

Developing Music Digital Library based on Polish Traditional Music Archives and dLibra

Marcin Werla
mwerla@man.poznan.pl
Poznan Supercomputing and
Networking Center

Jacek Jackowski
jacek.jackowski@ispan.pl
Institute of Art
Polish Academy of Sciences

Magdalena Chudy
magdalena.chudy@ispan.pl
Institute of Art
Polish Academy of Sciences

Ewa Łukasik
ewa.lukasik@cs.put.poznan.pl
Institute of Computing Science
Poznan University of Technology

Ewa Kuśmierk
ewa.kusmierk@man.poznan.pl
Poznan Supercomputing and
Networking Center

Ewa Dahlig-Turek
ewa.dahlig-turek@ispan.pl
Institute of Art
Polish Academy of Sciences

ABSTRACT

This paper presents an approach to developing a virtual environment for researchers interested in various aspects of traditional music. This undertaking draws on the combined expertise of the Institute of Art of the Polish Academy of Sciences (IAPAS) in maintaining the largest archive of Polish traditional music and of Poznan Supercomputing and Networking Center (PSNC) in developing digital library software (dLibra) and research tools for digital humanities. The work described in this paper is a part of the ongoing activities of DARIAH-PL Music Information Retrieval Working Group.

KEYWORDS

music information retrieval, traditional music, research environment, digital libraries

1 INTRODUCTION

The Institute of Art of the Polish Academy of Sciences (IAPAS) holds the largest sound archives of Polish traditional music. Database systems currently used by the Institute do not support more advanced research tasks. Therefore, IAPAS has initiated the Music Information Retrieval Working Group (MIR WG) within DARIAH-PL [2, 3]. The group's aim is to develop a virtual research environment (VRE) within Music Digital Library to assist the research process, from selection of source materials to automated extraction of desired audio features. The infrastructure of the VRE is based on dLibra Digital Library Framework, part of the DInGO [4] package developed by Poznan Supercomputing and Networking Center (PSNC), whose functional scope covers many aspects required to collect, manage and publish (meta)data of digital objects.

2 POLISH TRADITIONAL MUSIC

Polish traditional music was partially preserved in music notation thanks to Oskar Kolberg (1814-1890). Since 2014, his collection of

ca. 20,000 folk tunes has been continuously transcribed by IAPAS into the Essener Assoziativ Code (EsAC) format [6].

Phonograph recordings [7, 8] were introduced in Poland at the beginning of the 20th century. Unfortunately, the earliest collections (nearly 25,000 recordings) were lost during the WW2. The national Folk Music Collecting Campaign (1950-1954) resulted in 46,000 recorded items. They have been stored in the IAPAS folk music archive currently containing ca. 150,000 recordings. This major repository as well as numerous smaller archival collections of traditional music have been systematically digitized and made available via the IAPAS web application since 2014 [1].

3 OUR APPROACH

The aim of the MIR WG is to develop an environment which supports researchers working with the contents of music archives. In particular, the system should provide an extended support in three major areas such as: i) data collection and management; ii) data processing, analysis and visualisation; and iii) result publishing.

It should also offer tailored sets of functionalities for target user groups, e.g. researchers and educators, media professionals and artists, citizen scientists and general public [13].

3.1 System Overview

Figure 1 illustrates the logical structure of the new research platform. It is divided into three areas: data sources (orange), the main research environment (green) and users (blue). Regarding data sources, end-users can utilize a dedicated data management application to upload data from their private collections or closed archives and automatically ingest data from open archives. The collected data is deposited in a dedicated MIR repository with a profiled search index and data enrichment module. This module allows various data processing tasks to be executed, so that the imported data is enriched with features and information extracted automatically from the source data. The additional information is stored back in the repository and indexed, thus improving the information retrieval and analysis possibilities. The data management application is also used to manage collected data and organize it as needed. The system provides a public data portal with authorized access via web interface and via API to support reuse of research datasets. It also ensures that each published dataset has a registered DOI and is unambiguously citable.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

DLfM'18, 28 September 2018, Paris, France

© 2018 Copyright held by the owner/author(s).

