Our Law and State] dating from 1928, or the jubilee collection of the Všehrd association.

5. DATABASE CONCEPTS OF ACCESS AND METHODS OF USE

The content is made available in the database in full-text html form through the conversion of scanned originals using OCR tools, which subsequently – in view of the sometimes significantly differing quality of the period documents – undergo manual and partly also automatic correction. The texts are however also concurrently made available everywhere as a facsimile, meaning the scanned historic originals. This allows the user to fully profit from the advantages of both access methods depending on their needs and preferences. The database enables users to search both in an advanced mode for more complex queries, and in a simple mode for simpler ones. Various query types can be intuitively formulated and combined, searching is possible in the full-text and in the metadata attached to the texts in the database, the selection can be narrowed for example by time period, etc. As with all quality databases, the National Legal Heritage enables the user to browse the document structure. This method of use is particularly advantageous for the user who does not have a specific goal, who is interested in a specific document, type of document, or a historic period as a whole. A wide range of export options for the subsequent further use of the information by the user is an essential element of every information system.

Image 2. Search mask, pilot form.

We can divide this into the export of complete units for subsequent use offline, and the export of copied text directly from a browser. It is also possible to download e.g. articles, judicial decisions, and other texts in pdf format.

6. WIDE PUBLIC ACCESSIBILITY AS A PRINCIPLE

The database was conceived to be widely accessible to the public – and thus its content was limited to free copyrighted works where the proprietary rights have already expired (as a rule 70 years after the author's death⁵). This concept enables the database to be accessible by the public under open-access principles, including the non-commercial free dissemination of results (open source platform – see below). The main result of the project as well as all the related results were created by the research team as 'employee work' pursuant to Section 58 of the Copyright Act, meaning as part of the performance of rights and obligations arising from their employment. The legal nature of this selected mode offers, in terms of the scope of employer rights, a unique position for the subsequent exercise of proprietary rights to these works, including extensive possibilities for modification and other unrestricted (non-commercial) dissemination.⁶ From the perspective of the project's actual purpose (the publication of historical legal texts), we should also mention the special legal mode of free access to information, in particular Section 4b of Act No. 106/1999 Coll., pursuant to which the provider shall provide all information through publication in all formats and languages in which it was created, while at least one of these formats must be open and, if possible, machine-readable (if possible and appropriate, the obligated entity shall also publish related metadata together with the information. Both the format and the metadata should, as far as possible, comply with open formal standards). This statutory requirement is a result of the response to changes in European legislation,⁷ which in terms of form and publication does not restrict any user as regards the method of use, and entitles all users to disseminate it further, as long as the author of the data is indicated during such use and dissemination, and other users will have the same rights as regards the disposal of the disseminated data.

7. BASIC PRINCIPLES OF THE TECHNOLOGICAL SOLUTION

The National Legal Heritage database was designed as a web application. The objective was to create an open, stable, and reliable solution. Another goal was for it to be expandable in the future and to guarantee that the application would remain easy to maintain and develop over the years to come. In relation to the requirement for openness, the solution emphasises the use of libraries and other open software (open source, free software, see below). The indicated libraries and software allow the project to be developed without

⁵ Pursuant to Section 27 of Act No. 121/2000 Coll., the Copyright Act.

⁶ MYŠKA, M. Selected Legal Aspects of Open Access to Scientific Information. *Právní rozhledy* [Legal Perspectives], 2014, No. 18, pp. 611–619.

⁷ The adoption of Directive No. 37/2013, which was updated through Directive No. 2013/98 to ensure public access to 'open data' as specific content.

the restrictions that usually accompany closed software and its rules of use. The components used in the solution do not restrict either the user or future development. Further improvement will thus be possible, for example also with regard to the need for greater performance or new or modified functionality. Some technologies are reworked over the years (e.g. JavaScript libraries, e.g. Angular), which entail repeated modifications to the projects that depend on them. This is why we chose technologies that will not change significantly over time and that are already well established. Hence there is a high probability that there will be enough developers capable of working with them even many years into the future. We applied the same principle to the software. This will enable us to build on an already debugged solution and reduce the need to spend considerable time debugging the used libraries. The question of security should also not be neglected. This was taken into account in the technology solution (technologies that are seldom attacked were chosen) and in the system concept (weaker elements such as the application server and database are protected and are not directly externally available).

