



Masterthesis Langzeitarchiv für Audiowerke

Long time archive for audio works

Masterthesis



CC BY 4.0

Studiengang:	Master of Science in Engineering (MSE)
Autor/in:	Christoph Zimmermann
Betreuer/in:	Daniel Debrunner
Experte/in:	Fabian Page (Bozzio AG)
Auftraggeber/in:	Schweizerische Stiftung Public Domain
Datum:	23. Januar 2017

Management Summary

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 (Shellac), that are out of copyright.

In the course of this master thesis a data model was developed to represent the metadata 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 thereby 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 guarantee 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.

Zusammenfassung

In dieser Arbeit, die zusammen mit der Schweizerischen Stiftung Public Domain durchgeführt wurde, geht es um Audiowerke wie Musik, Hörspiele, Reden etc. die digitalisiert wurden und in der digitalen Domäne langfristig erhalten werden sollen, so dass nachfolgende Generationen darauf Zugriff haben.

Dieser Masterarbeit ist eine Projektarbeit vorausgegangen, in der die Grundlagen zur digitalen Langzeiterhaltung und des OAIS Referenzmodells erarbeitet wurden. Mit einem Audit gemäss CCSDS 652.0-M-1 wurde beurteilt, wie die Erhaltungsziele des Audioarchivs, das von der Stiftung betreut wird, erfüllt werden. Abschliessend wurde eine neue Systemarchitektur für die digitale Langzeiterhaltung von Audiowerken präsentiert.

Darauf aufbauend beschäftigt sich diese Masterarbeit mit dem öffentlichen Verfügbarmachen semantisch aufbereiteter Metadaten als Open Data mit Hilfe von Semantic MediaWiki und der Bestandssicherung des digitalen Archivs des Public Domain Projekts.

Im entwickelten Datenmodell werden die Metadaten als tripple (Subjekt, Prädikat, Objekt) modelliert, so wie es von den Semantic Web (Web 3.0) Standards vorgegeben wird. Dazu werden die Ontologien Dublin Core, Schema.org, Music Ontology, Creative Commons und Logistics Core eingesetzt. Auf die Metadaten kann per Web-API zugegriffen werden, zurückgeliefert wird RDF/XML oder Turtle. Dies fördert die Weiternutzung dieser Metadaten auf anderen Webseiten und Projekten, was wiederum den Wert der Metadaten und der Arbeit des Public Domain Projekts erhöht.

Dieses Datenmodell wurde implementiert als Set von neuen Templates und Formularen auf Basis von Semantic MediaWiki, einer Erweiterung zur Webapplikation MediaWiki, die im Projekt für die Erfassung von Tonträgern eingesetzt wird. Semantic MediaWiki ermöglicht das Anzeigen von Datenwerten auf anderen Wikiseiten mittels semantischer Abfragen. Ein Datenfeld kann eine Datenprüfung implementieren oder nur einen begrenzten Satz von Werten zulassen. Diese Funktionen erlauben es, dass die neuen Formulare den Dateneingabeprozess wesentlich vereinfachen und Fehler signifikant reduzieren.

Eine vertrauenswürdige Speicherlösung für digitalisierte Audiodateien muss die Anforderungen für die digitale Erhaltung definiert vom OAIS Model erfüllen. Eine neue Systemstruktur wurde evaluiert und die Migrationsstrategie dahin definiert. Als erster Schritt wurde das Betriebssystem des Speicherservers durch Gentoo GNU/Linux ersetzt, weil es die Source Codes von jeder installierten Software speichert. Dieser Source Code zusammen mit Spezifikationen zu Dateiformaten etc. wird *Repräsentationsinformationen* genannt, welche genau so wie die Audiodateien erhalten werden müssen, um die Interpretierbarkeit der Daten auf dem Datenträger bis hin zum letzten Bit zu garantieren.

Danksagung

Ich möchte mich bedanken bei folgenden Personen für ihre Unterstützung vor oder während dieser Arbeit:

- Adrian Gschwend
- Christoph Müller
- Hartwig Thomas
- Bei meinem Advisor Daniel Debrunner, der es mir ermöglicht hat, diese spannende horizontenerweiternde Arbeit zu machen.
- Angela, die nicht viel von mir hatte in der letzten Zeit, weil ich immer dann weg war, wenn sie zu Hause war.
- Marcel Jacomet, der mich schon bei meinem Grundstudium unterstützt hat und mich motiviert und bestärkt hat, mein Master Studium in Angriff zu nehmen.
- All die freiwilligen Entwickler von MediaWiki und Semantic MediaWiki; ohne sie wäre diese Arbeit nie möglich gewesen.

Bedanken möchte ich mich auch bei folgenden Organisationen, die mich bei dieser Arbeit und das Public Domain Projekt im allgemeinen unterstützen. Ohne sie wäre das Projekt heute nicht da, wo es jetzt ist:

- Wikimedia CH
- ZHDK
- SOS-ETH
- Digitale Allmend

Inhaltsverzeichnis

1	Einleitung	11
1.1	Struktur des Dokuments	11
1.2	Ziele der Masterthesis	12
1.3	Definitionen	13
1.3.1	Begriffsdefinitionen	13
1.3.2	Was ist ein digitales Langzeitarchiv	16
1.3.3	Was bedeutet vertrauenswürdig (Trustworthy)	16
1.3.3.1	Vertrauen in die Institution	16
1.3.3.2	Vertrauen in die Echtheit der Information (Authenticity)	17
1.3.4	Vorgesehene Zielgruppe (Designated Communities)	17
1.3.5	Inhaltsinformation (Content Information)	17
2	Kontext dieser Masterthesis	19
2.1	Das Public Domain Projekt und die Schweizerische Stiftung Public Domain	19
2.1.1	Schweizerische Stiftung Public Domain	19
2.2	Was bedeutet gemeinfrei (Public Domain)	20
2.3	Projektarbeit 2	21
3	Vorbereitende Arbeiten	23
3.1	Aufwandsanalyse der vorgeschlagenen Systemarchitektur	23
3.2	Evaluation eines vertrauenswürdigen Speichersystems (Archival storage)	24
3.2.1	Bestehende Struktur	25
3.2.2	Stufen der Datenerhaltung (preservation level)	26
3.2.3	Evaluation der Dateiintegritätsprüfung	27
3.2.3.1	Anforderungen	27
3.2.3.2	Recherche	27
3.2.3.3	Entscheid	29
3.2.4	Evaluation eines Monitoringsystems	30
3.2.4.1	Anforderungen	30
3.2.4.2	Recherche	30
3.2.4.3	Entscheid	31
3.2.5	Evaluation eines Versionsverwaltungssystems für Multimediadaten	32
3.2.5.1	Anforderungen	32
3.2.5.2	Recherche	33
3.2.5.3	Entscheid	39
3.2.6	Vorgeschlagene Struktur	40
3.3	Migration des Archivspeichers	42
4	Bestandssicherung des digitalen Archivs	43
4.1	Archivspeicher (Archival Storage)	43
4.1.1	Aktuelle Situation	43
4.1.2	Hardware aufrüsten	44
4.1.3	Gentoo Installation	44
4.1.3.1	CFLAGS	44
4.1.3.2	USE Flags	45
4.1.3.3	Wahl der Linux Kernel Version	46
4.1.3.4	E-Mail Weiterleitung	46
4.1.3.5	Einrichten des RAID Verbundes	47
4.1.3.6	Neues Logical Volume Schema	48
4.1.3.7	Überwachung der Festplatten und des RAID Verbundes	48
4.1.3.8	Umstellung des Archivspeichers	50
4.1.3.9	NFS	51
4.1.3.10	Sperren des Schreibzugriffs auf archivierte Daten	51

4.1.4	Fernwartung mit IPMI Karte	51
4.2	Repräsentationsinformation	53
4.2.1	Repräsentationsinformation in Papierform	54
4.3	Umsetzung des Monitoringsservers	55
4.3.1	Hardware aufrüsten	55
4.3.2	Installation	55
5	Verfügbarmachen der Metadaten als Linked Open Data	57
5.1	Semantic Web und Linked Open Data	57
5.1.1	Qualitätsniveau von Open Data	58
5.2	Anforderungen der vorgesehenen Zielgruppen	58
5.3	Aktueller Stand der Metadaten im Public Domain Projekt	60
5.3.1	Semantic MediaWiki	60
5.4	Evaluation der einzusetzenden Standards für die Metadaten	61
5.4.1	Grundlegende Ontologien	61
5.4.1.1	Dublin Core (Dc, Dcterms)	61
5.4.1.2	Schema.org (Schema)	62
5.4.2	Spezialisierte Ontologien relevant für das Public Domain Projekt	63
5.4.2.1	Creative Commons Rights Expression Language (CC)	63
5.4.2.2	Music Ontology (mo)	63
5.4.2.3	Studio Ontology	63
5.4.2.4	Logistics Core (LogiCO)	64
5.4.3	Verwendete Ontologien bei verwandten Projekten	65
5.4.4	Nicht weiterverfolgte Metadatenstandards	65
5.5	Zuweisung von Strings anstelle von URIs	66
5.6	Implementation eines Prototyps auf Basis von LogiCO	67
5.6.1	Storage box Template	67
5.6.2	Storage box Formular	69
5.7	Implementation der ausgewählten Ontologien für bestehende Daten	71
5.7.1	Mapping zwischen Metadatenfeldern und Ontologien	71
5.7.1.1	Audio file Template	72
5.7.1.2	Audio storage description Template	73
5.7.1.3	Phonograph cylinder description Template	73
5.7.1.4	Storage media location report Template	74
5.7.1.5	License tags	74
5.7.2	Nicht erweiterte Templates	75
5.7.3	Bemerkungen zur Implementierung	75
5.8	Definition eines Datenmodells für das Public Domain Projekt	76
5.8.1	Probleme der bestehenden Templates	76
5.8.2	Neu zu erfassende Daten	76
5.8.3	Datenmodell	77
5.9	Entwicklung der neuen Templates auf Basis des Datenmodells	80
5.9.1	Vorgehen bei der Entwicklung der neuen Templates	80
5.9.2	Implementation der neuen Templates	80
5.9.2.1	Hilfstemplates für beteiligte Personen	80
5.9.2.2	Template Audio record	81
5.9.2.3	Template Archived item	82
5.9.2.4	Template Audio track	82
5.9.2.5	Template Digitized audio track	83
5.9.2.6	Template License investigation	84
5.9.2.7	Template Audio release set	85
5.10	Entwicklung von Formularen zum Datenmodell	87
5.10.1	Einschränken der möglichen Werte	87
5.10.2	Konzepte (Semantic MediaWiki concepts)	87
5.10.3	Formular Audio record	88

5.10.4	Formular Audio track	89
5.10.5	Formular Audio release set	91
5.10.6	Arbeitsablauf mit den neuen Formularen	92
5.11	Weiterführende Arbeiten	93
5.11.1	Anpassen/Automatisieren des Uploads in Wikimedia Commons	93
5.11.2	Automatisiertes Erstellen von Ogg/Vorbis Dateien	93
5.11.3	Bereitstellen eines SPARQL Endpoints	93
5.11.4	Ausbau der Metadaten von Open Data zu Linked Open Data	93
5.11.5	Extraktion und Bereitstellung von Audiofeatures	94
6	Dokumentenverwaltungssystem (DMS)	96
6.1	Cloud Lösung	96
6.2	Installation	96
6.3	Verschlüsselten Zugriff einrichten	98
6.3.1	Zertifikate von Let's encrypt	98
6.3.2	Konfiguration des Webservers	98
6.3.3	Prüfung	99
6.3.4	Erhalt der Sicherheit (Security)	99
7	Zusätzlich ausgeführte Arbeiten	100
7.1	Authentizität des Zugriffs sicherstellen	100
7.1.1	Prüfung	100
7.2	Korrekturen im MediaWiki	100
8	Schlussfolgerungen/Fazit	101
9	Literaturverzeichnis	102
10	Tabellenverzeichnis	102
11	Abbildungsverzeichnis	103
12	Lizenz	104
12.1	Lizenz der verwendeten Bilder	104
13	Selbständigkeitserklärung	105
14	Anhang	106
14.1	Einrichtung des MediaWiki Testsystems	107
14.1.1	Testserver	107
14.1.2	Entwicklungsrechner	107
14.1.2.1	Installation dnsmasq	108
14.2	Verwendete Ontologien von verwandten Projekten	109
14.3	MediaWiki Code zu den LogiCO Prototypen	111
14.3.1.1	Definition der externen Ontologie: Smw import LogiCO	111
14.3.1.2	Template Unit conversion length	111
14.3.1.3	Property mit Einheitenumrechnung: LogiCO:hasLengthValue	111
14.3.1.4	Property mit Wertbereichseinschränkung: LogiCO:hasLocation	111
14.3.1.5	Archive Rütli	112
14.3.2	Template Storage box	112
14.3.3	Template Storage box size	114
14.3.4	Concept Storage box	114
14.3.5	Formular Storage box	115
14.4	Beispieldaten	116
14.4.1	Beispieldaten wie sie bisher menschenlesbar im Browser dargestellt werden	116
14.4.2	Beispieldaten eines Lieds aus der MediaWiki API	118
14.4.3	RDF/XML Beispieldaten eines Lieds vor der Masterthesis	119
14.4.4	RDF Beispieldaten eines Lieds nach der Masterthesis	122
14.4.5	RDF Beispieldaten eines Lagerbehälters	131
14.5	Listen der erstellten oder geänderten MediaWiki Seiten	135
14.5.1.1	Archive	135
14.5.1.2	Demoseiten	135
14.5.1.3	Liste der importierten Ontologien	135

14.5.2	Liste der erstellten Properties	135
14.5.3	Liste der geänderten Templates	138
14.5.4	Liste der neu erstellten Templates	139
14.5.5	Liste der erstellten Formulare	140
14.5.6	Liste der neu erstellten Kategorien	140
14.6	Reguläre Ausdrücke für Datenmigration in semantische Templates	141
14.7	Anpassungen an der Konfiguration des MediaWiki Servers	142
14.7.1	Anpassen der Limiten von Semantic MediaWiki	142
14.7.2	Installation von MassEditRegex	142
14.7.3	Fehler zend_mm_heap corrupted beheben:	143
14.7.4	Setzen der korrekten Serveradresse	143
14.8	Apache2 Konfiguration	144
14.8.1	Vhost Konfiguration	144
14.8.2	SSL.conf	146
14.9	Security Testberichte von Qualys SSL Labs	149
14.10	Detailliertere Planung und Aufwandsabschätzung der Systemarchitektur	150
14.11	Aufgabenstellung Masterthesis	151

1 Einleitung

Diese Arbeit ist in Kooperation mit der Schweizerischen Stiftung Public Domain entstanden, welche das ehrenamtliche Public Domain Projekt zur Erhaltung von gemeinfreien Tonträgern unterhält. Unter Audiowerken sind hier Aufnahmen von musikalischen Darbietungen, gesprochener Sprache (Reden, Lesungen etc.) oder Geräuschen (Vogelgesang, typische Geräuschkulisse eines Marktes etc.) zu verstehen.

In der vorgängigen Projektarbeit wurden die grundlegenden Anforderungen für ein digitales Langzeitarchiv für Audiowerke erarbeitet. Dazu dient das etablierte OAIS Modell als Basis zusammen mit weiteren Arbeiten aus dem Bereich digitale Langzeiterhaltung. Der derzeitige Stand des Public Domain Projekts wurde mit einem Audit gemäss CCSDS 652.0-M-1 analysiert. Mit diesen zwei Teilarbeiten wurde dann eine neue Systemarchitektur für das Projekt entwickelt.

In dieser Masterthesis wird nun diese vorgeschlagene Systemarchitektur detaillierter ausgearbeitet und erste Teile davon praktisch umgesetzt.

Bemerkung:

Betrachtet wird in dieser Arbeit nur die digitale Langzeiterhaltung der Bestände des Public Domain Projekts. Die physische Erhaltung und die Digitalisierung der analogen Tonträger ist nicht Gegenstand dieser Arbeit.

1.1 Struktur des Dokuments

Dieses Dokument ist unterteilt in die Hauptbereiche:

- Einleitung, Hintergrund der Arbeit
- Vorbereitende Arbeiten
- Bestandssicherung des digitalen Archivs des Public Domain Projekts
- Öffentliches Verfügbarmachen semantisch aufbereiteter Metadaten und die Implementierung mit Hilfe von Semantic MediaWiki
- Implementation eines Dokumentenverwaltungssystems
- Fazit
- Anhang

1.2 Ziele der Masterthesis

Die Ziele dieser Masterthesis sind gemäss Aufgabenstellung:

- Evaluation und Vergleich von Linked Data Ontologien, die relevant sind für das Public Domain Projekt
- Entwicklung einer vereinheitlichten Vorlage (Templates) für Semantic MediaWiki basierend auf diesen Ontologien, Migration bestehender Daten und Implementation auf einem Testserver
- Evaluation und Implementation eines semi-automatisierten Dokumentenverwaltungssystem (Document management system, DMS) zur internen Verwendung für Administration, Verwaltung und Wartung
- Einrichten eines sicheren externen Zugangs zu diesem DMS unter Verwendung von aktuellen Sicherheitsmechanismen
- Evaluation eines vertrauenswürdigen Speichersystems für die digitalisierten Audiowerke und Entwicklung einer Migrationsstrategie, um dieses zu implementieren

Die Aufgabenstellung ist im Anhang 14.11 wiedergegeben.