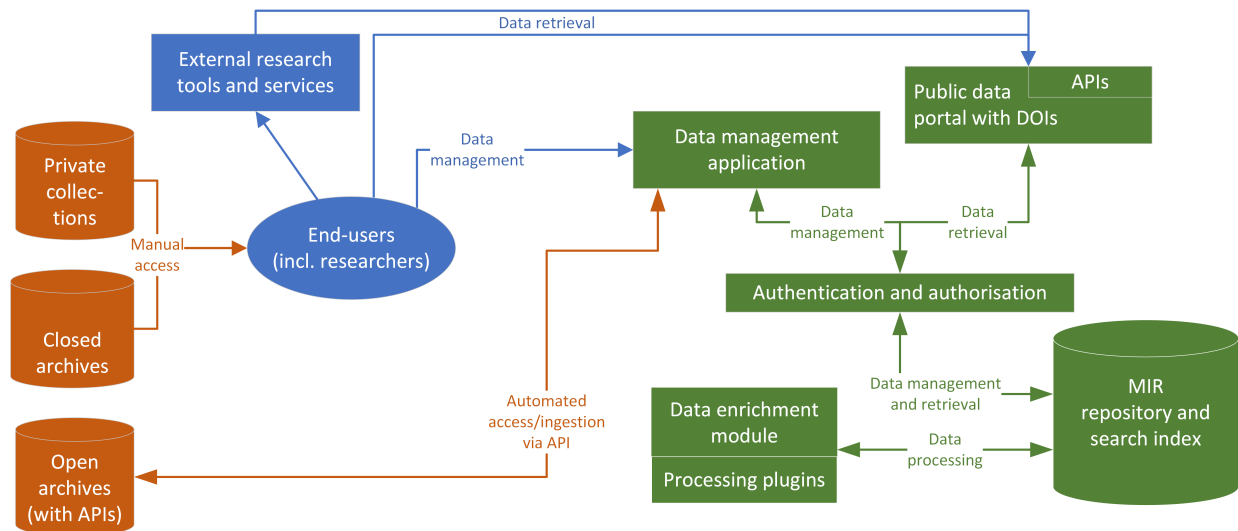


Figure 1: Overview of the main components of the MIR research environment

3.2 dLibra as Research Environment

The dLibra software is a distributed and portable digital library framework developed for building complex and highly interoperable digital libraries. It has a service-oriented modular architecture [12] and uses dLibra data model for organising digital objects [11]. The system has been chosen for implementing the main research environment as it offers configurable metadata schema, support for various types of content, collections and complex objects, basic and advanced search with facet-based filtering, content versioning and many other features. It also provides an ability to expose gathered data in several interoperable standards such as RDF, OpenSearch, OAI-PMH and OAIRO.

The implementation steps will focus on adapting and extending dLibra’s functionality for music content. In particular, on enabling various MIR tasks including music retrieval based on symbolic notation, audio content and metadata, symbolic notation conversion and use of open source music processing tools for automatic content description. Changes made to the dLibra data model will be reflected in the end-user interface and dedicated support for new types of content will be added.

An example of such extension is a specialized dLibra module which supports archiving and searching for traditional Polish music written in a symbolic text-based EsAC notation. Its prototype called WebEsAC [5] offers an online access to the database where each tune is saved as a separate record, with an easy-to-use interface. The implementation of EsAC-to-MIDI and MIDI-to-EsAC converter gives additional flexibility to the system and allows aural verification of the encoded melody and rhythm.

3.3 Content Resources

The MIR research environment has been primarily developed to assist ethnomusicological studies carried out in IAPAS and focused on its folk music collections. The sound archive is currently administered and operated using a dedicated system designed and built in IAPAS to address its specific functional requirements [9]. The

system includes a dedicated database with a structure optimized for developing ethno-phonographic collections and a specialized user application.

The database stores descriptive, administrative and technical metadata which describes each digitized medium, its creators and performers, the object’s structure, the date of object creation, technical data related to the object, and paradata regarding its digitization process. The metadata also includes information connecting related documents of various types such as digitized graphics associated with the recording (e.g. music or song transcriptions), medium description in the form of a track-list, protocol or label scan, photographs of an artist or instrument. The system supports document versioning for tracking history of changes.