8. BASIC PRINCIPLES OF SOFTWARE SOLUTION

The basic decision was to carefully choose suitable open software solutions (typically with open source licences). A decision of this type has positive consequences in terms of cost and long-term sustainability. An appropriate solution permits the use, modification, or other application of source code or other parts from third parties, while these parts can also be examined and modified for other use, as a rule free of charge, meaning without any other licencing or other fees, if certain minimal obligations are complied with. We have used MIT and BSD licences that allow users to dispose of the project practically as they see fit. Only the distribution of the derived code under the same name is forbidden. We have also used the GPL licence: use is permitted, but derivative works must be disseminated under the GPL licence – once free, the software must remain free. The greatest requirement placed on such a work is the obligation to provide the user of the work with the source code upon request. The whole licence is conceived as a competitive advantage for open source, while GPL works may use GPL libraries but not closed libraries, as the manufacturer does not want to provide its source code. We do not use GPL licence software in the application to avoid impacting the licence. One of the relatively standard requirements for the use of such advantages, or more precisely related user rights, is the publication of the source code (some licences however do not include arrangements regarding making the source code accessible). The use of a GPL software for monitoring, for example, is not the creation of a derived work, and thus does not impact the project licence. We also used an LGPL licence, a milder form of the GPL used for libraries, where the author wants to enable the library to remain free even in a closed work. A key requirement is to retain the possibility for the user of the target work to replace this library, e.g. with a modified or new version, and of course to provide the source code to this library upon request. We also used a CDDL licence, a licence that is similar to the GPL but supplemented with protection from software patents. Finally, we used the Apache Licence (ASL), a very loose licence that allows any type of use. Only the use of its own trademarks identifying products or other services is forbidden.

9. MORE DETAILS CONCERNING INDIVIDUAL SOFTWARE SOLUTIONS

PostgreSQL 43 was chosen for the database solution primarily because of its stability, speed, reliability, expandability, and the PostgreSQL licence, similar to the MIT and BSD.⁸ On the server side, we used a solution typical for enterprise systems that emphasises reliability – the Java Enterprise Edition 7 platform. Its wide use ensures it will be possible to continue to develop the system for a long time to come, as there will undoubtedly still be supported servers as well as programmers who have experience with this technology. Of the available servers, we used Glassfish, a reference implementation from Oracle.⁹ For the future, we are also considering transferring to a Payara server that offers somewhat better community support. Payara was created as a community fork of Glassfish with faster development, faster addressing of security issues, and additional extensions, while the licences are CDDL and GPL. On the application server, we deployed two types of monitoring – at both system (availability and notifications of issues by mail) and Java levels. This second application state monitoring is provided by the Java-Melody library.¹⁰ The use of JavaMelody enables us to monitor application state, memory use and performance requirements, while the licence is Apache License 2.0. The application server provides the required services yet unfortunately lags behind in terms of security. As the range of services does not permit perfect security, communication with the client is performed via a web server, in our case nginx, while the licence is a 2-clause BSD license.¹¹ This is a typical ‘reverse proxy’ setting. In addition to closing security holes, the web server also mediates HTTPS termination. Data are stored on a physical disk array made up of 4 physical disks in a RAID 10 array. The virtualisation system itself uses other disks. This guarantees that only this project gets all the performance of the disks. The project uses virtualisation possibilities, in this specific case through the KVM project. Thanks to this, it has been possible to deploy four virtual machines for the National Legal Heritage project on a single physical server (IBM Lenovo). Virtualisation is used here primarily because the configuration of the physical server provides significant excess performance for the operation of a single production server. The performance is thus split among multiple virtual servers, which definitely do not suffer from a lack of resources; in this perspective, we can consider virtualisation (similarly to cloud solutions) to be a current global trend. Although these are virtual servers, they are all of course monitored using special software. In our case, we have selected the Zabbix monitoring system, which is open source yet also an enterprise-class solution. It allows us not only to monitor the current state of the servers, but also to intelligently predict how individual trends will develop in the future, and it thus also allows us to regulate the resources needed or e.g. additionally purchase needed infrastructure before overloading occurs.