1.3 Definitionen

1.3.1 Begriffsdefinitionen

Wie in jedem Wissenschaftsbereich sind exakte Begriffsdefinitionen wichtig, um eine gemeinsame Basis zu schaffen, damit unter dem selben Begriff auch das Selbe verstanden wird. Diese Definitionen dienen auch dazu, abstrahierte Konzepte zu definieren und zu etablieren.

In dieser Arbeit werden durchgehend die in der deutschsprachigen Übersetzung des OAIS Referenzmodells [NES13] definierten Begriffe benutzt. Bei der ersten Verwendung innerhalb eines Kapitels wird jeweils der englischsprachige Begriff in Klammern dazu gestellt. Dies soll die Einordnung und den Vergleich zu internationalen Publikationen vereinfachen sowie den fachlichen Austausch fördern.

Deutschsprachiger Begriff	English term	Bedeutung
OAIS	OAIS	Ein OAIS ist ein Archiv, das aus einer Organisation, die Teil einer größeren Organisation sein kann, aus Menschen und Systemen besteht, das die Verantwortung übernommen hat, Information zu erhalten und sie einer vorgesehenen Zielgruppe zugänglich zu machen.
Langfristig	Long Term	Eine Zeitspanne, die lange genug andauert, um sich mit den Auswirkungen des Technologiewandels inklusive der Unterstützung von neuen Datenträgern und Datenformaten sowie einer sich verändernden vorgesehenen Zielgruppe auf die Information im OAIS auseinander zu setzen. Diese Zeitspanne reicht bis in die unbestimmte Zukunft.
Langzeiterhaltung	Long Term Preservation	Die langfristige Erhaltung von Information in einer für die vorgesehene Zielgruppe unmittelbar verstehbaren Form, und mit Evidenznachweisen, die ihre Authentizität langfristig unterstützen.
Inhaltsinformation	Content Information	Ein Satz an Informationen, der das eigentliche Ziel der Erhaltung ist oder der Teile der oder die komplette Information enthält. Es ist ein Informationsobjekt, das sich aus dem Inhaltsdatenobjekt und seiner Repräsentationsinformation zusammensetzt.
Authentizität	Authenticity	Das Ausmass, in dem eine Person (oder System) ein Objekt als das ansieht, was es vorgibt zu sein. Authentizität wird auf der Basis von Evidenz beurteilt.
Unmittelbarverstehbar	Independently Understandable	Ein Merkmal von Information, die hinreichend vollständig ist, um von der vorgesehenen Zielgruppe interpretiert, verstanden und verwendet zu werden, ohne dass diese auf spezielle, nicht weit verbreitete, Hilfsmittel, einschliesslich benannter Personen, zurückgreifen muss.

Vorgesehene Zielgruppe	Designated Community	Eine ausgewiesene Gruppe potenzieller Endnutzer, die in der Lage sein soll, einen bestimmten Satz an Informationen zu verstehen. Die vorgesehene Zielgruppe kann sich aus mehreren Benutzergruppen zusammensetzen. Eine vorgesehene Zielgruppe wird vom Archiv definiert und diese Definition kann sich über die Zeit verändern.
Grundwissen	Knowledge Base	Eine Reihe von Informationen, verinnerlicht in einer Person oder einem System, die es der Person oder dem System erlauben, empfangene Information zu verstehen.
Repräsentationsinformation	Representation Information	Die Information, die ein Datenobjekt in für Menschen aussagekräftigere Konzepte übersetzt. Ein Beispiel von Repräsentationsinformation für eine Bitsequenz, die eine FITS-Datei ist, könnte sich aus einem FITS-Standard, der das Format definiert und einem Wörterbuch, das die Bedeutung von Schlüsselbegriffen definiert, die nicht Bestandteil des Standards sind, zusammensetzen. Ein weiteres Beispiel ist JPEG Software, die benutzt wird, um eine JPEG-Datei anzuzeigen. Die JPEG-Datei als Bits anzuzeigen ist für den Menschen nicht sehr aussagekräftig, aber die Software, die ein Verständnis des JPEG-Standards verkörpert, überträgt die Bits in Pixel, die dann als Bild für die menschliche Betrachtung angezeigt werden können.
Erhaltungsmetadaten (PDI)	Preservation Description Information (PDI)	Die Information, die benötigt wird, um die Inhaltsinformation angemessen zu erhalten, und die als Provenienz, Referenz, Beständigkeit, Kontext und Information über Zugriffsrechte kategorisiert werden kann.
Archivinformationspaket (AIP)	Archival Information Package (AIP)	Ein Informationspaket, bestehend aus der Inhaltsinformation und den dazugehörigen Erhaltungsmetadaten, das innerhalb eines OAIS aufbewahrt wird.
Übergabeinformationspaket (SIP)	Submission Information Package (SIP)	Ein Informationspaket, das vom Produzenten an das OAIS geliefert wird, um es zur Konstruktion oder zur Aktualisierung eines oder mehrerer AIPs und/oder den dazugehörigen Erschließungsinformationen zu benutzen.
Auslieferungsinformationspaket (DIP)	Dissemination Information Package (DIP)	Ein Informationspaket, abgeleitet aus einem oder mehreren AIPs, das als Antwort auf eine Anfrage an das OAIS von dem Archiv an den Endnutzer gesendet wird.
Digitale Migration	Digital Migration	Der Transfer digitaler Information innerhalb des OAIS mit dem Ziel ihrer Erhaltung. Sie unterscheidet sich von Transfer im Allgemeinen in drei Punkten: <ul style="list-style-type: none"> • einem Fokus auf der Erhaltung des gesamten Informationsgehalts, der Erhaltung benötigt; • einer Perspektive, dass die neue archivistische Erscheinung der Information ein Ersatz für die alte ist; und • dem Verständnis, dass die volle Kontrolle und Verantwortung über alle Aspekte des Transfers bei dem OAIS liegen.

Persistenz-information	Fixity Information	Die Information, welche die Mechanismen dokumentiert, die sicherstellen, dass das Inhaltsinformationsobjekt nicht unerlaubt verändert wurde. Ein Beispiel ist ein Schlüssel aus einer zyklischen Redundanzüberprüfung (CRC) für eine Datei.
Archivverbund	Federated Archives	Eine Gruppe von Archiven, die sich darauf verständigt hat, Zugriff auf ihre Bestände über eine oder mehrere gemeinsame Findmittel zu ermöglichen
Zugriffsprogramm	Access Software	Eine Art von Software, die Teile des oder den gesamten Informationsgehalt eines Informationsobjekts in für Menschen oder Systeme verstehbaren Formen präsentiert.

Tabelle 1: Die wichtigsten Begriffe für diese Projektarbeit aus dem OAIS Referenzmodell in deutsch und englisch. Aus [NES13], Seiten 8 bis 16

Die verschiedenen Metadatenstandards verwenden unterschiedliche Begriffe um das selbe Audioobjekt zu bezeichnen oder verwenden den gleichen Begriff für unterschiedliche Objekte. Darum werden hier die verwendeten deutschsprachigen Begriffe den verschiedenen Standards gegenübergestellt. Diese werden im Kapitel 5.4 näher beschrieben. Dublin Core ist nicht aufgeführt, weil alle aufgeführten Objekte zur Klasse Dcterms:Sound gehören.

Deutschsprachiger Begriff	Beschreibung	MusicBrainz	Schema.org	Music Ontology
Lied	Einzelnes Lied (Track) auf einem Tonträger oder aus einem digitalen Angebot. Nicht zu verwechseln mit dem abstrakteren Begriff des komponierten Werkes. Kann Musik sein, aber auch Reden, Naturaufnahmen etc.	Release	MusicRecording	Track
Tonträger	Veröffentlichter Tonträger (CD, Schallplatte, Zip Datei etc.). Es können Tonträger aus verschiedenen Ländern oder Auflagen separat existieren.	Release	MusicRelease	Record
Serie, Set	Gruppe von Tonträgern, die gemeinsam veröffentlicht wurden. Ein Album auf einer CD wird beispielsweise auf zwei Vinylplatten veröffentlicht.	Release	MusicRelease	Release
Veröffentlichung	Eine Single, EP oder Album. Zusammenstellung von aufgezeichneten Aufführungen (recorded performances), zusammen mit Begleitmaterial wie Booklets, Cover Bild etc.	Release	MusicAlbum	Release

Tabelle 2: Vergleich der Begriffe zur Beschreibung von Tonträgern in den Ontologien von MusicBrainz, Schema.org und Music Ontology

1.3.2 Was ist ein digitales Langzeitarchiv

Die DIN 31644 definiert ein digitales Langzeitarchiv so:

Organisation (bestehend aus Personen und technischen Systemen), die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit von Information in digitaler Form sowie die Bereitstellung für eine bestimmte Zielgruppe übernommen hat. [KEI13]

Was der Definition im englischsprachigen OAIS Referenzmodell entspricht:

[...] an Archive, consisting of an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. [OAIS12]

Wie aus der Begriffsdefinition zu entnehmen ist, bezeichnet *Langzeit* (long term) keine bestimmte Zeitspanne (Fünf Jahre, drei Generationen etc.), sondern ob die „*Auswirkungen des Technologiewandels inklusive der Unterstützung von neuen Datenträgern und Datenformaten sowie einer sich verändernden vorgesehenen Zielgruppe*“ relevant sind. Diese offener Definition greift damit viel weiter als klassische Archive/Bibliotheken und erfasst (bewusst) viel mehr Institutionen, Firmen, ja sogar einzelne Abteilungen/Personen.

Daraus folgt auch: Um diese Verantwortung übernehmen zu können, ist es lohnenswert, schon zum Zeitpunkt der Erschaffung von Information Überlegungen zu deren Langzeiterhaltung anzustellen.

1.3.3 Was bedeutet vertrauenswürdig (Trustworthy)

Damit von einer vertrauenswürdigen Langzeiterhaltung (Long Term Preservation) gesprochen werden kann, muss Vertrauen vorhanden sein sowohl gegenüber der archivierenden Institution, als auch in die Information, die von dieser archivierenden Institution erhalten wird.

1.3.3.1 Vertrauen in die Institution

Die vorgesehenen Zielgruppen und die Geldgeber müssen das Vertrauen haben, dass die Institution der übernommenen Verantwortung gewachsen ist. Es ist die Aufgabe der Institution, dies in geeigneter Weise darzulegen. Dazu gehören die Arbeitsweise, die Finanzierung, die Fähigkeiten der Mitarbeitenden, die technische Ausrüstung und wie die Institution diese Aspekte kommuniziert.

Daraus ergibt sich (ganz natürlich) die Erwartungshaltung der vorgesehenen Zielgruppen und der Geldgeber, dass eine Institution zur Transparenz verpflichtet ist und eine weitreichende Offenlegungspflicht hat.

1.3.3.2 Vertrauen in die Echtheit der Information (Authenticity)

Damit die Informationen aus einem Archiv genutzt werden können, muss von der vorgesehenen Zielgruppe nachvollzogen werden können, dass die Information wirklich das ist, was sie vorgibt zu sein: Die Authentizität muss belegt werden können.

Das ist keine neue Anforderung im Digitalen, dieser Anspruch stellt sich generell an ein Archiv. Daher ist dies ein altes Thema von Archiven mit einer eigenen Unterdisziplin in den Geschichtswissenschaften namens Diplomatie. Generell ist die Anforderung, dass die Herkunft (Wer, wann, warum, wo) und der Kontext der Information nachvollzogen werden kann und dass innerhalb des Archivs die Information nicht verändert wurde.

1.3.4 Vorgesehene Zielgruppe (Designated Communities)

Das Public Domain Projekt hat folgende vorgesehenen Zielgruppen definiert:

- Allgemeine Nutzergruppe (Global Community) mit Zugang zu einem Web Browser, HTML 4.0 fähig, Realschulabschluss oder höher, Sprachniveau für Englisch: A2
- Musikwissenschaftler, Historiker, Interpretationsforscher mit Zugang zu einem Web Browser, HTML 4.0 fähig, Schulabschluss: Abitur oder vergleichbar, Grundkenntnisse von Dublin Core, Sprachniveau für Englisch: B2
- Suchmaschinen, Metaarchive¹, Datenanalyseprogramme (Bots), die Abfragen per HTTP 1.1 stellen können und als Antwort HTML 4.0 oder RDF 1.1 (Serialisiert als RDF/XML) akzeptieren.

1.3.5 Inhaltsinformation (Content Information)

Das Public Domain Projekt übernimmt die Verantwortung, die ihr übertragenen digitalen Audiowerke zu erhalten. Die Inhaltsinformation von Audiowerken wird definiert als:

- Die akustische Information im Frequenzbereich, der von Menschen hörbar ist (15 Hz bis 20 kHz)
- Alle nötigen Begleitinformationen (Metadaten), um die Identität, die Herkunft, die Entstehung und die Authentizität zu bestimmen.

¹ Zugriff über gemeinsame Findmittel, z. B. per Wikimedia Commons, Memoriav Memobase oder europeana

2 Kontext dieser Masterthesis

2.1 Das Public Domain Projekt und die Schweizerische Stiftung Public Domain

Das Public Domain Projekt ist ein 2009 gegründetes Projekt *zur Erhaltung und Nutzbarmachung von gemeinfreien Musik- und Filmaufzeichnungen.*²

In diesem Projekt engagieren sich Freiwillige dafür, Tonträger aus den Anfängen der Schallaufzeichnung zu erhalten und diese Werke im Internet für alle zugänglich zu machen. Das Projekt schlägt so eine Brücke zwischen den Anfängen der Aufzeichnung von Tönen und dem Internetzeitalter.

Damit folgt dieses Projekt vergleichbaren Vorhaben, die mit ehrenamtlicher Arbeit Kulturgut erhalten, allen voran ist hier das Gutenberg-Projekt³ zu nennen, das seit 1971 gemeinfreie Bücher digitalisiert und verfügbar macht. Im Bereich Musik darf das International Music Scores Library Project (IMSLP)⁴ hervorgehoben werden, das Noten von derzeit über 100 000 Werken online zur Verfügung stellt. In weiten Kreisen bekannte Projekte, die mit freiwilligen Helfern grosse Leistungen erbringen, sind die freie Enzyklopädie Wikipedia oder die freie Weltkarte OpenStreetMap.

Ein Überblick zum Werdegang des Projektes ist in der Chronologie des Projektes⁵ nachzulesen.

2.1.1 Schweizerische Stiftung Public Domain

Das Projekt wird getragen von der Schweizerischen Stiftung Public Domain, die bezweckt:

- *gemeinfreie Ton- und Bildaufzeichnungen, insbesondere von Musik, Filmen und Rundfunksendungen schweizerischer Herkunft, zu sammeln, zu erhalten, und in der Schweiz zu verbreiten sowie öffentlich nutzbar und bekannt zu machen.*
- *Informationen über gemeinfreie Ton- und Bildaufzeichnungen zu sammeln, zu katalogisieren, und in der Schweiz zu verbreiten sowie öffentlich nutzbar und bekannt zu machen.*⁶

Die Stiftung übernimmt die Verwaltungsaufgaben und als langfristige Institution kümmert sie sich vor allem um die analogen Tonträgersammlungen, die der Stiftung von mehreren Sammlern überschrieben wurden. So sind derzeit über 50 000 Tonträger im Besitz der Stiftung.

² http://de.publicdomainproject.org/index.php/PD:%C3%9Cber_PUBLIC_DOMAIN_PROJEKT

³ <http://gutenberg.org>

⁴ <http://imslp.org>

⁵ http://de.publicdomainproject.org/index.php/PD:Chronologie_des_Projekts

⁶ http://de.publicdomainproject.org/index.php/Statuten#Art._2_Zweck

2.2 Was bedeutet gemeinfrei (Public Domain)

Im Namen des Projekts und der Stiftung kommt jeweils der Begriff *Public Domain* vor, es muss darum kurz darauf eingegangen werden, was darunter genau zu verstehen ist:

Public Domain (eigentlich «öffentlicher Grund», «Allmend») steht für diejenigen Inhalte, die nicht oder nicht mehr urheberrechtlich geschützt und damit frei verfügbar sind. Diese Inhalte sind gemeinfrei. Der Zugang zu Ihnen kann nicht durch das Urheberrecht begrenzt oder kostenpflichtig ausgestaltet werden.⁷

Der deutschsprachige Begriff, der rechtlich das Selbe meint, ist *gemeinfrei*.

In der Schweiz sind Werke gemeinfrei, wenn der Urheber mehr als 70 Jahre tot ist und 50 Jahre seit der ersten Veröffentlichung vergangen sind. Dies ist einer der Gründe wieso sich das Public Domain Projekt auf Tonträger aus der Frühzeit der Tonaufzeichnung konzentriert. So kann das Public Domain Projekt zu den archivierten Werken uneingeschränkten Zugang bieten.



