

The Czech Digital Library - Fedora Commons based solution for aggregation, reuse, dissemination and archiving of digital documents

Tomas Foltyn

Collection's Management Division Director
National Library of Czech Republic
Klementinum 190, Prague 1, 11000, Czech Republic
+420221663526
tomas.foltyn@nkp.cz

Martin Lhotak

Institute of Information Studies and Librarianship
Faculty of Arts of the Charles University in Prague
Jan Palach Square 2, 116 38 Prague 1
+420221403260
lhotak@knav.cz

ABSTRACT

How to effectively ensure complex digitization processes including post-processing, workflow monitoring, archiving and making the content available is still a current issue from both national and international perspectives. To support the culture heritage institutions with open-source solutions, make their activities more effective and become the central access point to digitized documents stored in the libraries is the main aim of the project "Czech Digital Library", one of the ground pillars needed to provide centralized digital services in the Czech Republic defined by the "Library Development Strategy of the Czech Republic for 2011 to 2015," approved by the Czech Ministry of Culture. The project will provide tools for the whole process from digitization to data processing to dissemination and archiving.

Czech Digital Library Project Description

The goal of the project is to create the Czech Digital Library ("CDL") which will aggregate content of digital libraries in the Czech Republic. It will serve both as a uniform interface for end-users and as a primary data provider for international projects, especially for Europeana and TEL – The European Library.

The open source Kramerius system is the initial software solution for the Czech Digital Library. Kramerius is based on the Fedora Commons repository system and is widely used as a digital library system in the Czech Republic. It was developed jointly through the cooperation of the Library of the Academy of Sciences and the National Library of the Czech Republic. Besides data harvesting from different instances of the Kramerius system, it is also necessary to arrange a connection with other systems used in Czech libraries (e.g., Dspace, Eprints). The Czech Digital Library will serve as an OAI-PMH provider compatible with Europeana. Some other OAI-PMH profiles might also be implemented to facilitate cooperation with other projects.

The Digitization Registry, which was built formerly as a project of the Library of the Academy of Sciences and the National Library of the Czech Republic, is used as an interconnecting system and relevant source of information. The Registry is publicly accessible online at <http://www.registrdigitalizace.cz/>. It holds a large amount of information about digitized documents in the Czech Republic.

Included is the identification of original printed documents, owner and location of the digital library where the digital document is available, persistent identification and other relevant entries. The main aim of the national registry of the digitized documents is to avoid unwanted duplication and to enable the sharing of digitization results throughout the Czech Republic. The Digitization Registry could also provide tools for digitization workflow management to simplify the process of monitoring the digitization. This solution could serve to end-users as the central access point to digitized documents. Very important is the fact that it cooperates with library catalogue systems as well as with digital document repositories. In the context of interoperability and cooperation with library information systems, the Registry is designed to communicate and cooperate automatically with other library information systems as much as possible. It uploads bibliographic records of items chosen for digitization in batches exported from the Aleph catalogue in MARCXML. The Registry is able to harvest data from digital libraries via OAI-PMH to import data describing digitized documents. Finally, it provides information about completed digitization to library OPACs together with a link to digital documents. Information is subsequently sent from library OPACs to the Union catalogue of the Czech Republic.

Other open source tools to support complex digitization processes are developed in the frame of the CDL project. Included is the digital documents processing and archiving solution ProArc based on the Fedora Commons repository and cooperating with Archivematica. The goal is to use these tools to increase the number of materials available in the Czech Digital Library. Using the same production and archiving tools will enhance interoperability and data sharing between individual digitization projects. The rapid semi-automatic creation of the standard metadata is enabled by the production system. It involves structural, descriptive and archival metadata, OCR and conversion to specific graphical formats. With regard to the archival part of the solution, standards for long term archiving, such as the OAIS model, are implemented.

Mutual interoperability between all developed systems and tools is accented in the frame of the project as well as interoperability with solutions already existing on the market. The aim is to share, use, reuse and archive digital content as easily and effectively as possible with open source solutions.