# CD-ARK

## *A Tool for Cooperative Processing of Optical Discs*

**Zdeněk Hruška**

*The Moravian Library*
*Czech Republic*
*zdenek.hruska@mzk.cz*
*0000-0001-7606-5693*

**Abstract – Optical discs are a valuable and unique part of library collections, but they are vulnerable to a variety of risks. CD-Ark is a collaborative software tool for processing and storing data from optical data discs in Czech libraries. It creates data packages that include ISO disc image, technical and bibliographic metadata, and checksums. This helps protect data stored on optical discs.**

**Keywords – Optical discs, Cooperation, Iso image, Bit-level preservation, Libraries.**

**Conference Topics – Collaboration; Building Capacity, Capability and Community**

### I. INTRODUCTION

Library collections in the Czech Republic contain tens of thousands of optical discs (such as CD-ROMs, DVD-ROMs, Audio CD, DVD-Video and others). These are at risk of physical damage, loss, theft, or poor storage conditions. Little attention has been paid to the issue of long-term preservation of optical discs in Czech libraries. Although the carrier is also a cultural artifact and must be protected, it is primarily the data stored on it that must be preserved for a long time. Therefore, the Moravian Library has decided to develop tools that will enable the processing and long-term protection of optical disc contents.

### II. CD-ARK

The CD-Arch system enables cooperative processing of optical data discs (mainly CD-ROMs and DVD-ROMs) and storage of bit copy on the central server. Tools are designed for libraries that can collectively process all the optical data disc collections. To avoid duplication in the carrier processing, bibliographic metadata and the disc image checksum check is set. This should prevent the same disc from being processed in different libraries.

For the sake of simplicity and to maximize and cost savings, the existing application has been modified (from "Book cover" project). Instead of scanning book covers, it now processes disc images.

### III. WORKFLOW

Carrier processing begins by downloading the metadata record from the electronic catalog, followed by CD-Ark-client application creating a data image from the optical disc (the so-called ISO image based on ISO 9660). Then, the top of the disc is scanned, the cover and booklet (because the disc's visual appearance must also be kept) and the data, along with the checksums, are sent to the server for further processing.

CD-Ark-server continues with processing, adding the output from the DROID tool into the data package, creating OCR from the scans, and wrapping everything in BagIt package format. The package is stored on the server and, thanks to checksums, it is possible to perform regular automatic controls to check for data corruption.

Thus, at least bit-level data protection from optical data discs is solved. In the future, it will also be necessary tackle the problem of logical long-term preservation, but it is now important to get data from the discs. Many of them are approaching the end of their lifespan (e.g. burned discs) and others are at risk of physical damage, loss or other threats.

iPRES
2019

# REFERENCES

[1] "Obálky knih.cz", GitHub, c2019. [Online]. Available: https://github.com/cbvk/obalkyknih. [Accessed: 17-Mar.-2019].

[2] "CDArcha-client", GitHub, c2019. [Online]. Available: https://github.com/moravianlibrary/CDArcha-client. [Accessed: 17-Mar.-2019].

[3] "ISO 9660", The File Format Wiki, 2019. [Online]. Available: http://fileformats.archiveteam.org/wiki/ISO_9660. [Accessed: 17-Mar.-2019].

[4] "CDArcha-server", GitHub, c2019. [Online]. Available: https://github.com/moravianlibrary/CDArcha-server. [Accessed: 17-Mar.-2019].

[5] "DROID: (Digital Record and Object Identification)", GitHub, 2018. [Online]. Available: https://digital-preservation.github.io/droid/. [Accessed: 17-Mar.-2019].

[6] J. Kunze, J. Littman, L. Madden, E. Summers, A. Boyko, and B. Vargas, The BagIt File Packaging Format (V0.97). 2016. [Online]. Available: https://tools.ietf.org/id/draft-kunze-bagit-14.txt. [Accessed: 17-Mar.-2019].

iPRES 2019 - 16th International Conference on Digital Preservation
September 16- 20, 2019, Amsterdam, The Netherlands.

2