Abbildung 1: Beispiel einer Zelluloid Walze: Ansicht auf die Verpackung und die Rippen

Abbildung 2: Beispiel einer Zelluloid Walze: Ansicht auf die Stirnseite.

Quelle: Nr. 4M 1047 der U.S. Everlasting Records. Titel: My Heart Has Learned To Love You. Musik komponiert von Ernest R. Ball (1878-1927), Liedtext von Dave Reed

<http://pool.publicdomainproject.org/index.php/Everlasting-4m-1047>

⁷ Antworten auf häufige Fragen Urheberrecht – Public Domain <https://www.ige.ch/urheberrecht/haeufige-fragen/public-domain.html>

2.3 Projektarbeit 2

In der vorgängigen Projektarbeit, [CHZ16], wurden die grundlegenden Anforderungen für ein digitales Langzeitarchiv für Audiowerke erarbeitet. Dazu diente als Basis das etablierte OAIS Modell, [NES13], zusammen mit weiteren Arbeiten aus dem Bereich digitale Langzeiterhaltung. Der derzeitige Stand des Public Domain Projekts wurde mit einem Audit gemäss CCSDS 652.0-M-1, [CCSDS652], analysiert.

Die Projektarbeit und das Audit wurden auf der Webseite des Public Domain Projekts veröffentlicht⁸.

Mit diesen zwei Teilarbeiten wurde anschliessend eine neue Systemarchitektur für die Langzeitarchivierung der digitalen Daten des Public Domain Projekt entwickelt. Abbildung 3 zeigt die Hauptkomponenten dieser Architektur und deren Schnittstellen.

⁸ http://pool.publicdomainproject.org/index.php/File:Projekt_2_-_Langzeitarchiv_für_digitale_Audiowerke_Christoph_Zimmermann_mit_Anhang.pdf
http://en.publicdomainproject.org/index.php/PD:Internal_CCSDS_652.0-M-1_audit

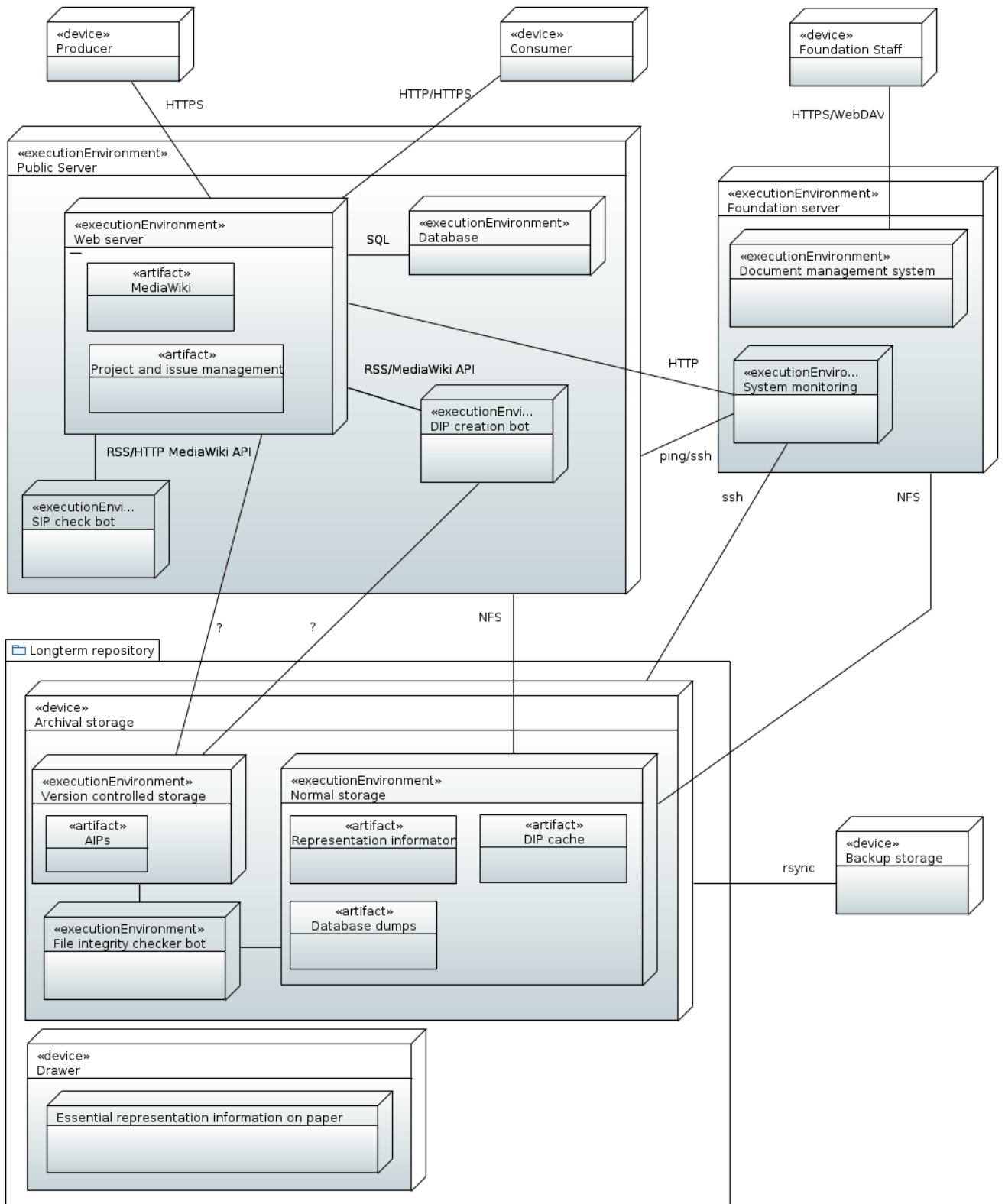


Abbildung 3: Vorgeschlagene Systemarchitektur zur Langzeiterhaltung im Public Domain Projekt [CHZ16], Seite 50

3 Vorbereitende Arbeiten

3.1 Aufwandsanalyse der vorgeschlagenen Systemarchitektur

In der Projektarbeit, die dieser Arbeit voranging, wurde nach eingängiger Analyse der akzeptierten Erhaltungsstrategien und Vorgehensweisen für digitale Archive eine Systemarchitektur vorgeschlagen, um diese Erhaltungsstrategien für das Public Domain Projekt umzusetzen.

Als erste Aufgabe in dieser Masterthesis wurde diese Systemarchitektur detaillierter geplant und in einzelne Projekte, Teilprojekte und Arbeitspakete zerlegt, um für die Umsetzung eine Aufwandsabschätzung machen zu können. Die Planungstabellen sind im Anhang 14.10 zu finden.

Es hat sich gezeigt, dass eine seriöse Entwicklung und Definition eines neuen AIP, basierend auf den Standards Matroska und Flac⁹, einen Aufwand entspricht von 540 h und entsprechend eine eigene Projektarbeit bzw. Masterthesis rechtfertigen würde.

⁹ Details dazu sind der Projektarbeit [CHZ16] ab Seite 48 zu finden.

3.2 Evaluation eines vertrauenswürdigen Speichersystems (Archival storage)

Der Archivspeicher ist eine der Kernkomponenten in der Langzeiterhaltung. Wie aus dem Audit hervorgeht, besteht beim Archivspeicher Verbesserungsbedarf. In [CHZ16] wurde vorgeschlagen, dass die archivierten Audiodateien versionsverwaltet werden sollen und dass die Infrastruktur überwacht (Monitoring) werden soll. Zudem müssen die Audiodateien regelmässig auf ihre Korrektheit überprüft werden. Diese Teile der neuen Struktur müssen genauer evaluiert werden.

Zudem soll aufgezeigt werden, wie die Migration vom jetzigen Archivspeicher zur angestrebten Struktur erfolgen soll.

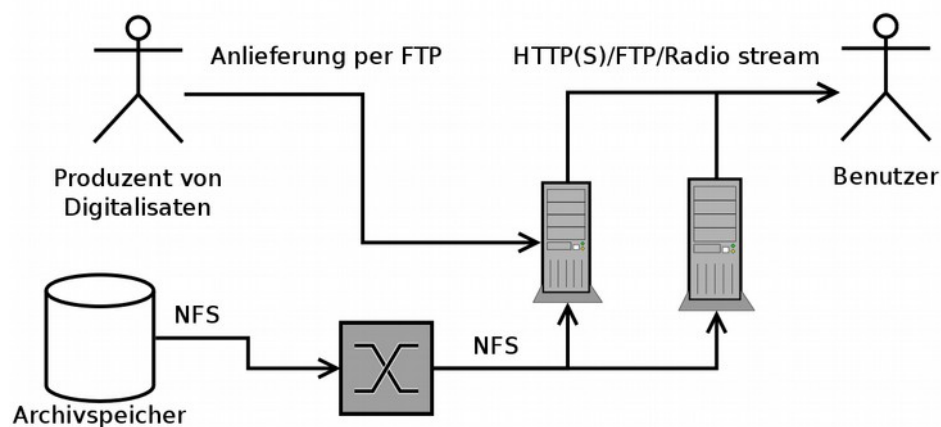


Abbildung 4: Verbindung des Archivspeichers mit den Produzenten und Benutzern

Die generellen Anforderungen an den Archivspeicher für das Public Domain Projekt sind in der Projektarbeit erarbeitet worden:

- *Die benötigte Repräsentationsinformation bis hinunter zum Bit-Level muss vorhanden sein*
- *Es muss nachvollziehbar sein wer, was, wann und wieso hinzugefügt oder geändert hat → Versionsverwaltung der gespeicherten Daten*
- *Verhindern von unautorisierten Veränderungen*
- *Risikomanagement für alle denkbaren Vorkommnisse, die die Daten gefährden könnten, dazu gehört mindestens ein Backup an einem geographisch entfernten Ort*
- *Einsatz von Free and Open Source Software (FOSS) [CHZ16], Seite 38*

Die Forderung für FOSS leitet sich direkt aus der ersten Anforderung, dem Vorhandensein der Repräsentationsinformation ab.

3.2.1 Bestehende Struktur

Derzeit setzt sich der Archivspeicher aus zwei Servern vom Hersteller Coraid zusammen, bestehend aus einem Server mit 7 Festplatten (Disc shelf) und einem zweiten Server (Storage controller), der auf den Disc shelf per ATA over Ethernet (AoE) zugreift.

Der Disc shelf hat 7 Festplatten, 6 davon in einem RAID 5 Verbund, die siebte Platte als aktive Reserveplatte (Hot-spare). Dieser Server läuft unter dem Betriebssystem Plan 9¹⁰, das von den Bell Labs entwickelt wurde und jetzt von Alcatel-Lucent weiterentwickelt wird (Alcatel-Lucent hatte 2006 Bell aufgekauft). Der Source Code von Plan 9 ist als Open Source veröffentlicht, trotzdem konnten keine genaueren Informationen zur Festplattenverwaltung oder zum RAID Format gefunden werden, das in Plan 9 verwendet wird. Es konnte auch kein Programm oder Treiber gefunden werden, der dieses Format in einem anderen Betriebssystem auslesen kann. Es ist daher derzeit unklar, ob der Disc shelf ein Software RAID von Plan 9 oder eine proprietäre Implementation vom Hersteller Coraid verwendet.

Es besteht also keine Möglichkeit, die geforderten Repräsentationsinformationen zu beschaffen. Diese wären nötig um den RAID Verbund in einem anderen System einzubinden, Reparaturen durchzuführen oder herauszufinden, ob beim Einsatz dieser RAID Implementation spezielle Risiken bestehen könnten, die beachtet werden müssen. Zudem bietet der Disc shelf auch keine Möglichkeit, die Datenstrukturen, die für den Betrieb des RAIDs auf der Festplatte benötigt werden, regelmässig zu überprüfen und Berichte darüber zu erstellen. Die Daten der Self-Monitoring, Analysis and Reporting Technology (SMART) der Festplatten im Disc shelf sind auch nicht zugänglich, was verhindert, dass Festplattenausfälle frühzeitig erkannt werden können oder noch vor dem Ausfall eingegriffen werden kann.

Der Storage controller greift per ATA over Ethernet (AoE) auf den Disc shelf zu. AoE ist vergleichbar mit iSCSI, beides sind Storage Area Network (SAN) Technologien, die auf bewährter, weit verbreiteter und daher günstiger LAN Technik basieren, im Gegensatz z. B. zu Fibre channel. AoE ist offen spezifiziert¹¹ und es existieren Open Source Implementationen für diverse Betriebssysteme. Der Storage controller läuft unter dem Betriebssystem Debian GNU/Linux 5 (Lenny). Seine Aufgabe ist es, die Disc shelves (in diesem Fall nur einen) zu verwalten und ihren Speicherplatz im LAN zur Verfügung zu stellen. Zur flexiblen Verwaltung des Speicherplatzes des RAID Verbundes wird das Linux eigene Logical volume mana-

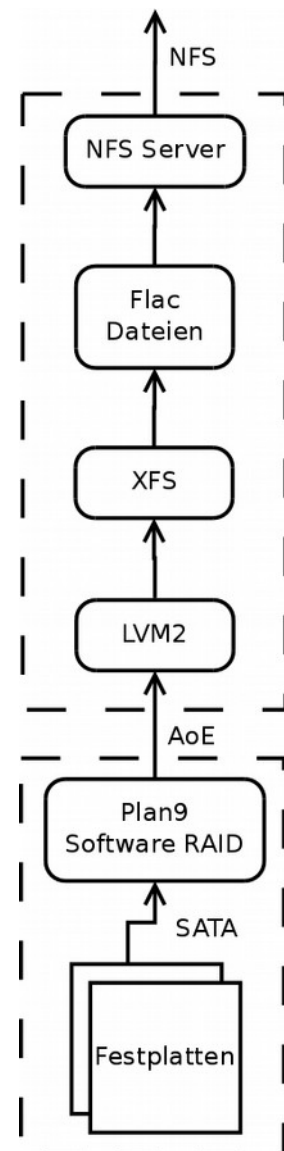


Abbildung 5: Bisherige Struktur des Archivspeichers bestehend aus Disc Shelf und Storage controller

¹⁰ <http://9p.io/plan9/>

¹¹ AoE protocol definition: <http://brantleycoilecompany.com/AoEr11.pdf>

gement 2 (LVM2) verwendet. Eine formale Spezifikation wurde dazu nicht gefunden, aber es ist schon seit vielen Jahren in Linux enthalten und hat eine grosse Verbreitung und Akzeptanz. Es wird von allen Desktop oder Serverdistributionen von Haus aus unterstützt und die entsprechenden Live-Versionen sind auch geeignet, um auf LVM2 Systeme zu greifen oder sie zu reparieren. Die mit LVM2 erstellten logischen Partitionen verwenden XFS¹² als Dateisystem, das ursprünglich von Silicon Graphics (SGI) entwickelt wurde und noch länger als LVM2 unter Linux eingesetzt wird. Die Dokumentation für das Bit-level Format, das auf der Festplatte verwendet wird, umfasst 80 Seiten und ist unter einer Creative Commons Lizenz veröffentlicht. Deshalb ist XFS für einen Archivspeicher gut geeignet.

Nicht mehr relevant für die Vollständigkeit der Repräsentationsinformationen ist das eingesetzte Protokoll, um die Daten des Archivspeichers im Netzwerk zur Verfügung zu stellen. Es soll hier trotzdem der Vollständigkeit halber erwähnt werden. Zum Einsatz kommt das Network File System (NFS), das ursprünglich von Sun Microsystems entwickelt wurde und als offener Internetstandard öffentlich verfügbar ist (RFC 1094, RFC 1813 und RFC 7931).

3.2.2 Stufen der Datenerhaltung (preservation level)

Bei der Auslegung eines Archivspeichers geht es vor allem um das Risikomanagement der Datenspeicherung auf unzuverlässigen Medien und die Sicherstellung der Interpretierbarkeit der Daten. Aber wie weit soll man dabei gehen? Wann ist ein Risiko genug minimiert?

Eine vermutlich vernünftige Antwort dazu wurde im PREMIS Standard gefunden. Bei PREMIS (Preservation Metadata: Implementation Strategies)¹³ handelt es sich um einen Standard, der gezielt für Erhaltungsmetadaten entwickelt wurde. Im PREMIS 3.0 Standard Dokument wird zur Datenerhaltung ein Vorschlag von drei Stufen mit steigendem Aufwand und kleiner werdendem Risiko gemacht:

One possible preservation level type is "Bit preservation level". This might have values of 'Low', 'Medium' or 'High', where, for example, in 2015 technology examples for:

- *'Low' means ordinary on-site backup*
- *'Medium' means two copies on different media types with a minimum of 150 km distance between the stored copies, with separate checksums that are integrity checked annually*
- *'High' means solutions a minimum of 5 independent copies on a variety of storage media distributed over different organizations in several continents with quarterly integrity checks. Quelle: PREMIS 3.0 final, Seite 52*

Mittel- bis Langfristig strebt das Public Domain Projekt die Stufe Mittel an.

¹² <http://xfs.org>

¹³ <http://www.loc.gov/standards/premis/v3/index.html>

3.2.3 Evaluation der Dateiintegritätsprüfung

3.2.3.1 Anforderungen

Da es sich um Audiodateien mit eingebetteten Metadaten handelt, soll es möglich sein, die Metadaten zu ergänzen/korrigieren und trotzdem den Audioinhalt auf unerwünschte Veränderungen zu prüfen.