As shown in Figure 1, the IAPAS archive represents a closed archive while Polish Digital Libraries Federation [10] exemplifies open archives. The latter is a service operated by PSNC, which aggregates metadata of about 5 mln heritage objects available online from digitized collections of hundreds of Polish cultural and scientific institutions. It will provide an extensive source of additional information contexts for music collections managed by IAPAS.

4 CONCLUSIONS AND FUTURE WORK

The paper presented an initiative undertaken by DARIAH-PL MIR Working Group to develop a research environment which facilitates access to Polish traditional music archives with extended support for advanced research tasks. The system is built based on dLibra, the existing multimedia library software. In the first step it will be integrated with dedicated modules designed and built in IAPAS for managing and sharing ethno-phonographic collections owned by the Institute as well as those obtained from other institutions. The system components are already being built, e.g. WebEsAC for encoding and archiving tunes in EsAC symbolic notation. In the future, it is planned to link the developed music digital library with traditional music archives of other regions of Central and Eastern Europe.

REFERENCES

- [1] CADIS – Centre for Archivisation and Digitisation of Image and Sound. <http://cadis.ispan.pl/en/>.
- [2] DARIAH-PL Consortium. <http://dariah.pl>.
- [3] DARIAH-PL Music Information Retrieval Working Group. <http://mir-wg.dariah.pl/en/main-page/>.
- [4] DInGO Software Package. <http://dingo.psnc.pl>.
- [5] Marcin Chmiel, Joanna Solomiewicz, Ewa Lukasik, and Ewa Dahlig-Turek. 2018. *WebEsAC Web-based system for archiving and retrieval of Polish traditional music in EsAC format (in Polish)*. Technical Report RB-10/18. Institute of Computing Science Poznan University of Technology.
- [6] Ewa Dahlig. 2000. *EsAC. An Integrated System of Encoding, Analysing and Processing of One-part Melodies*. Universitätsverlag Rasch, Osnabruck, 427–433.
- [7] Jacek Jackowski. 2006. The Oldest Sound Archive of Traditional Music in Poland in Relation to Current Challenges. *Tautosakos darbai / Folklore Studies* 31 (2006), 25–39.
- [8] Jacek Jackowski. 2014. Sound archives in Poland - resources and availability (in Polish). <http://www.muzykatradycyjna.pl/pl/leksykon/artykuly/articles/archiwa-dzwiekowe-w-polsce-zasob-i-dostepnosc-679>.
- [9] Jacek Jackowski (Ed.). 2017. *Polish traditional music – phonographic heritage. The current state – preserving and sharing (in Polish)*. Liber Pro Arte, Warsaw.
- [10] Cezary Mazurek, Marcin Mielnicki, Aleksandra Nowak, Krzysztof Sielski, Maciej Stroiński, Marcin Werla, and Jan Węglarz. 2014. CLEPSYDRA Data Aggregation and Enrichment Framework: Design, Implementation and Deployment in the PIONIER Network Digital Libraries Federation. In *Intelligent Tools for Building a Scientific Information Platform: From Research to Implementation*, Robert Bembenik, Łukasz Skonieczny, Henryk Rybiński, Marzena Kryszkiewicz, and Marek Niezgódka (Eds.). Springer International Publishing, Switzerland, 275–288.
- [11] Cezary Mazurek, Tomasz Parkola, and Marcin Werla. 2010. Storage, management and online presentation of multi-domain cultural heritage dLibra software case study. In *2nd International Conference on Information Technology (ICIT)*. 249–252.
- [12] Cezary Mazurek and Marcin Werla. 2005. Distributed Services Architecture in dLibra Digital Library Framework. In *8th International Workshop of the DELOS Network of Excellence on Digital Libraries on Future Digital Library Management Systems*, Vol. 29.
- [13] Josephine Simonnot and Stephanie Khoury. 2014. Applications and implications of digital audio databases for the field of ethnomusicology, a discussion on the CNRS -Musée de l’Homme sound archives. <https://halshs.archives-ouvertes.fr/halshs-01280831>. *First Monday* (Oct. 2014).