



Scientific and Technical Information

Hungarian journal of library and information science

Vol. 60 • 2013/07

Summaries

BOROS, I.: Digitisation – experience from the Digitised Legislation Knowledge Store (DTT) project 283

The Digitised Legislation Knowledge Store (DTT) project of the Library of the Hungarian Parliament was implemented in the period January 4, 2010 to November 30, 2012, as a priority project, in the framework of the Electronic Administration Operational Programme (EKOP), supported by the European Union and co-financed by the European Regional Development Fund. This article describes the workflow of digitisation. After creating the IT background the general technical parameters were defined, and the physical and logical presentation of digital documents started. This was followed by the workflow for mass uploading: METS/XML was prepared based on templates for various groups of materials, and plans for quality control were made using mathematical-statistical methods. Principles for the selection of works for digitisation were defined, with categories according to fields of science, document types, language, time, etc. Hungarian-language materials were selected from the main collecting areas of the Library (law, history and political science). The size, structure and condition of the volumes selected were registered on status sheets. Books were described individually, while journals, official gazettes and decisions were described in groups in a conservation status database. In order to select suitable copies it was necessary to identify the bibliographic, copy and publishing data in catalogues, and to check, improve and prepare for conversion the records of books and journals concerned, to enter their copyright status and collection organisation codes into the database. Metadata were prepared. The digitisation company was chosen within a public procurement process. Because of the value of works selected for digitisation, their uniqueness and in many cases irreplaceable nature, as well as preservation considerations, digitisation took place on site, mostly using a Kirtas KABIS III. robotic scanner. Large volumes and fold-out attachments were digitised by a flatbed scanner. The processing of the images, their cutting and correction was carried out by the Book Scan Editor (BSE) software. Metadata were assigned to each page of each volume by the GLOBE-Index software. This was followed by optical character recognition (OCR): after image processing the data were transferred into the OCR database, and the OCR Engine automatically created the final two-layer PDF file format. The pages prepared were then input into the DigiTool software in this format. The first step of checking was the automatic control of TIFF images. During manual inspection the general quality control of scanning results took place. Within the project two million pages have been digitised, the total number of volumes was 5272. The documents can be accessed in accordance with the copyright legislation in force. A major part of works (40%) is under copyright protection; consequently they can be displayed on the Library's computers only for scientific research or private study. The works that are not subject to copyright protection are available without any limitation to the public on the Internet. The DTT portal is barrier-free; those visually impaired can use it properly as well.

TÓTH, É.: Copyright aspects of the Digitised Legislation Knowledge Store 291

Libraries planning digitisation activities must consider three main issues in terms of copyright: May they digitise their holdings at all? What rights arise for the library after digitisation? What can the library do with the material within legal limits? Before answering these questions a number of tasks has to be solved, in order to ensure that the works are utilised in

accordance with copyright regulations: first, it has to be established whether the work in question is protected under copyright or related rights; second, if so, whether the period of protection has already ended; third, whether the planned action is qualified as authorized or free use; and, after this three-step test, whether its free use (digitisation and provision) is possible. The DTT portal includes two million pages of historical parliamentary documents, sources of law, literature on law, history and political science. About 60% of these materials does not either belong under the Copyright Law of 1999 (no. LXXVI.), or the specified protection time has already elapsed, consequently, they can be made accessible to all. These materials are provided by the Library of the Hungarian Parliament organised into a database. A smaller part of documents, about 40%, are still under copyright protection. These digitised pages are provided solely on a closed network, and one cannot replicate (copy, print or save) them. As libraries and the book publishing industry have got different views on some issues, it is important to continue their dialogue begun last year, as it is their common interest to design a modern service system.

KARDOS, H.: Content provision on the DDT portal297

One of the main goals of the Digitised Legislation Knowledge Store project was to digitise outstanding works on law, political science and the history of the Hungarian Parliament, for the purposes of legislation, education and interested individuals. It was found most justified to digitise legal literature between the two world wars as it is the literature, arguments and term used in the period in question that best reflect the continuity of language and thinking within the literature on law and political science. The parliamentary literature collection includes books on parliamentarism, on the history of the Parliament, its organisation and operation. The works selected from the Library's collection for digitisation were mostly published between 1850 and 1945, in Hungarian. In the Knowledge Store compiled legal sources represent a significant part: the official gazettes of seven ministries, and collections of decisions with the theoretical foundations by courts. During the project a total of 73 renowned journals in the field of law and political science were digitised, together with their articles (overall, 20,000 issues containing approximately 90,000 articles) which can be searched by author, title and subject. The documents are arranged in subject groups (called collections) and are managed by the DigiTool system. Information is provided by the integrated library system database, the connection between the two systems – through the same record identifiers. 40% of materials are protected by copyright, therefore, they are accessible only on site. In case of journals the protection period was set at 110 years from publication date. This means in practice that from January 1, 2013 the journal issues published in 1902 and earlier are accessible. The DTT portal constitutes a linked content management and collection management system which contains 1) digitised content with metadata organised into subject collections on the one hand, 2) related reviews about digitised content on the other, 3) plus earlier catalogued parliamentary materials in databases with a link to the new portal.

Z. KARVALICS, L.: About data hugging and beyond300

In the world of data earthquake-like changes are taking place, accompanied by reflections from extensive special literature. This brief review does not, however, review recent developments; it strives to introduce as many new, frequently closely intertwined concepts making up conceptual word clusters from data culture, as possible.