An die Dateiintegritätsprüfung werden folgende Anforderungen gestellt:

- Zu neuen Dateien müssen automatisiert Prüfsummen generiert und aus dem Audioinhalt extrahiert werden.
- Die Prüfsummen müssen auf einem separaten Speichermedium abgelegt sein.
- Die Prüfsummen müssen ausserhalb des Schreibzugriffs von Produzenten oder Benutzern sein.
- Zur Überprüfung der Authentizität sollen Kopien der Prüfsummen den Benutzern zur Verfügung gestellt werden.
- Verwendung von Free and Open Source Software (FOSS)
- Die Prüfung soll die Integrität der archivierten Datei prüfen und ob sie auf unerlaubtem Weg verändert wurde (Korrekte Datei aber Prüfsumme stimmt nicht überein)
- Wenn möglich, sollen die geforderten, regelmässigen Dateiintegritätsprüfungen durch das Monitoringsystem (siehe Kapitel 3.2.4) ausgelöst werden und die Ergebnisse per E-Mail mitgeteilt werden.

3.2.3.2 Recherche

Die Recherche hat ergeben, dass es kein fertiges Werkzeug gibt, das alle genannten Anforderungen erfüllt. Es hat sich aber gezeigt, dass die Kombination aus den Flac Standardwerkzeugen und einem Programm zur Prüfsummenberechnung/prüfung ausreicht.

Eine ausführliche Erläuterung zu Prüfsummen, die in der Flac Datei enthalten sind, und einen Vergleich zu Prüfsummen über die gesamte Datei bietet etree.org:

Flac Fingerprint

- *A FLAC Fingerprint is generated only for the audio data portion of the file. (Therefore, changing the filename or the tags or FlacMetadata does not change the fingerprint calculation.) In contrast, an .md5 is generated against the whole file, including header portions.*
- *To create an ffp file using the standard .flac command line tools type "metaflac --show-md5sum flac_file_names > ffp.txt" When you generate a flac fingerprint file, it is merely a readout and compilation of the internally stored checksums from each of the flac files.*

SPECIAL NOTE ABOUT .MD5 FILES AND FLAC: Whether or not to make an .md5 checksum file for a .flac fileset is a confusing topic!

- *Why whole-file flac .md5s can be a hindrance: Under FLAC, you are allowed to change the compression ratio and add/remove meta data to .flac files without changing the actual audio. The audio may be identical, but the extra data will completely change the .md5 checksum.
[...]*
- *Why whole-file .md5 can be a help: They can still serve as a quick "parts list" since they will have a line for each flac file that is supposed to be in the fileset. If any files are missing, you can tell quickly. They help as a quick check for integrity of the whole file (not just the data part), so you can spot simple corruptions during uploads/downloads. It's also currently a good idea to add whole-file .md5 files to .flac sets that are uploaded to archive.org; it makes the whole upload/contribution process there run smoothly at several steps.
[...]*

Note that a flac fingerprint isn't a checksum of the encoded flac data – it's a checksum of the decoded music data. So to test the file, flac decodes the data in the file and verifies that the checksum of the music data matches the (internally stored) flac fingerprint. Quelle: <http://wiki.etree.org/?page=FlacFingerprint>

Auch um die Integrität einer Flac Datei zu prüfen, braucht es nichts weiteres als das Standard flac Programm, mit dem Flac Dateien en-/decodiert werden:

-d, --decode Decode (flac encodes by default). flac will exit with an exit code of 1 (and print a message, even in silent mode) if there were any errors during decoding, including when the MD5 checksum does not match the decoded output. Otherwise the exit code will be 0.

-t, --test Test (same as -d except no decoded file is written). The exit codes are the same as in decode mode. Quelle:

https://xiph.org/flac/documentation_tools_flac.html#decoding_options

Mit der Option -t wird also eine Flac decodiert und dabei geprüft, ob alle enthaltenen Prüfsummen übereinstimmen. Dazu muss die Datei decodiert werden, weil der oben erwähnte Fingerprint eine Prüfsumme der ursprünglichen Daten vor dem encodieren ist, der entsprechend nach dem decodieren wieder die selbe sein muss.

Es soll auch kurz diskutiert werden, welcher Algorithmus für die Prüfsummen verwendet werden soll. Einfache Cyclic-Redundancy-Check (CRC) sind weniger gut geeignet für größere Dateien. Der sehr verbreitete MD5 ist vergleichsweise schnell und sehr weit verbreitet. Er kann aber nicht mehr als kryptographischer Hashalgorithmus bezeichnet werden, in Security Anwendungen ist er daher zu vermeiden. Neuere Algorithmen wie SHA1, SHA256 sind

dazu besser geeignet, benötigen aber auch signifikant mehr Rechenleistung. Da in dieser Anwendung hauptsächlich Fehler durch Datenträgeralterung oder Anwenderfehler erwartet werden und grössere Datenmengen zu prüfen sind, wird MD5 als sinnvoller erachtet.

3.2.3.3 Entscheid

Da kein fertiges Werkzeug existiert, sollen zwei Werkzeuge entwickelt werden. Da die benötigten Arbeitsschritte alle von Kommandozeilenprogrammen erledigt werden können, bieten sich dazu Shellscripts oder andere Skriptsprachen an.

Das erste Skript dient zum Generieren von MD5 Werten der ganzen Flac Datei, extrahieren des Fingerprints aus dem Flac und speichern auf einem separatem Datenträger. Dies wird sinnvollerweise kombiniert mit der Überprüfung der geforderten technischen Anforderungen an archivierte Flac Dateien. Für dies kommt das Open Source Werkzeug MediaConch¹⁴ in Frage, das auch von mehreren Archiven/Bibliotheken eingesetzt und gefördert wird. Obwohl es für Matroska basierte AIPs ausgelegt ist, kann es auch für die Prüfung von anderen Formaten eingesetzt werden, da es auf MediaInfo¹⁵ basiert, das auch Flac unterstützt. In der Online Version von MediaConch ist auch ein Beispiel für einen Regelsatz publiziert¹⁶, der zeigt, wie die Abtastrate und die Bitbreite geprüft werden kann.

MediaConch is an extensible, open source software project consisting of an implementation checker, policy checker, reporter, and fixer that targets preservation-level audio-visual files (specifically Matroska, Linear Pulse Code Modulation (LPCM) and FF Video Codec 1 (FFV1)) for use in memory institutions, providing detailed and batch-level conformance checking via an adaptable and flexible application program interface accessible by the command line, a graphical user interface, or a web-based shell. Quelle:

<https://mediaarea.net/MediaConch/about.html>

Das zweite Skript dient dazu die archivierten Dateien gegen ihre gespeicherten Fingerprints und Prüfsummen zu prüfen. Es soll verschiedene Fehlercodes ausgeben, je nachdem ob nur die Prüfsumme oder auch der Fingerprint unterschiedlich ist. Dieses Skript kann in das Monitoringsystem integriert werden, was den Vorteil hat, dass der Gesamtzustand des Archivspeichers an einem zentralen Ort visualisiert werden kann.

¹⁴ <https://mediaarea.net/MediaConch>

¹⁵ <https://mediaarea.net/en/MediaInfo>

¹⁶ <https://mediaarea.net/MediaConchOnline/publicPolicies>

3.2.4 Evaluation eines Monitoringsystems

3.2.4.1 Anforderungen

Mit Monitoringlösungen bestehen noch keine Erfahrungen, weder beim Projektpartner noch beim Autor. Darum konnten keine sehr detaillierten Anforderungen formuliert werden.

An das Monitoring wurden folgende Anforderungen gestellt:

- Bei Fehlern sollen die Administratoren per E-Mail informiert werden
- Web-GUI zur Kontrolle und Fehleranalyse
- Erweiterbar, um nicht nur den Archivspeicher, sondern auch alle anderen eingesetzten Server, virtuellen Maschinen und deren Dienste zu überwachen
- Free and Open Source Software (FOSS)
- Wenn möglich, sollen die geforderten regelmässigen Dateiintegritätsprüfungen durch das Monitoring selbständig ausgelöst werden und die Ergebnisse per E-Mail mitgeteilt werden.

3.2.4.2 Recherche

Zum Vergleich von verschiedenen Systemen wurde neben einer Internetrecherche das Linux-Magazin Ausgabe 6/2015 benutzt, das zu Monitoringlösungen mehrere Artikel publizierte:

- Monitoring-Exit: Warum, wie und mit welchem Resultat Anwender Nagios den Rücken kehren¹⁷
- Warum Anwender ihre Monitoring-Lösung wechseln¹⁸
- Senu: Eine leistungsfähige und skalierbare Monitoring-Lösung¹⁹

Die bekannteste Lösung heisst „Nagios“²⁰. Folgende Argumente aus den Artikeln sprachen dagegen:

- *Überdies möchten manche Systembetreiber Status-quo-Messungen über längere Zeit speichern und einer Trendanalyse unterziehen* -> Dies ist mit Nagios nicht möglich, aber zur Überwachung des Archivspeichers wäre es sehr wertvoll/nötig.
- *Da das klassische Nagios kein Multithreading beherrscht und seine Checks seriell abarbeitet, zwingen die dabei anfallenden Laufzeiten zu sehr großen Checkintervallen.* -> Dies bedeutet, dass keine Dateiintegritätsprüfungen durchgeführt werden können, da diese Stunden bis Tage dauern.
- Viele Anwender berichteten von einer unübersichtlichen und mühsamen Konfiguration.

¹⁷ <http://www.linux-magazin.de/Ausgaben/2015/06/Einfuehrung>

¹⁸ <http://www.linux-magazin.de/Ausgaben/2015/06/Nagios-Umsteiger>

¹⁹ <http://www.linux-magazin.de/Ausgaben/2015/06/Sensu>

²⁰ <https://www.nagios.org/>

Gegen die vorgestellte Lösung Check_Mk sprach, dass sie aus mehreren Komponenten (Wato, PNP4Nagios, Nagvis) zusammengesetzt werden muss, was sie zwar sehr flexibel macht aber vermutlich schwieriger in der Einarbeitung und Pflege wäre.

Dem gegenüber steht Zabbix²¹, das einen kleineren Installationsaufwand besitzt. Es ist bei mehreren interviewten Systemadministratoren beliebt wegen der Einfachheit im Vergleich zu Nagios, wobei sich Zabbix bei ihnen auch im Alltag bewährt hat.

Im dritten Artikel im genannten Linux-Magazin wurde die Monitoringlösung „Sensu“²² vorgestellt. Diese Lösung basiert auf einer modernen, verteilten Softwarearchitektur, die in Abbildung 6 zu sehen ist. Gemäss diesem Artikel ist Sensu (auch) kompatibel mit Plugins von Nagios und es gibt auch eine grössere Anzahl von nativen Sensu Plugins²³. Es stehen Plugins für das im Projekt relevante LVM2, Linux RAID und SMART zur Verfügung. Das Plugin sensu-plugins-filesystem-checks²⁴ bietet auch die Möglichkeit, Dateien gegenüber ihren Hashwerten zu prüfen. Zusätzlich kann z. B. auch der online Radioserver Icecast2 überwacht werden.

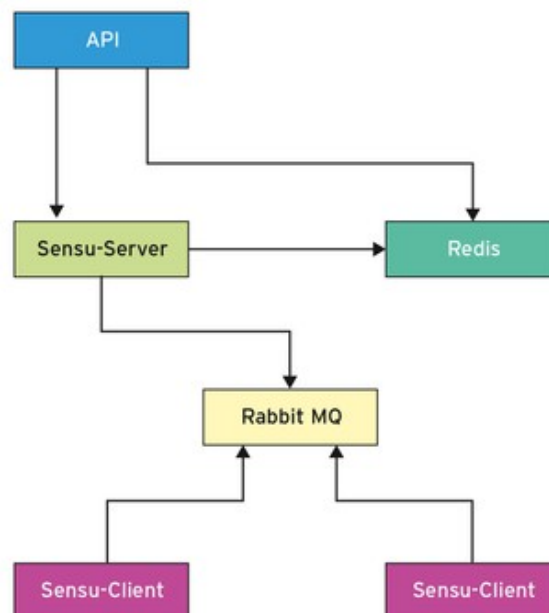


Abbildung 6: Architektur der Monitoringlösung Sensu

3.2.4.3 Entscheid

Es wurde entschieden, für weitere Tests und Untersuchungen Sensu zu installieren. Die moderne Architektur, die mit langen Prüfzeiten umgehen kann, ohne die anderen Tests zu beeinträchtigen, begünstigt diesen Entscheid. Für einen weiteren Test ist Zabbix in die engere Auswahl genommen worden.

Da es sich dabei erst um eine Testinstallation handelt, soll auf dem Archivspeicher trotzdem die Überwachung der Zustände der Festplatten und des RAID Verbundes eingerichtet werden mit den dafür geeigneten Werkzeugen smartd und mdadm. So entsteht keine Überwachungslücke während die Monitoringlösung ausgetestet wird.

²¹ <http://www.zabbix.com/>

²² <https://sensuapp.org>

²³ <https://github.com/sensu-plugins>

²⁴ <https://github.com/sensu-plugins/sensu-plugins-filesystem-checks>

3.2.5 Evaluation eines Versionsverwaltungssystems für Multimediadaten

3.2.5.1 Anforderungen

Wie in den generellen Anforderungen definiert, muss nachvollziehbar sein wer, was, wann und wieso hinzugefügt oder geändert hat. Das CCSDS 652.0-M-1 Audit fordert zusätzlich, dass eine dauerhafte Verknüpfung zwischen angelieferten Daten und den archivierten Daten (AIP) bestehen muss. Diese Anforderungen werden durch eine Versionsverwaltung²⁵ der gespeicherten Daten erfüllt.

4.4.1.1 The repository shall preserve the Content Information of AIPs.

[...] Discussion

The repository should be able to demonstrate that the AIPs faithfully reflect the information that was captured during ingest and that any subsequent or future planned transformations will continue to preserve all the required Information Properties of the Content Information.

One approach to this requirement assumes that the repository has a policy specifying that AIPs cannot be deleted at any time. [...] Either way, persistent links between the ingested object and the resulting AIP should be maintained. Quelle: [CCSDS652], Seite 52

An das Versionsverwaltungssystem wurden folgende Anforderungen gestellt:

- Umgang mit grossen Dateien mit binärem Inhalt (Die Flac Dateien sind je > 100 MiB)
- Skalierbarkeit, die Verwaltung von mehr als 50 000 Dateien (7,5 TiB) muss möglich sein. (Dies wurde aus der vorhandenen Anzahl Tonträger in der Sammlung der Schweizerischen Stiftung Public Domain abgeleitet)
- Unterstützung von Replikation auf ortsgetrennte Backupsysteme oder Backupdatenträger ohne Verlust der Versionsgeschichte
- Um Risiken zu minimieren, sollen die archivierten Dateien auch ohne Versionsverwaltungssystem auf dem Datenträger wiedergefunden werden (mindestens die aktuellste Version)
- Auf absehbare Zeit wird MediaWiki das Frontend bleiben. Darum müssen die archivierten Audiodateien aus MediaWiki²⁶ abgerufen werden können. Entweder direkt oder durch einen (zu entwickelnden) DIP Server als Gateway.
- Der Anlieferungsprozess (Ingest) muss über das Internet möglich sein mit den dazu nötigen Security Massnahmen. (Die Digitalisierung findet an verteilten Standorten statt)

²⁵ Unter dem Begriff Versionsverwaltung (version control) wird hier das Konzept gemeint, das aus dem Bereich der Softwareentwicklung bekannt ist.

²⁶ [https://www.mediawiki.org/wiki/Manual:\\$wgForeignFileRepos#Using_files_from_a_local_folder_-_FileRepo](https://www.mediawiki.org/wiki/Manual:$wgForeignFileRepos#Using_files_from_a_local_folder_-_FileRepo)

3.2.5.2 Recherche

Invenio

Aus dem Bereich der Systeme zur Verwaltung von digitalen Archiven wurden Dspace, Fedora Commons und Invenio genauer untersucht.

