# Development of open source systems for digital libraries – ProArc the system for digital document production and archiving

## Martin Lhoták

Library of the Academy of Sciences of the Czech Republic / Institute of Library and Information Science, Faculty of Arts, Charles University in Prague

**Abstract:** Open source solutions are widely used to run a digital library (for example DSpace, EPrints and Fedora Commons repository). Nevertheless there is not many open source tools available for a digital document production and there is not many tools for long term archiving of digital documents. In the frame of the Czech Digital Library project the open source ProArc system is developed which enables to create standardized metadata, export digital documents to a digital library and provide data for long term archiving. The solution follows international standards and it can be freely used by all interested institutions which want to run digitization projects.

**Keywords:** digital library, open source, metadata, long term archiving,

## 1. Introduction

There is a significant activity in open source system development around the world for several years already. Systems like DSpace, EPrints and Fedora Commons repository are widely used in many libraries as a digital library or repository. On the other hand, there are not many open source tolls available to create digital documents in a standard way, with standardized metadata and with possibility to create Submission or Archival Information Package to archive the data or Dissemination Information Package to provide the data to end users. Such solution was developed in the Czech Republic in The Library of the Academy of Sciences in cooperation with IT company INCAD. The source code is in Java and Fedora Commons is used as a repository.

## 2. Production and archiving system ProArc description

Production and archiving system ProArc is freely available open source solution that enables to create and edit descriptive, administrative and technical metadata for both – digitized and digital born – documents. System exports standardized data for dissemination in a digital library as well as Submission Information
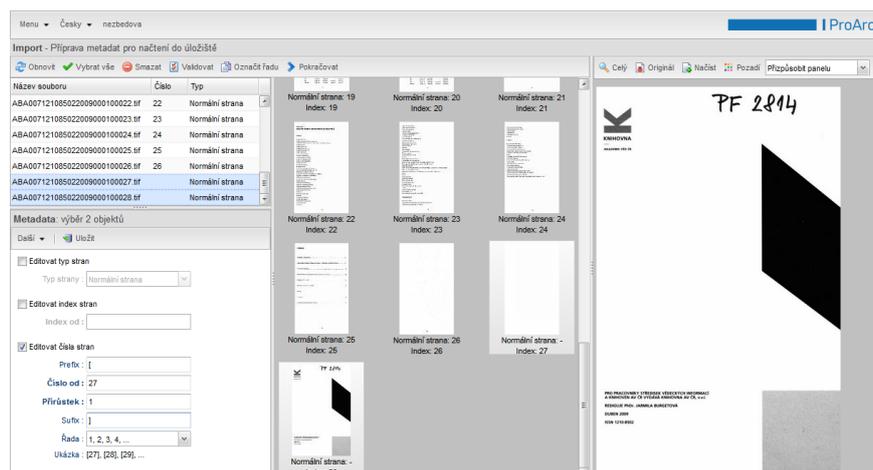
Package which might be used as import data for archiving system in compliance with OAIS reference model. ProArc follows standards of the Czech National Library which are based on international standards such as MODS, METS, PREMIS, MIX and Dublin Core.

ProArc is interoperable system (Athanasopoulos, 2010) that is able to reuse data from other library information system, MARC bibliographic data are used to create descriptive metadata in MODS. ProArc also share data about digitized documents in digitization registry RegistrDigitalizace.cz described by T. Foltyn (2013). ProArc is also able to create links from documents published a digital library to the corresponding records in library information systems.

RDflow is a part of the ProArc which serves to monitor digitization workflow. Such tool is necessary for efficient work of the digitization centers and workplaces.

ProArc is in full production in the Digitization center of the Library of the Academy of Sciences (Lhotak, 2007) since end of 2013. More than 400 thousand pages were processed within one year. Other workplaces are currently starting to use the system, which is permanently improved and enhanced (e.g. new types of documents are being added.)

*Image 1 – Example of metadata creation in ProArc*



### 3. Archiving part of the system

With regard to the archival part of the solution, standards for long term archiving, such as the OAIS model are implemented.

Functionality which will support long term preservation and smart cooperation with Archivematica LTP system will be finished in 2015.

ProArc will provide standard SIP – submission information package which will be transferred to Archivematica and converted to AIP – archival information package and stored for long term archiving.

## 4.  Conclusions

ProArc system is one of the main outputs of the Czech Digital Library (CDL) project funded by Ministry of Culture program NAKI. Besides ProArc, two another solutions – Kramerius digital library system and RDflow digitization workflow system – are developed in the frame of CDL project.

ProArc enables to create standardized metadata and prepare complex digital document, which could be used as an import package for long term archiving solution. It also enables to create DIP – dissemination information package which serves as an input data to public accessible interface in a digital library.

More information:
https://code.google.com/p/archivacni-system/
http://www.czechdigitallibrary.cz/en/

**References**

Foltyn, T. (2013) Has It Been Already Digitized? How to Find Information about Digitized Documents, *Review of the National Center for Digitization*, Faculty of Mathematics, Belgrade, 2013, Issue: 22, ISSN: 1820-0109.

Athanasopoulos, G. ; Fox, E. ; Ioannidis, Y. ; Kakaletris, G. ; Manola, N. ; Meghini, C. ; Rauber, A. ; Soergel, D. (2010) A Functionality Perspective on Digital Library Interoperability. In Research and Advanced Technology for Digital Libraries. Heidelberg: Springer Berlin. P. 405-408. ISBN 978-3-642-15463-8. Available on the World Wide Web: http://dx.doi.org/10.1007/978-3-642-15464-5_43

Lhotak, M. (2007) Das Digitalisierungszentrum und die Digitale Bibliothek an der Akademie der Wissenschaften, Prag. In Digitalisieren - Internationale Projekte in Bibliotheken und Archiven. Berlin : BibSpider Info-Networking für Bibliotheken,. P. 92-101. ISBN 978-3-936960-17-4.