

Long time archive for audio works

Degree programme: Master of Science in Engineering | Vertiefung: Industrielle Technologien

Thesis advisor: Daniel Debrunner

Expert: Fabian Page, Bozzio AG

External project partner: Schweizerische Stiftung Public Domain, Küsnacht

The Swiss Foundation Public Domain is responsible for the long time data archive of the volunteer driven Public Domain Project. The volunteers are collecting, digitizing and capturing metadata of old audio records, mainly 78 rpms, that are out of copyright.

In this master thesis a data model was developed to represent the meta-data as Open Linked Data. Also a trustworthy archival storage according to OAIS was evaluated and first migration steps were undertaken.

Following the semantic web (Web 3.0) standards the Metadata (title, creator, publication date, images etc.) is modeled as triples (subject, predicate, object) using the ontologies Dublin Core, Schema.org, Music Ontology, Creative Commons and Logistics Core. The new data model is accessible via a web API that delivers RDF/XML or turtle. This fosters the reuse of this metadata on other websites and projects, which increases the overall value of the metadata and the work of the Public Domain Project itself.

This model is implemented as a set of new templates and forms using Semantic MediaWiki (SMW). SMW allows the value of a data field to be shown on other wiki pages with a semantic query. A data field may have data validation or can have only a limited set of values. These features simplify data entry and reduce errors significantly.

A trustworthy storage system for the digitized audio files must fulfill digital preservation requirements defined by the OAIS model. A new system structure was evaluated and a migration strategy was defined. As a first step the operating system of the file server was replaced by Gentoo GNU/Linux because it stores the source code of every installed software. The source code together with file format specifications etc. is called **representation information** and which needs to be preserved together with the audio files to guar-

antee the understandability of the bits on the storage media.

A document management system (DMS) for the internal document handling of the foundation was evaluated and the selected NextCloud was implemented on a new virtual machine (VM) secured with TLS and certificates from Let's encrypt.

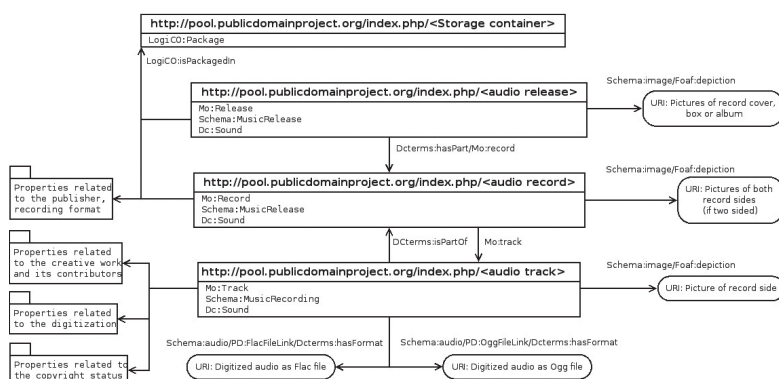
The full report can be downloaded from:
<https://publicdomainproject.org> (License: CC BY)



Christoph Zimmermann
+41 77 433 24 75



Example of a 78 rpm (shellac) record from 1938 that gets digitised by a volunteer with a laser turntable



The data model for audio works developed for the Public Domain Project