Bei Invenio handelt es sich um eine Entwicklung des CERN:

Invenio was born at CERN as a digital library software solution to run the CERN document server, managing over 1,000,000 bibliographic records in high-energy physics since 2002. Covering articles, books, journals, photos, videos, and more. [...] Manage audio, photo and video material. Create thumbnails and derived formats. Quelle:

<http://inveniosoftware.org/>

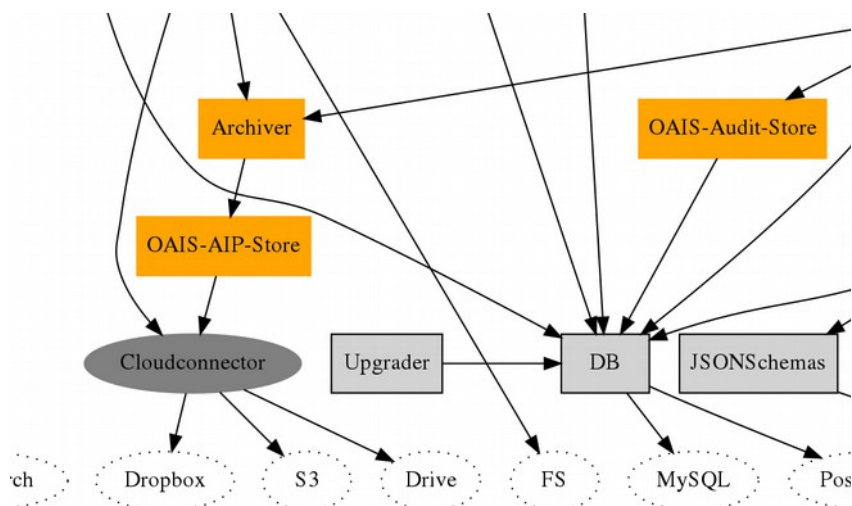


Abbildung 7: Ausschnitt aus der Systemübersicht von Invenio v3. Quelle: <https://invenio.readthedocs.io/en/latest/architecture/system-overview.html>

Die Untersuchung zeigte, dass die Systemstruktur sehr komplex ist und es wurden keine genauen Informationen gefunden, wie die Dateien verwaltet werden. Zudem wurden keine Informationen zu den versprochenen Features gefunden, um abgeleitete Formate zu generieren.

Dspace

Dspace is an out-of-the-box open source software package for creating repositories focused on delivering digital content to end users and providing a full set of tools for managing and preserving content within the application. DSpace is the most widely used repository software platform (open source or proprietary), with more than 2,000 installations worldwide representing a continuously growing and active user community. Quelle: <http://www.dspace.org/techspecs>

Der Aufbau von Dspace entspricht dem Systemmodell aus dem OAIS Standard [NES13], mit getrennten Einheiten für Anlieferung, Archivierung und Auslieferung. Die archivierten Daten werden direkt auf einem Dateisystem abgelegt, begleitet von Metadaten in einer relationalen

Datenbank. Die Untersuchung zeigte, dass zu allen Dateien Prüfsummen gespeichert werden und es darauf aufbauend eine integrierte Dateiintegritätsprüfung gibt. Eine Versionierung ist möglich und die neueste Version der Dateien kann zusammen mit ihren Metadaten als METS/XML in eine AIP Ordnerstruktur exportiert werden²⁷. Es wurden keine Informationen gefunden, wie die Dateien auf dem Dateisystem organisiert werden und es scheint schwierig zu sein, das Archiv mit vollständiger Versionsgeschichte auf ein entferntes System zu replizieren²⁸.

What metadata standards does DSpace support? Can I create metadata using the [SCORM or VRA or FGDC or MARC or myOwnSchema]?

In this context support for a given metadata schema means that metadata can be entered into DSpace, stored in the database, indexed appropriately, and made searchable through the public user interface. [...] Currently DSpace supports only the Dublin Core metadata element set with a few qualifications conforming to the library application profile. Quelle: <https://wiki.duraspace.org/display/DSPACE/User+FAQ>

Dspace verwaltet Daten und die dazugehörigen Metadaten. Es ist dabei aber beschränkt auf den Metadatenstandard Dublin Core (siehe Kapitel 5.4.1.1). Wie schon in [CHZ16] gezeigt wurde, reicht Dublin Core nicht aus, um alle Metadaten des Public Domain Projekts zu erfassen, was Dspace für das Projekt uninteressant macht.

Fedora Commons

Fedora Commons (kurz: Fedora) ist ein System zur Verwaltung eines Archivs, das vollständig auf RDF Daten basiert. Im Gegensatz zu Dspace können in Fedora auch Ressourcen verwaltet werden, zu denen keine Binärdaten sondern nur Metadaten vorhanden sind. Die Funktionsliste ist lang und sie wird hier nur Auszugsweise wiedergegeben:

- *Fixity Checking*
- *Versioning — Within Fedora 4, snapshots of the current state of a resource may be saved into the version history. The RDF for historic version snapshots may be browsed and old non-RDF content may be downloaded. Furthermore, an object or subgraph may be reverted to the state that it existed in a historic version.*
- *Backup and Restore — The Fedora 4 Backup capability allows a user, such as the repository manager, make a REST call to have the repository binaries and metadata exported to the local file system. [...] files are stored on disk named according to their SHA1 digest, so a filesystem backup approach can also be used.*
- *Linked Data Platform[...] Fedora 4 implements the LDP specification for create, read, update and delete (CRUD) operations, allowing HTTP, REST, and linked*

²⁷ <https://wiki.duraspace.org/display/DSDOC6x/AIP+Backup+and+Restore>

²⁸ <https://wiki.duraspace.org/display/DSPACE/BackupRestore>

data clients to make requests to Fedora 4. Quelle:
<https://wiki.duraspace.org/display/FEDORA47/Features>

Und weiter:

The Audit Service provides a mechanism for tracking repository events, recording critical details about them using predicates from the PREMIS ontology:

Which user performed the event, and using what client software

What kind of action was taken (creating objects, deleting files, etc.)

When was the action performed

What repository resource was changed

Quelle: <https://wiki.duraspace.org/display/FEDORA47/Optional+-+Audit+Service>

Die Anforderungen für Versionierung und Nachvollziehbarkeit werden also von Fedora erfüllt. Auch ist die Integritätsprüfung der archivierten Daten damit möglich.

In der Dokumentation sind auch detaillierte Informationen enthalten, wie die Daten verwaltet werden, welche Optionen für Binärdaten bereitstehen und wie darauf zugegriffen werden kann:

ModeShape is a distributed, hierarchical, transactional, and consistent data store with support for events, versioning, references, and flexible schemas. It is very fast, highly available, extremely scalable, and it is 100% open source and written in Java.[...]

Modeshape stores object metadata in an Infinispan cache. Binary content MAY be stored in Infinispan, or in an alternative BinaryStore. [...] Modeshape provides additional points of extensibility (e.g. Sequencers), and support for widely implemented APIs like CMIS, WebDAV, etc. [...]

Infinispan is the storage subsystem used by Modeshape for storing object structure, and (optionally) binary content. [...] Fedora 4 ships with a handful of example Infinispan configurations to get up and running quickly.

file no-frills, FileCacheStore backed

clustered a trivial, cluster-ready example; replicates metadata, distributes 2 copies of content Quelle: <https://wiki.duraspace.org/display/FEDORA47/FAQs+and+How-Tos>

Fedora kann also nicht nur die Dateien speichern, sondern auch gleich sicherstellen, dass mehrere Kopien davon existieren. Die Möglichkeit per WebDAV darauf zu greifen, vereinfacht das Einbinden in andere Frontendsysteme.

Fedora eignet sich dazu, nicht nur die zu archivierenden Flac Dateien, sondern gleich auch die dazugehörigen Metadaten (mit) zu speichern. Dazu bleiben derzeit viele Fragen offen, denn erst im Verlauf dieser Masterarbeit werden die Metadaten des Public Domain Projekts überhaupt im RDF Format vorliegen.

Die Systeme aus diesem Bereich liefern auch ihre eigenen Frontends mit, die durch eine eigene Gatewaylösung (DIP Server) zur Einbindung ins MediaWiki ersetzt werden könnten. Speziell Fedora Commons mit seiner REST API würde dies vereinfachen.

Boar

Aus dem Bereich der Versionsverwaltungssysteme für Binärdaten, die auf den Konzepten der Softwareentwicklung basieren, wurden boar und git-annex untersucht.

Boar²⁹ setzt auf ein zentrales Repository und ist vom Konzept her daher vergleichbar mit Subversion³⁰, ohne die Funktionalitäten um Textdateien zu vergleichen und zusammen zu führen, da dies bei Binärdaten nicht im generellen möglich ist.

Boar aims to be the perfect way to make sure your most important digital information, like pictures, movies and documents, are stored safely.

- *Boar makes it possible for you to restore any or all of your files from any point in time.*
- *Boar makes it easy to maintain verified backups of your data, including file history.*
- *Boar imposes no limits on file or repository sizes.*

[...] Boar stores snapshots of directory trees in a local or remote repository and provides tools to ensure that your data is consistent and complete. You can keep just some or all of your data checked out for viewing and editing.

The repository has a simple layout to ensure that the data can easily be extracted even if the original software should be unavailable. This simplicity makes boar ideal for data that needs safe long-term storage. Quelle:

https://bitbucket.org/mats_ekberg/boar/wiki/Home

Das bedeutet, dass die Arbeitsweise für die Datenübernahme eine andere ist, als bei allen anderen Systemen. Ein Produzent muss bei sich eine (Teil-) Kopie des Repositories (Working copy) haben, damit neue Daten hinzugefügt werden können. Es ist unklar, ob nur die Ordnerstruktur ohne die grossen Flac Dateien kopiert werden können. Im Prinzip würde dies auch für die Auslieferung an Endbenutzer durch einen Webserver gelten, aber Boar kann einzelne Dateien auch ohne Working copy mit dem Befehl `boar cat <Dateipfad>` ausliefern.

²⁹ https://bitbucket.org/mats_ekberg/boar/wiki/Home

³⁰ <https://subversion.apache.org/>

Git-annex

Bei git-annex³¹ handelt es sich um einen Zusatz zum Versionsverwaltungssystem git:

use case: The Archivist

Bob has many drives to archive his data, most of them kept offline, in a safe place.

With git-annex, Bob has a single directory tree that includes all his files, even if their content is being stored offline. He can reorganize his files using that tree, committing new versions to git, without worry about accidentally deleting anything.

When Bob needs access to some files, git-annex can tell him which drive(s) they're on, and easily make them available. Indeed, every drive knows what is on every other drive.

Bob thinks long-term, and so he appreciates that git-annex uses a simple repository format. He knows his files will be accessible in the future even if the world has forgotten about git-annex and git.

Run in a cron job, git-annex adds new files to archival drives at night. It also helps Bob keep track of intentional and unintentional copies of files, and logs information he can use to decide when it's time to duplicate the content of old drives.

Quelle: <https://git-annex.branchable.com/>

Wenn bei git-annex eine Working copy erstellt wird, wird alles was von git verwaltet wird, wie gewohnt auf den Client kopiert inkl. Versionsgeschichte. Zusätzlich wird die Versionsgeschichte der von git-annex verwalteten Dateien kopiert aber nicht die Daten selber. Es existieren nur symbolische Verknüpfungen zu diesen Dateien. Git-annex Dateien können wenn gewünscht (`git annex get`) kopiert werden. Um diese zu verändern müssen sie zuerst entsperrt werden (`git annex unlock`).

Git-annex ist in der Lage sicherzustellen, dass eine Mindestanzahl von Kopien existiert. Von den anderen untersuchten Systemen kann das nur Fedora Commons. Git-annex verweigert das Entfernen der Daten und der Versionsgeschichte, falls dies nicht garantiert werden kann. Dazu werden die entfernten Repositories zuerst kontaktiert, ob dies auch stimmt:

It's important that data not get lost by an ill-considered git annex drop command. So, git-annex can be configured to try to keep N copies of a file's content available across all repositories. [...] git annex drop attempts to check with other git remotes, to check that N copies of the file exist. If enough repositories cannot be verified to have it, it will retain the file content to avoid data loss. Quelle: <https://git-annex.branchable.com/copies/>

³¹ <https://git-annex.branchable.com>

Auf der Webseite von git-annex gibt es auch eine Liste von Vergleichen mit verwandten Projekten, darunter auch boar:

- *git-annex is not git-media, although they both approach the same problem from a similar direction. [...] It lacks git-annex's support for widely distributed storage, using only a single backend data store. It also does not support partial checkouts of file contents, like git-annex does. [...]*
- *Similarly, git-annex is not git-lfs, which also uses git smudge filters, and appears to lack git-annex's widely distributed storage and partial checkouts.*
- *git-annex is also not boar, although it shares many of its goals and characteristics. Boar implements its own version control system, rather than simply embracing and extending git. And while boar supports distributed clones of a repository, it does not support keeping different files in different clones of the same repository, which git-annex does, and is an important feature for large-scale archiving. Quelle: <https://git-annex.branchable.com/not/>*

Damit ein Webserver die verwalteten Dateien an Endbenutzer ausliefern kann, benötigt er einen Lesezugriff auf ein git-annex Repository, wo alle Dateien in der Working copy vorhanden sind. Dies wird vorzugsweise das Hauptrepository sein, in das alle Produzenten ihre Daten hinzufügen.

Digital Asset Management / weitere

Es wurden auch Lösungen aus dem Bereich „Digital Asset Management“ in die Recherche aufgenommen. Dabei sind Razuna³² und ResourceSpace³³ zu erwähnen, die beide das Konvertieren von verwalteten Audiodateien in andere Formate unterstützen (bei ResourceSpace auch automatisiert). Ein spezialisierter Vertreter aus dieser Kategorie ist Knora³⁴, eine Entwicklung des „Digital Humanities Lab“ der Universität Basel:

*Knora and Sipi (Simple Image Presentation Interface) are two **complementary** software projects. Whereas Knora deals with data that is written to and read from a triplestore (metadata and annotations), Sipi takes care of storing, converting and serving image files as well as other types of files such as audio, video, or documents (binary files it just stores and serves). Quelle:*

<http://www.knora.org/documentation/manual/rst/sipi/sipi-and-knora.html>

Es hat sich aber gezeigt, dass diese Produktgruppe nicht für den Archiveinsatz geeignet ist, weil generell zu wenig Aufmerksamkeit den Belangen der Langzeitarchivierung wie Versionierung und Nachvollziehbarkeit gewidmet wird. Razuna kann zwar versionieren, dies ist aber optional und muss für jede einzelne Datei manuell genutzt werden. Bei Knora/Sipi konnte in der Dokumentation nichts zu Versionierung gefunden werden.

³² <http://www.razuna.org/>

³³ <http://www.resourcespace.com>

³⁴ <http://www.knora.org/>

3.2.5.3 Entscheid

Als erster Schritt soll der Schreibzugriff eingeschränkt werden. Nach einer Bearbeitungsperiode von 30 Tagen sollen nur noch Administratoren die Rechte besitzen, eine archivierte Datei auf Antrag hin zu löschen oder zu ersetzen. Diese einfache Massnahme erzwingt ein Vier-Augen-Prinzip, was das Risiko einer unerlaubten oder unbemerkten Veränderung reduziert. Es verhindert effektiv einfache Benutzerfehler wie z. B. das versehentliche Löschen im FTP Programm.

Die Empfehlungen aus der Evaluation der Dateiintegritätsprüfung sollen unabhängig von einer Versionsverwaltung umgesetzt werden. Es soll sichergestellt sein, dass nur korrekte Daten in die Versionsverwaltung importiert werden.

Da die Einführung eines Versionsverwaltungssystems einen grossen Einfluss auf den Arbeitsablauf hat und im Fall von Matroska zusätzliche projektspezifische Software (DIP Server) entwickelt werden muss, müssen weitere Abklärungen und vor allem Tests durchgeführt werden. Darum wird vorgeschlagen, dass die Systeme git-annex und Fedora Commons in kleineren Testsystemen über eine längere Zeit ausgiebig getestet werden, wobei sie ihre Tauglichkeit beweisen müssen. Diese Tests müssen Themen der Handhabbarkeit für Produzenten, Replikation/ Backup, Vollständigkeit der Repräsentationsinformationen und Prüfung der Dateiintegrität umfassen.

3.2.6 Vorgeschlagene Struktur

Auf Basis der Struktur der vorgängigen Projektarbeit und der durchgeführten Evaluationen wird folgende Struktur für den Archivspeicher vorgeschlagen.

Der Archivspeicher besteht aus zwei Teilen:

Einem Gentoo GNU/Linux³⁵ basierten Server der die magnetischen Festplatten verwaltet, den Speicherplatz im Netzwerk zur Verfügung stellt und die Daten regelmässig auf undokumentierte Veränderungen prüft. Auf dem Server werden die AIPs in einem versionsverwalteten Speichersystem abgelegt. Dieses System stellt sicher, dass alle Änderungen an AIPs aufgezeichnet werden (Wer hat wann was geändert) und speichert von jeder AIP Version die Prüfsumme, um unerwünschte Veränderungen erkennen zu können. Auf dem Server werden in digitaler Form Repräsentationsinformationen zu den eingesetzten Dateiformaten, Dateisystem, Metadatenstandards und Vokabulare manuell abgelegt. Ergänzt wird die Repräsentationsinformation durch alle Quellcodes der auf dem Server vorhandenen Software. Dies geschieht automatisch, da Gentoo Linux jede Softwarekomponente als Quellcode herunterlädt und automatisiert kompiliert und installiert. So ist sichergestellt, dass die Repräsentationsinformation zur Software immer vollständig und aktuell ist. Anstatt Gentoo könnte auch FreeBSD eingesetzt werden, um dies zu erreichen.