⁸ PostgreSQL [online]. [2020-01-02]. Available at: <<https://www.postgresql.org/>>.

⁹ Glassfish [online]. [2020-01-02]. Available at: <<https://javaee.github.io/glassfish/>>.

¹⁰ Javamelody [online]. [2020-01-02]. Available at: <<https://github.com/javamelody/>>.

¹¹ Nginx [online]. [2020-01-02]. Available at: <<http://nginx.org/>>.

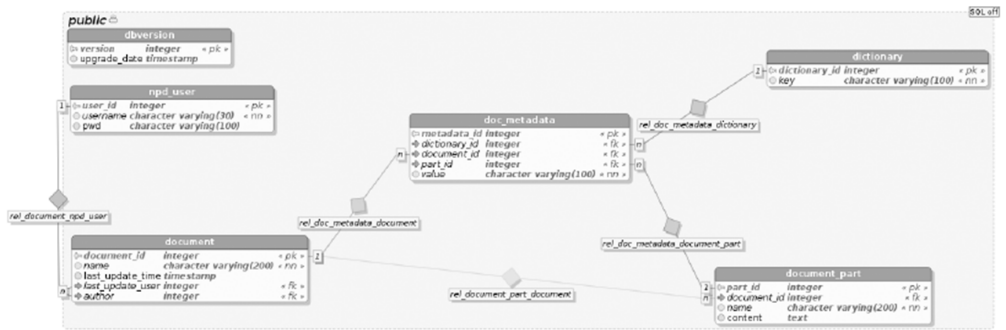


Image 3. Schematic sketch of the database

Backing up is always one of the most important elements of a system. Files are backed up daily using a deduplication system, ensuring that each file is physically backed up only once, while in all other backups there is only a link to that file. This dramatically reduces the space needed for the backups. The database (both production and for Git tasks) is also backed up to a geographically separated data centre near Montreal in Canada (chosen to provide backups in another continent), which speeds up recovery options.

10. CONCLUSION

The evaluation of the initial situation, namely difficult access to older Czech legal literature, indicated that the systematic digitisation of historical specialised legal literature was a possible solution. The National Legal Heritage project is an attempt to address this. It represents a new approach in terms of type, which does not want to be only digitisation, but rather aims at the creation of a quality legal information system that will represent a marked improvement in the accessibility of historical Czech legal information.¹² The basic benefits will be the possibility to perform advanced searches in the database and to perform analytical work with digital text. This will open up new possibilities for explorative work and research. There will of course be other benefits, typical for all digitisation projects – ubiquity, remote access, the removal of the need to protect valuable originals when they are being used by readers, etc. In Czech conditions, which are in many respects atypical compared to other states, these benefits will also represent a path towards the future fundamental and final healing both of historical deficiencies, and also the losses that libraries have suffered (e.g. during floods). One benefit – that is actually a consequence of the advantages mentioned above – is a fundamental social benefit: it relates to the inventories of Czech and Slovak legal faculties and universities in general. It will provide equality of access for members of academic communities of individual faculties and universities to historical legal texts. So far there have always been significant differences resulting from lack of inventory at faculty libraries – at least as regards older literature. More generally, it

¹² Searches in the National Legal Heritage database can be performed at: <http://www.npd.cz>.

will also mean the establishment of access equality to these texts between ‘centres’ (large towns with important historical and science libraries) and the periphery. Making historical legal information widely available for the needs of courts, state and local authorities, and the legal profession in general, will also be useful. Enabling access to historical legal texts for additional specialised non-legal groups (researchers in history, linguistics, political science, economics) may also help improve the interdisciplinary consideration of historical law and historical legal science by other disciplines.