Der zweite Teil ist eine Schublade die im Serverraum neben dem Gentoo GNU/Linux Server eingebaut wird. Sie enthält in Papierform die essentiellen Repräsentationsinformationen um die Daten im Server auslesen und interpretieren zu können. Dazu gehören ASCII Tabelle, Zahlensystembeschreibung, Spezifikation der Festplattenschnittstelle, Dateisystem und Dateiformatspezifikationen, Sprachreferenz für die Programmiersprache C und x86 Befehlssatz. [CHZ16], Seite 51

In der Bestandsaufnahme (Kapitel 3.2.1) wurde festgestellt, dass das bisher eingesetzte Dateisystem XFS und das System zur flexiblen Speicherplatzverwaltung LVM2 die Anforderungen an ein vertrauenswürdiges Langzeitarchiv erfüllen und entsprechend weiterhin verwendet werden sollen. Neu soll der RAID Verbund mit dem im Linux Kernel enthaltenen multi device Treiber (md) und den dazu gehörenden Werkzeugen (mdadm) eingerichtet werden. Md ist wie LVM2 ein etabliertes System, das schon viele Jahre im Linux Kernel enthalten ist und eine entsprechende Verbreitung und Bekanntheit besitzt. Das On-Disc Format ist nicht formal spezifiziert, das Metadaten Format (genannt Superblock) aber schon³⁶.

³⁵ <https://gentoo.org/>

³⁶ https://raid.wiki.kernel.org/index.php/RAID_superblock_formats

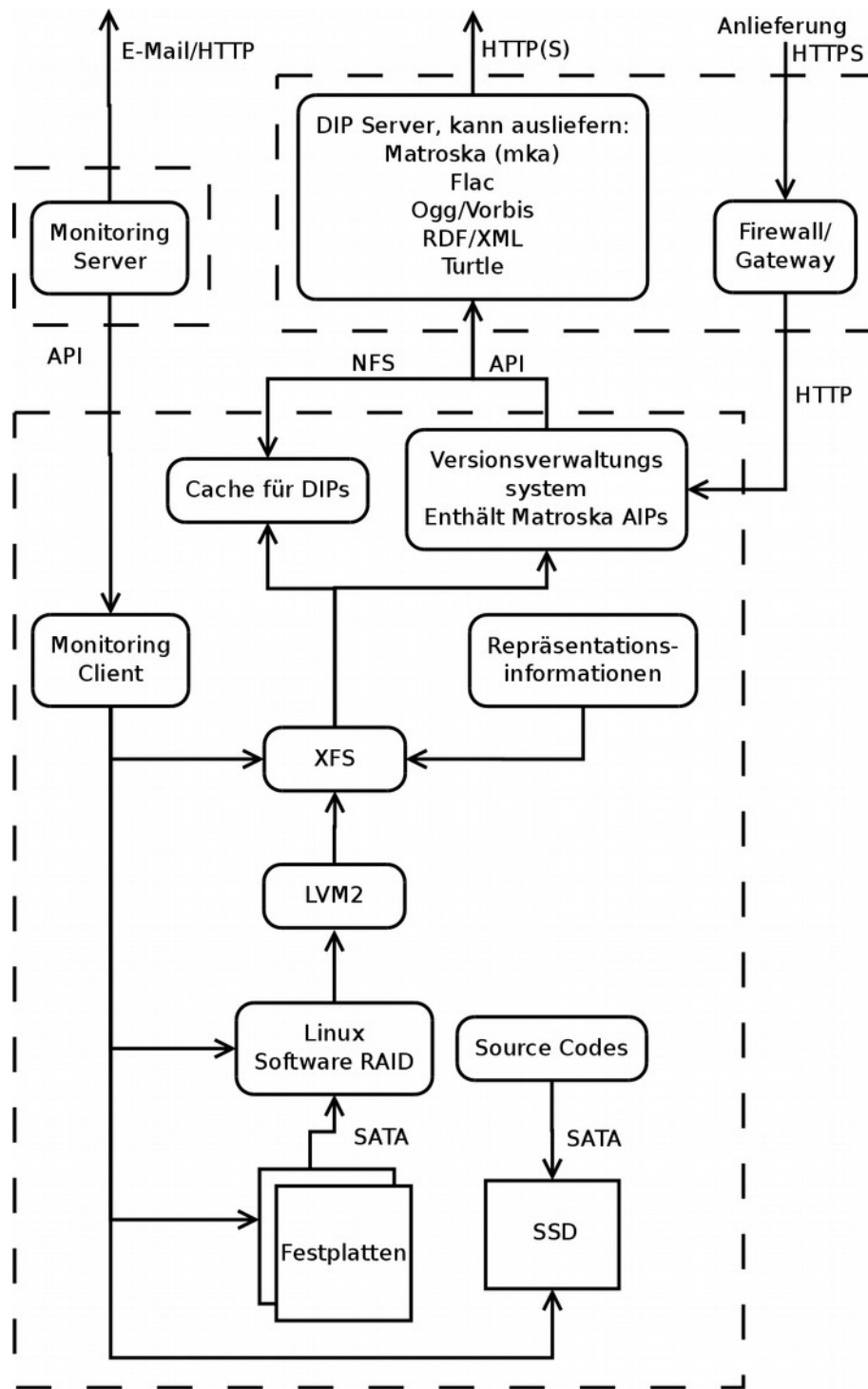


Abbildung 8: Geplante Struktur des Archivspeichers

Zusätzlich soll ein weiteres Speichersystem eingerichtet werden oder eine Partnerschaft eingegangen werden, um Sicherheitskopien aller Daten an einem getrennten Ort zu lagern. Die Distanz soll gross genug sein, dass Naturkatastrophen wie Erdbeben, Überschwemmungen etc. nur einen Standort betreffen (Ziel > 150 km).

3.3 Migration des Archivspeichers

Die Resultate des Audits, das im Rahmen der Projektarbeit durchgeführt wurde, haben gezeigt, dass die digitale Bestandssicherung ein dringendes Thema ist. Dazu gehört das regelmäßige Überprüfen der Daten auf ungewollte Veränderungen, das Vorhandensein genügend detaillierter Repräsentationsinformation und der Schutz gegen unbeabsichtigtes/unerlaubtes Löschen. Dieses Thema muss angegangen werden, bevor Migrationsschritte in Richtung eines neuen AIP oder versionierten Archivspeichers gemacht werden. Diese digitale Bestandssicherung stellt den ersten Schritt in Richtung neue Systemarchitektur dar.

Dazu wird das Speichersystem mit Gentoo GNU/Linux neu aufgesetzt, ohne an den Archivdaten auf den Festplatten etwas zu ändern, aber um Überwachungsfunktionen der verschiedenen Komponenten ergänzt. Die Zustände der Festplatten und des RAID Verbundes werden mit den dafür geeigneten Werkzeugen smartd und mdadm überwacht und Fehler per E-Mail den Administratoren gemeldet. Diese Migration wurde während dieser Masterarbeit durchgeführt und ist in Kapitel 4 beschrieben.

Als nächster Schritt soll ein Monitoringsystem implementiert werden, das den Systemzustand kontinuierlich überwacht und regelmässig die Dateiintegrität der archivierten Daten prüft. Dazu ist es nötig, dass zu neu angelieferten Dateien Prüfsummen erstellt werden.

Danach bestehen mehrere Optionen:

1. Es wird ein AIP auf Basis von Matroska definiert und die bestehenden Flac Dateien migriert und in einem späteren Schritt wird ein System zur Versionsverwaltung eingeführt
2. Es wird erst ein System zur Versionsverwaltung eingeführt und die Flac Dateien darin untergebracht und danach wird ein AIP auf Basis von Matroska definiert und die bestehenden Flac Dateien migriert.

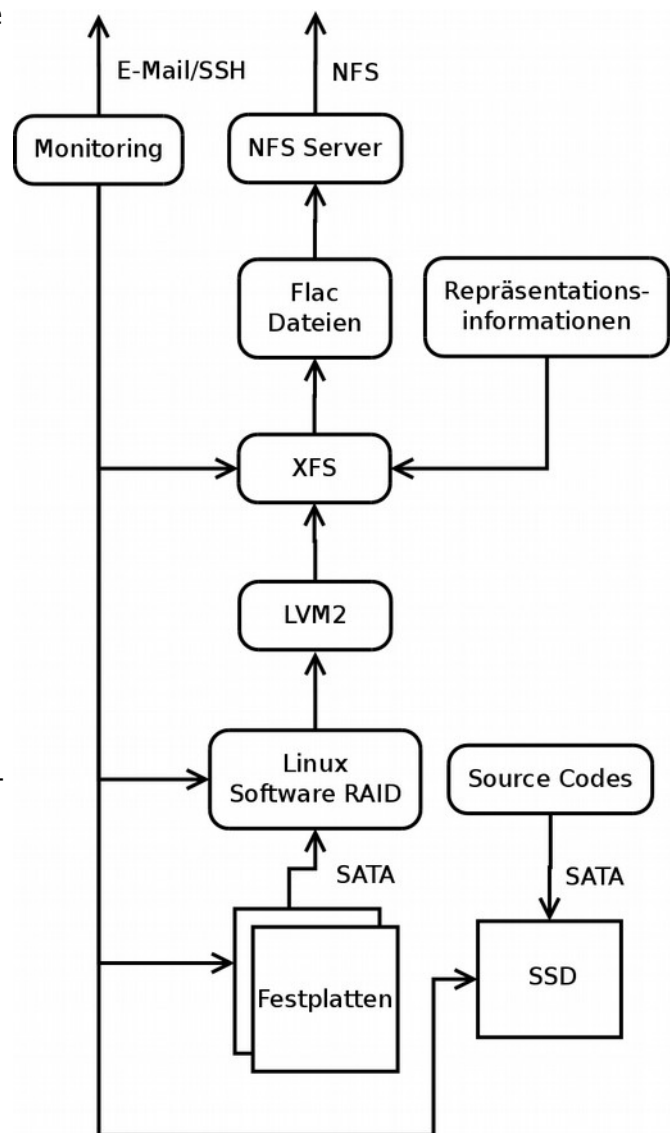


Abbildung 9: Erster Migrationsschritt hin zur neuen Struktur des Archivspeicher

Vorteil der zweiten Option ist, dass der Migrationsschritt schon im Versionsverwaltungssystem nachvollzogen werden kann.

4 Bestandssicherung des digitalen Archivs

Das Ziel war es, den ersten Migrationsschritt umzusetzen, wie er in Abbildung 9 zu sehen ist.

4.1 Archivspeicher (Archival Storage)

4.1.1 Aktuelle Situation

In den Unterlagen zum Coraid Speichersystem konnten keine Angaben gefunden werden zur verbauten Hardware. Einzig, dass es sich um normale x86 Serverhardware handeln muss.

Informationen zu den verwendeten Komponenten im SR1521 konnten von Fotos gewonnen werden, die im Internet veröffentlicht wurden. Für den angedachten Umbau war dies aber nicht ausreichend. Daher wurden bei einem Besuch im Serverraum der Zürcher Hochschule der Künste (ZHDK) die fehlenden Informationen vor Ort erfasst.



Abbildung 10: Coraid SR1521 Disc shelf und CLN/21 Storage controller

Hardware im Coraid SR1521 Disc shelf:

- Mainboard Super Micro X7DVL-E
 - Dual Intel® 64-bit Xeon LGA 771 quad core/ dual core processors at a front side bus speed of 1333 MHz/1067 MHz/667MHz
 - Memory Six 240-pin DIMM sockets with support up to 24 GB ECC Fully Buffered DDR2 667/533 Memory
 - SIMLC IPMI Slot
- 2 x SATA Controller mit je 8 Anschlüssen: Super Micro SAT2-MV8
 - Marvell 88SX6081 Chip
 - Passender Linux Treiber heisst: sata_mv
- Von Coraid angepasstes Plan9³⁷ Betriebssystem auf einem parallel ATA Flashmodul

³⁷ „The SR1521 is a Plan 9 system installed on a PC that serves up to 15 750Gbytes SATA disks over 1000bT using AoE (ATA over Ethernet).“ <http://lsub.org/sr.html>

- Unbekannte Menge RAM
- Unbekannter CPU Typ

4.1.2 Hardware aufrüsten

Um den Disc shelf für die zusätzlichen Aufgaben als Archivspeicher fit zu machen, wurde die Hardware aufgerüstet.

Damit der Archivspeicher unabhängig vom Betriebssystem aus der Ferne gewartet werden kann, wurde eine IPMI Karte nachgerüstet. Die passende Zusatzkarte von Super Micro hat die Typbezeichnung AOC-SIM1U+.

Da die Grösse des Arbeitsspeichers im SR1521 Disc shelf unbekannt war, wurden vorsorglich 2 x 4 GiB RAM gemäss Kompatibilitätsliste von Super Micro bestellt und installiert (Micron FB-DIMM 4 GB PC2-5300F ECC). Da ein unbekannter CPU Typ verbaut war, wurde ein passender Quad-Core CPU bestellt und installiert (Intel XEON E5430 4x 2.66 GHz / 12 MiB L2 / 1333 MHz FSB / Socket 771)

Das alte parallel ATA Flashmodul besitzt nur 32 MiB Speicherplatz und wurde ersetzt durch eine 60 GiB SATA SSD (Typ Corsair Force LS). Dies ermöglichte es auch, das neue Betriebssystem für den Archivspeicher auf dem Testrechner aufzusetzen, ohne den Speicher vom Netzwerk nehmen zu müssen. Wenn bei der Migration ein Problem aufgetreten wäre, hätte bis zum Erstellen des RAID Verbunds jederzeit wieder das alte Flashmodul eingesetzt werden können, um dann in Ruhe das Problem zu lösen.

4.1.3 Gentoo Installation

Im Vergleich zu anderen Distributionen besitzt Gentoo kein Installationsprogramm, sondern es müssen einzelne Installationsschritte von Hand gemäss Anleitung³⁸ abgearbeitet werden.

Dafür bestand die Möglichkeit, während der Installation Entscheidungen zu treffen, die den Einsatz als Archivspeicher begünstigen.

4.1.3.1 CFLAGS

In Gentoo können die Compiler Einstellungen, die für das ganze Betriebssystem genutzt werden, angepasst werden. Üblicherweise dient das zur Leistungssteigerung, da jede Software optimal für den verwendeten CPU Typ gebaut wird.

Für einen Archivspeicher ist dies nicht erwünscht, da dann im Problemfall das Betriebssystem nicht auf einer anderen Mainboard/CPU Kombination lauffähig ist. Darum wurde bewusst kein spezifischer CPU Typ ausgewählt, damit nur für das generelle Ziel x86/AMD64 kompiliert wird.

Dies wird über folgende Einträge in der Datei /etc/portage/make.conf konfiguriert:

```
CHOST="x86_64-pc-linux-gnu"
CFLAGS="-O2 -pipe"
```

³⁸ <https://wiki.gentoo.org/wiki/Handbook:AMD64/Full/Installation>

4.1.3.2 USE Flags

Ein grundlegendes Konzept von Gentoo ist der Einsatz von sogenannten USE Flags:

Such a flag is a keyword that embodies support and dependency-information for a certain concept. If someone defines a certain USE flag, Portage will know that they want support for the chosen keyword. Of course this also alters the dependency information for a package. Quelle: <https://wiki.gentoo.org/wiki/Handbook:AMD64/Working/USE>

Auf dem Archivspeicher wurden die USE Flags so ausgewählt, dass:

- die Repräsentationsinformation möglichst ausführlich ist
- keine Binärpakete installiert werden können, für die kein Source Code vorliegt
- das installierte System nicht abhängig ist von der aktuell vorhanden Hardware

Details zu einzelnen ausgewählten USE Flags:

```
-bindist - Flag to enable or disable options for prebuilt (GRP) packages
(eg. due to licensing issues) -> deaktivieren, damit keine Binärpakete be-
nutzt werden, wo keine Repräsentationsinformationen vorhanden ist.
-branding - Enable Gentoo specific branding
+doc - Add extra documentation (API, Javadoc, etc). It is recommended to en-
able per package instead of globally
+examples - Install examples, usually source code
+ipv6
+nfsdclld Enable nfsdclld NFSv4 clientid tracking daemon
+nfsidmap Enable support for newer nfsidmap helper
+nfsv4 Enable support for NFSv4
+nfsv41 Enable support for NFSv4.1
```

Die vollständige USE Flag Konfiguration:

```
USE="acpi -adobe-cff bash-completion -bindist -branding -cairo doc examples
hddtemp -introspection ipv6 lm_sensors latex nfsdclld nfsidmap nfsv4 nfsv41
uuid -wayland -X"
```

Nach der Konfiguration der CFLAGSs und der USE Flags muss das ganze Betriebssystem aus den Source Codes neu gebaut werden, damit die Änderungen wirksam werden. Dies nimmt etwas Zeit in Anspruch, läuft aber vollständig automatisiert ab. Dazu dient folgendes Kommando:

```
emerge --ask --update --deep --newuse @world
```

Leider hat das Aktivieren des USE Flag „doc“ zu zirkulären Abhängigkeiten geführt, die das Buildsystem nicht automatisch auflösen konnte. Um diese aufzulösen, musste temporär für gewisse Pakete dieses USE Flag deaktiviert werden. Dies geschah durch folgende Einträge in der Datei /etc/portage/package.use:

```
dev-python/Babel -doc
dev-python/six -doc
dev-python/jinja -doc
dev-python/pygments -doc
dev-tex/xcolor -doc
app-doc/doxygen -doc -latex
```

```
dev-lang/nasm -doc
dev-libs/icu -doc
media-libs/fontconfig -doc
app-text/texlive-core -xetex
>=app-text/texlive-core-2014-r4 -xetex
dev-python/sphinx -latex
```

Nach dem das Betriebssystem erfolgreich neu gebaut wurde, konnte diese Datei wieder entfernt werden und das Buildsystem war in der Lage in einem zweiten Durchgang diese Pakete mit dem USE Flag „doc“ neu zu bauen.

4.1.3.3 Wahl der Linux Kernel Version

Eine weitere Auswahlmöglichkeit besteht bei der Linux Kernel Version. Für den Archivspeicher wurde ein Kernel aus dem Longterm Zweig gewählt. Diese werden länger als andere Versionen gepflegt und sind entsprechend besser getestet und enthalten weniger Fehler.

There are usually several "longterm maintenance" kernel releases provided for the purposes of backporting bugfixes for older kernel trees. Quelle:

<https://www.kernel.org/category/releases.html>

Aktuell ist das der Linux Kernel 4.4 mit Support bis Februar 2018.

4.1.3.4 E-Mail Weiterleitung

Damit Fehlermeldungen per E-Mail an die Administratoren gemeldet werden können, muss eine E-Mail Weiterleitung an einen richtigen E-Mail Server (Smarthost) eingerichtet werden. Ohne diese Weiterleitung würden solche E-Mails nicht beim Empfänger ankommen, da sie übliche Eigenschaften von Spam Mails besitzen (Der Quellserver ist nicht als E-Mail Server bekannt, hat keinen zugehörigen DNS Eintrag etc.).

Dazu wird die kleine Software sSMTP eingesetzt, dessen einziger Zweck genau das Weiterleiten von E-Mails an einen Smarthost ist.

Die Konfiguration erfolgt in der Datei /etc/ssmtp/ssmtp.conf:

```
# /etc/ssmtp.conf -- a config file for sSMTP sendmail.
#

# The person who gets all mail for userids < MinUserId
# Make this empty to disable rewriting.
root=archiveadmins@pdproject.org

# All mail delivered to userid >= MinUserId goes to user, not root.
#MinUserId=1000

# The place where the mail goes. The actual machine name is required
# no MX records are consulted. Commonly mailhosts are named mail.domain.com
# The example will fit if you are in domain.com and your mailhub is so named.
mailhub=apollon.swiss-server.net
```

```

AuthUser=archive@pdproject.org
AuthPass=xxxxxxxxxxxxxxxx
AuthMethod=CRAM-MD5

# The full hostname

# Gentoo bug #47562
# Commenting the following line will force ssmtp to figure
# out the hostname itself.

hostname=archive.publicdomainproject.org

# Set this to never rewrite the "From:" line (unless not given) and to
# use that address in the "from line" of the envelope.
FromLineOverride=YES

# Use SSL/TLS to send secure messages to server.
UseTLS=YES
useSTARTTLS=YES

/etc/ssmtp/revaliases
# sSMTP aliases
#
# Format:      local_account:outgoing_address:mailhub
#
# Example: root:your_login@your.domain:mailhub.your.domain[:port]
# where [:port] is an optional port number that defaults to 25.
root:archiveadmins@pdproject.org:apollon.swiss-server.net
mainuser:archiveadmins@pdproject.org:apollon.swiss-server.net

```

Nach der Einrichtung wurde festgestellt, dass der cron Dämon³⁹ versuchte, viele E-Mails zu senden, die für die Administratoren unerwünscht und uninteressant wären. Cron wurde umkonfiguriert damit, diese Informationen ins Systemlog (syslog) umgeleitet werden. Dies wird einfach durch die Option „-s“ in der Datei /etc/conf.d/cronie erreicht⁴⁰:

```

# Settings for the CRON daemon.
# CRONDARGS= : any extra command-line startup arguments for crond
CRONDARGS=-s

```

4.1.3.5 Einrichten des RAID Verbundes

Da für die eingesetzten Festplatten keine Ersatzteile beim Public Domain Projekt an Lager sind und der Archivspeicher genügend freien Platz aufwies, wurde entschieden, dass der neue RAID Verbund aus einer Festplatte weniger als der Alte besteht. Somit konnte nach Abschluss der Umstellung eine Festplatte aus dem Archivspeicher entfernt werden und zum Ersatzteillager hinzugefügt werden.

Es wurde somit ein RAID5 Verbund mit sechs Festplatten mit je 750 GiB eingerichtet⁴¹. Dieser kann danach unter /dev/md0 angesprochen werden:

```
mdadm --create /dev/md0 --level=5 --raid-devices=6 /dev/sdc /dev/sde ...
```

³⁹ Der Cron Dämon ist für das regelmässige, zeitgesteuerte Ausführen von Programmen/Skripts zuständig

⁴⁰ <https://bbs.archlinux.org/viewtopic.php?id=56587>

⁴¹ https://wiki.gentoo.org/wiki/Complete_Handbook/Software_RAID

Wie beim alten RAID Verbund wurde eine weitere Festplatte als Hot-spare zugewiesen. Diese Festplatte wird vom RAID System automatisch aktiviert, wenn eine andere Festplatte ausfällt und die Redundanz wieder hergestellt:

```
mdadm /dev/md0 --add /dev/sdi
```

4.1.3.6 Neues Logical Volume Schema

Die Migration des Archivspeichers konnte auch gleich genutzt werden, um die Logical Volume besser den aktuellen Bedürfnissen anzupassen. Auf dem alten System existierte nur ein einzelnes Logical Volume für alles. Dies war ein Kompromiss, da noch niemand wusste, wie die Bedürfnisse sind und wie sich diese entwickeln werden.

Neu wurden mehrere Logical Volumes erstellt⁴² gemäss den Bedürfnissen die aus der Speicherplatzbelegung abgelesen werden konnten. Wie üblich, wurde nicht aller Festplattenplatz zugeteilt, so dass die einzelnen Logical Volumens flexibel nach ihrem zukünftigen Bedarf vergrössert werden können:

```
lvcreate -L 300G -n archive vg0
lvcreate -L 300G -n backup vg0
lvcreate -L 50G -n swissfoundationpublicdomain vg0
```

4.1.3.7 Überwachung der Festplatten und des RAID Verbundes

Bei Problemen mit Festplatten sollen die Administratoren per E-Mail informiert werden. Dazu wird der Zustand des RAID Verbundes mit dem Programm mdadm überwacht, welches Alarm schlägt, wenn eine Festplatte komplett ausfällt, entfernt wird etc. Solche Ausfälle werden von mdadm zeitnah erkannt und gemeldet. Die Konfiguration in der Datei /etc/conf.d/mdadm muss dazu wie folgt angepasst werden:

```
# /etc/conf.d/mdadm: config file for /etc/init.d/mdadm

# Misc options to pass to mdadm in monitor mode.
# For more info, run `mdadm --monitor --help` or see
# the mdadm(8) manpage.

MDADM_OPTS="--syslog --mail=archiveadmins@pdproject.org"
```

Die meisten Ausfälle von Festplatten kündigen sich an, noch während die Festplatte funktionstüchtig ist. Dazu lassen sich von der Festplatte diverse Parameter abfragen und sie kann auch dazu veranlasst werden, kürzere Prüfsequenzen durchzuführen. Zusammengefasst werden diese Techniken unter dem Begriff „Self-Monitoring, Analysis, and Reporting Technology used in computer storage devices“ (SMART). Um SMART zu nutzen, existiert unter Linux das Hintergrundprogramm „smartd“.

⁴² <https://wiki.gentoo.org/wiki/LVM>

Die Konfiguration dazu befindet sich in der Datei `/etc/smartd.conf`. Diese wurde für die Bedürfnisse des Archivspeichers wie folgt angepasst:

```
# All ATA/SATA or SCSI/SAS disks. Monitor all attributes, enable
# automatic online data collection, automatic Attribute autosave, and
# start a short self-test every day between 2-3am, and a long self test
# Saturdays between 3-4am.
/dev/sdc -a -o on -S on -s (S/../../../../02|L/../../../../6/03) -m archiveadmins@pd-
project.org
/dev/sdd -a -o on -S on -s (S/../../../../02|L/../../../../6/03) -m archiveadmins@pd-
project.org
/dev/sde -a -o on -S on -s (S/../../../../02|L/../../../../6/03) -m archiveadmins@pd-
project.org
/dev/sdf -a -o on -S on -s (S/../../../../02|L/../../../../6/03) -m archiveadmins@pd-
project.org
/dev/sdg -a -o on -S on -s (S/../../../../02|L/../../../../6/03) -m archiveadmins@pd-
project.org
/dev/sdh -a -o on -S on -s (S/../../../../02|L/../../../../6/03) -m archiveadmins@pd-
project.org
/dev/sdi -a -o on -S on -s (S/../../../../02|L/../../../../6/03) -m archiveadmins@pd-
project.org

# On the operating system SSD the parameter 197 is deactivated because
# it is not supported
/dev/sda -a -o on -S on -s (S/../../../../18|L/../../../../4/01) -I 197 -m archivead-
mins@pdproject.org
```

Es muss sichergestellt werden, dass diese beiden Überwachungsprogramme, die eine essen-
tielle Rolle spielen zur Datensicherung, auch wirklich ausgeführt werden und nicht unbe-
merkt abgestürzt sind. Dazu wurden zwei kleine Shellscripts erstellt, die testen, ob der
Dienst noch läuft und ihn sonst neu starten:

```
check_smartd.sh
#!/bin/sh

export PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin

pgrep -x smartd > /dev/null;
result=$?
#echo "exit code: ${result}"
if ! [ "${result}" -eq "0" ] ;
then
    /etc/init.d/smartd stop;
    if /etc/init.d/smartd start; then
        logger local1.info smartd restarted
    else
        logger local1.err unable to start smartd
    fi
fi

check_mdadm.sh
#!/bin/sh

export PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin

pgrep -x mdadm > /dev/null;
```

```

result=$?
#echo "exit code: ${result}"
if ! [ "${result}" -eq "0" ] ;
then
    /etc/init.d/mdadm stop;
    if /etc/init.d/mdadm start; then
        logger local1.info mdadm restarted
    else
        logger local1.err unable to start mdadm
    fi
fi

```

Diese Scripts werden per Cronjob jede Stunde ausgeführt:

```

# check every hour if the RAID and smart monitoring is still running
30 * * * *    root    /root/check_mdadm.sh
35 * * * *    root    /root/check_smartd.sh

```

4.1.3.8 Umstellung des Archivspeichers

Der Hardwareumbau und die Umstellung auf Gentoo GNU/Linux erfolgten am 4. Januar 2017. Vorgängig wurden dazu zwei Backups der bestehenden Daten angefertigt und die Dienste, die direkt auf den Archivspeicher zugreifen, vorübergehend deaktiviert (Online Radio, NextCloud). Nach erfolgter Umstellung und Einrichtung des Betriebssystems, RAID und LVM2 wurden die Backupdaten auf den neuen Archivspeicher kopiert. Danach wurden Schritt für Schritt wieder alle Dienste des Public Domain Projekts in Betrieb genommen und die Überwachung mit mdadm und smartd eingerichtet. Am 6. Januar konnte das neue System wieder dem produktiven Einsatz übergeben werden.

```

nuess0r@archive ~ $ uname -a
Linux archive.publicdomainproject.org 4.4.26-gentoo #4 SMP Tue Dec 6 05:59:18 C
ET 2016 x86_64 Intel(R) Xeon(R) CPU E5430 @ 2.66GHz GenuineIntel GNU/Linux
nuess0r@archive ~ $ uptime
12:22:28 up 6 days, 22:29, 1 user, load average: 0.00, 0.00, 0.00
nuess0r@archive ~ $ cat /proc/mdstat
Personalities : [linear] [raid0] [raid1] [raid10] [raid6] [raid5] [raid4]
md0 : active raid5 sdi[7](S) sdh[6] sdg[4] sdf[3] sde[2] sdd[1] sdc[0]
      3662215680 blocks super 1.2 level 5, 512k chunk, algorithm 2 [6/6] [UUUUU
U]
      bitmap: 0/6 pages [0KB], 65536KB chunk

unused devices: <none>
nuess0r@archive ~ $ df -h
Filesystem                Size      Used Avail Use% Mounted on
udev                     10M        4.0K   10M   1% /dev
/dev/sda4                 55G        6.0G   46G  12% /
tmpfs                    900M       540K   899M   1% /run
shm                      4.4G         0   4.4G   0% /dev/shm
cgroup_root              10M         0    10M   0% /sys/fs/cgro
up
/dev/mapper/vg0-archive   300G      186G   115G  62% /media/archi
ve
/dev/mapper/vg0-backup    300G      160G   141G  54% /media/backu
p
/dev/mapper/vg0-swissfoundationpublicdomain  50G       2.7G   48G   6% /media/swiss
foundationpublicdomain
/dev/mapper/vg0-swissradio 300G      177G   124G  59% /media/swiss
radio
/dev/sda2                 124M       32M   87M  27% /boot
nuess0r@archive ~ $ █

```

Abbildung 11: Ausgaben zu Kernelversion, RAID Status und Festplattenbelegung nach der Migration

4.1.3.9 NFS

Damit die vorhandenen Systeme ihre Daten im Netzwerk wie gewohnt wiederfinden, wurden die neuen Logical Volumes in den per NFS exportierten Ordner /mnt/publicdomain/ per Bind-Mount eingehängt und entsprechend in der /etc/fstab eingetragen. Verknüpfungen (soft links) können dafür nicht eingesetzt werden, da NFS Verknüpfungen nicht auflöst, sondern sie unverändert an den Client weiterreicht. Damit die per Bind-Mount eingehängten Volumes per NFS erreichbar sind, musste unter /etc/exports zusätzlich die Option crossmnt eingetragen werden.

4.1.3.10 Sperren des Schreibzugriffs auf archivierte Daten

Wie gefordert werden Flac Dateien, die länger als 30 Tage auf dem Archivserver sind, schreibgeschützt.

Gelöst wurde diese Anforderung mittels eines find Aufrufs, der die zu schützenden Dateien sucht und dann die Programme chown und chmod aufruft, um die Zugriffsrechte zu modifizieren. Das Shellscript forbid-writes-to-archived-files.sh enthält diesen find Aufruf. Es wird per cron einmal im Tag ausgeführt. Dazu wurde das Script im Ordner /etc/cron.daily/ gespeichert:

```
#!/bin/sh
## This script removes write permissions of archived files after 30 days
## they are stored on the server and changes the owner to root.
##
## After this time only administrators are allowed to modify or delete such
## files

find /mnt/publicdomain/audio/public/flac/ -type f -mtime +30 -exec chown
root '{}' \; -exec chmod 644 '{}' \;
```

4.1.4 Fernwartung mit IPMI Karte

Nach dem Einbau der IPMI Karte und einem BIOS Update des Mainboards konnte die IP Adresse im BIOS Menu konfiguriert werden. Wie befürchtet wurde, war die IPMI Karte vor dem Verkauf nicht zurückgesetzt worden und das Administrator Passwort war nicht bekannt (Die Karte ist älter und kann nur noch gebraucht gekauft werden). Mit dem Konfigurationsprogramm „IPMICFG“⁴³ vom Kartenhersteller (Super Micro) können neben vielen anderen Dingen auch die Benutzer verwaltet werden. So kann auch ein unbekanntes Passwort überschrieben werden.

⁴³ <ftp://ftp.supermicro.com/utility/IPMICFG/>

Die enthaltene Anleitung für Linux hat exakt so funktioniert wie beschrieben:

If your system has installed OpenIPMI driver, you can enabled Linux IPMI driver:

```
# /etc/init.d/ipmi start  
or  
# modprobe ipmi_msghandler  
# modprobe ipmi_devintf  
# modprobe ipmi_si
```

Then execute IPMICFG-Linux.x86 or IPMICFG-Linux.x86_64

Somit konnte mit `IPMICFG-Linux.x86_64 -user list` die Benutzer angezeigt werden und mit `IPMICFG-Linux.x86_64 -user setpwd` ein neues Passwort gesetzt werden.

Danach war es möglich, die IPMI Karte wie gewohnt per Webbrowser anzusprechen und zu testen. Damit kann z. B. der Archivspeicher gestartet werden oder auf das System zugegriffen werden, auch wenn das Betriebssystem/Netzwerkstack etc. ein Problem hat.

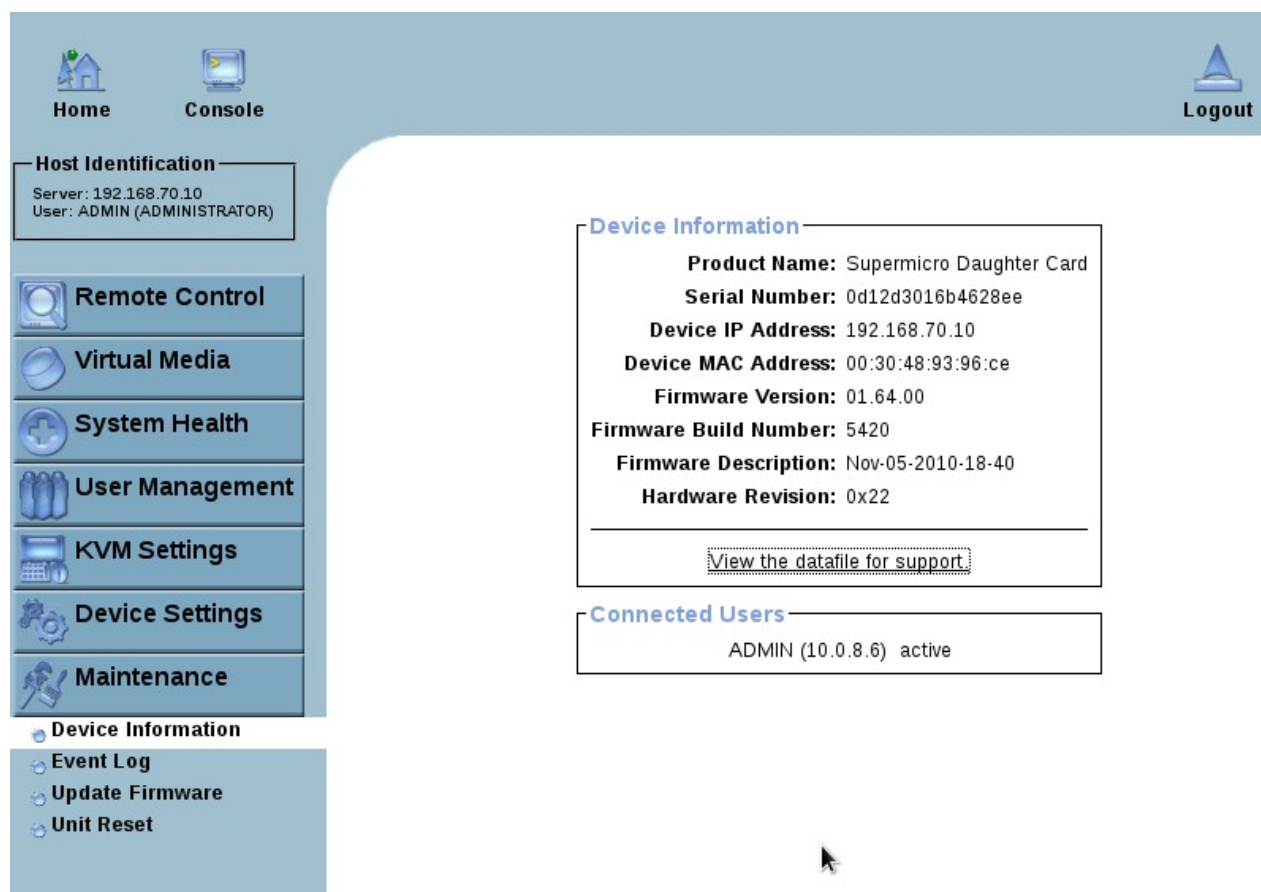


Abbildung 12: Ausgabe der IPMI Management Karte nach der Einrichtung

4.2 Repräsentationsinformation

Eine sehr wichtige Anforderung an einen Archivspeicher ist das Vorhandensein der nötigen Repräsentationsinformationen. Also all die Informationen, die nötig sind, um aus dem Archivspeicher die zu erhaltende Inhaltsinformation auszulesen und zu interpretieren.

Wie beschrieben, ist durch den Einsatz von Gentoo GNU/Linux zu jeder installierten Software auf dem Archivserver der Source Code abgelegt. Diese bleiben auch beim Aufräumen mit dem Tool `eclean-dist` erhalten⁴⁴.

Bevorzugt gegenüber Source Codes werden, wo vorhanden, Standards und formale Definitionen zu den eingesetzten Techniken und Dateiformaten. Diese wurden von Hand zusammengestellt und im Dokumentenverwaltungssystem (siehe Kapitel 6) abgelegt. Derzeit sind diese Informationen erst in digitaler Form abgelegt, da die vorgesehene Schublade mit den essentiellen Repräsentationsinformationen in Papierform nicht mehr während dieser Arbeit zusammengestellt und installiert werden konnte.

Dokumente, die auf dem Archivspeicher abgelegt sind:

- XFS Dokumentation. Quelle: <https://git.kernel.org/cgit/fs/xfs/xfsdocs-xml-dev.git> und http://xfs.org/index.php/XFS_Papers_and_Documentation
- FLAC Dokumentation. Quelle: <https://xiph.org/flac/format.html>
- tar Dokumentation. Quelle: <https://www.gnu.org/software/tar/manual/>
- gz Dokumentation. Quelle: <https://www.gnu.org/software/gzip/manual/gzip.html> <https://www.ietf.org/rfc/rfc1952.txt> <https://www.ietf.org/rfc/rfc1951.txt>
- bz2 Dokumentation. Quellen: <http://bzip.org/docs.html> www.hpl.hp.com/techreports/Compaq-DEC/SRC-RR-124.pdf www.ece.iit.edu/.../Efficient%20Decoding%20of%20Prefix%20Codes.pdf und das Paper *Fast Algorithms for Sorting and Searching Strings* <https://www.cs.princeton.edu/~rs/strings/paper.ps> Das referenzierte Paper `breed3.ps` wurde nicht gefunden!
- MD5 Message-Digest Algorithmus. Quelle: <https://www.ietf.org/rfc/rfc1321.txt>
- PNG Dokumentation. Quelle: <https://www.w3.org/TR/PNG> <http://www.libpng.org/pub/png/pngbook.html> <https://tools.ietf.org/rfc/rfc1950.txt>
- Unicode Standard. Quelle: <http://www.unicode.org/Public/9.0.0/ucd/>
- PDF 1.7 Standard, umgewandelt in Textdatei. Quelle: https://www.adobe.com/devnet/pdf/pdf_reference.html
- Linux RAID (md), Spezifikation auf Bit-Ebene nur für die Metadaten (Superblock) gefunden. Quelle: https://raid.wiki.kernel.org/index.php/RAID_superblock_formats
- LVM2, keine Spezifikation auf Bit-Ebene gefunden, aber Source Code vorhanden.
- MySQL Spezifikation auf Bit-Ebene für das InnoDB Dateiformat fehlt, aber Source Code vorhanden. MariaDB hat diese auch nicht publiziert.

⁴⁴ https://wiki.gentoo.org/wiki/Knowledge_Base:Remove_obsoleted_distfiles

- C Standard. Quelle: <http://www.open-std.org/jtc1/sc22/wg14/www/docs/n1570.pdf> Working Draft 4 zum C11 Standard. C Standard müsste sonst gekauft werden.
- NFS Standard. Quelle: <https://tools.ietf.org/html/rfc1094> <https://tools.ietf.org/html/rfc1813> <https://tools.ietf.org/html/rfc7931>

Die fehlenden Repräsentationsinformation sind:

- SATA Verbindungsstandard fehlt, nicht frei zugänglich (kommerziell zu erwerben)
- ATA Kommandostandard fehlt, nicht frei zugänglich (kommerziell zu erwerben)
- Information wie die Daten auf der Oberfläche der eingesetzten Festplatte organisiert sind. Diese Information besitzen nur der Festplattenhersteller und wenige spezialisierte Datenrettungsfirmen.

4.2.1 Repräsentationsinformation in Papierform

Der zweite Teil zur Umsetzung eines vertrauenswürdigen Archivspeichers ist eine Schublade, die im Serverraum neben dem Gentoo GNU/Linux Server eingebaut wird. Sie enthält in Papierform die essentiellsten Repräsentationsinformationen. Dies beinhaltet alle Informationen die nötig sind, um die digital abgelegten Repräsentationsinformationen aus dem Server auslesen und interpretieren zu können.

Dazu gehören:

- ASCII Tabelle
- Zahlensystembeschreibung
- ATA Kommando- und SATA Schnittstellenstandard zum Ansprechen der Festplatten
- XFS Dateisystemspezifikationen
- Linux RAID und LVM2 Spezifikationen
- Sprachreferenz für die Programmiersprache C und x86 Befehlssatz
- Beschreibung der Ordnerstruktur von Linux und des Public Domain Projekts
- Liste mit den Zugangspasswörtern
- Englischwörterbuch

4.3 Umsetzung des Monitoringsservers

Als Monitoringserver ist der ehemalige Storage controller von Coraid vorgesehen, der durch die Migration auf Gentoo GNU/Linux frei geworden ist.

Hardware im Coraid CLN/21 Storage controller

- Mainboard Super Micro PDSMi+
 - Single Intel Xeon Quad Core/Dual Core 3000 Series/Core 2 Duo/Pentium D (Dual-Core)/Pentium 4 Extreme Edition/Pentium 4/Celeron D LGA (Land Grid Array) 775 Processors at system bus speeds of 1066 MHz/800 MHz/533 MHz.
 - Four DIMM slots support Dual/Single Channel DDR2 667/533/400 MHz up to 8 GB of ECC/Non-ECC Unbuffered DDR2 SDRAM
 - Chipset: Intel 3000, Intel ICH7R, Intel PXH-V
 - One IPMI 2.0 Slot
- Debian GNU/Linux 5 (Lenny) Betriebssystem auf einem parallel ATA Flashmodul
- 2 GiB RAM
- CPU Typ: Intel(R) Pentium(R) D 3.00GHz (Dual-Core Pentium 4)

4.3.1 Hardware aufrüsten

Der vorhandene Arbeitsspeicher von 2 GiB wurde für das Monitoring als ausreichend betrachtet. Das Mainboard besitzt zwar einen IPMI 2.0 Slot, Fernwartung wurde aber nicht nachgerüstet, weil das 1 HE Gehäuse nicht für den Einbau einer IPMI Karte vorgesehen ist. Dies wurde auch als nur kleiner Mangel empfunden, da das Monitoring selber nicht kritisch ist für den Betrieb der Plattform des Public Domain Projekts und entsprechend auch längere Unterbrechungszeiten akzeptiert werden könnten.

Das alte parallel ATA Flashmodul besitzt nur 512 MiB Speicherplatz und wurde wie beim Archivspeicher ersetzt durch eine 60 GiB SATA SSD (Typ Corsair Force LS).

4.3.2 Installation

Für die Installation wurde der zur Verfügung stehende Testrechner benutzt. Nach der Installation von Debian GNU/Linux (Jessie) wurde Sensu und die Weboberfläche Uchiwa gemäss Anleitung⁴⁵ eingerichtet.

Die eigentlichen Tests und die Plugins dazu konnten nicht auf dem Testrechner eingerichtet werden, da dazu die zu testenden Server im gleichen Netzwerk vorhanden sein müssen.

Der Umbau des Storage controllers zum Monitoringserver wurde auch am 4. Januar 2017 im Serverraum der ZHDK durchgeführt. Danach blieb aber keine Zeit mehr um die vorgesehenen Tests einzurichten. Darum ist auch der Screenshot der Weboberfläche eher uninteressant.

⁴⁵ <https://sensuapp.org/docs/latest/platforms/sensu-on-ubuntu-debian.html>

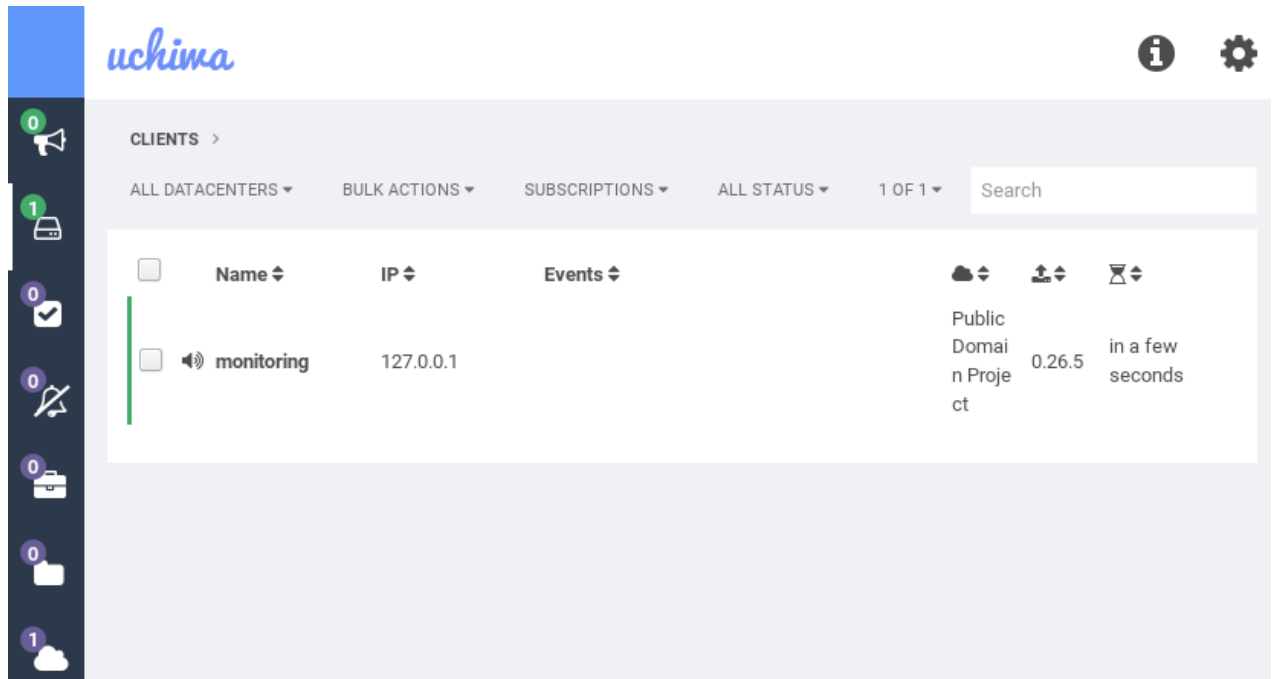


Abbildung 13: Die zu Sensu gehörende Weboberfläche Uchiwa nach der Installation im Serverraum

5 Verfügbarmachen der Metadaten als Linked Open Data

5.1 Semantic Web und Linked Open Data

Für das Verständnis des ganzen Kapitels 5 ist ein rudimentäres Basiswissen zu Semantic Web und RDF hilfreich. Es wurde davon abgesehen, dieses Wissen in diesem Bericht zu vermitteln, da es den Rahmen dieses Berichts sprengen würde und es ausserhalb des Themas dieser Masterthesis liegt. Der Fokus wird auf Fragen der Implementierung gelegt und nicht auf die zu Grunde liegenden Basistechnologien wie XML und RDF.

Einen Überblick vermitteln folgende Webseiten:

- Basics of the Semantic Web <http://strangelove.netlabs.org/semantic-web-basics/>
- RDF 1.1 Primer <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

Weiterführende Literatur:

- Die Masterthesis *Bibliographische Daten und ihre Zugänglichkeit im Web der Daten* von Fabian M. Fürste beschäftigt sich ausführlich mit dem Thema Linked Open Library Data [FUR11]
- An der 132. AES Convention (2012) wurde das ausführliche Tutorial *Semantic Web and Semantic Audio technologies* veranstaltet, das nach einer Einführung in die Grundlagen des Semantic Web auf die möglichen Anwendungsfelder im Audiobereich eingeht [GYÖ12]

5.1.1 Qualitätsniveau von Open Data

Ein aktuelles Bewertungsschema der Open Data Bewegung für öffentlich zugängliche Daten ist das 5-Sterne Modell:

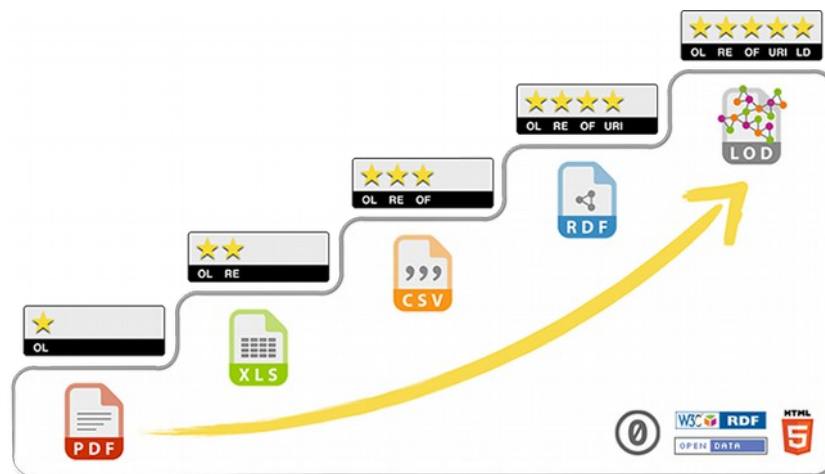


Abbildung 14: [5-Sterne-Modell](http://5stardata.info/de/) für Offene Daten (Open Data), CC0, <http://5stardata.info/de/>

- ★ *stelle deine Daten im Web unter einer offenen Lizenz bereit. Das Format ist dabei egal (OL = Open License)*
- ★★ *stelle Daten in einem strukturierten Format bereit (z. B. Excel anstelle eines eingescannten Bildes einer Tabelle) (RE = Regular Expression)*
- ★★★ *verwende offene, nicht proprietäre Formate (z. B. CSV statt Excel) (OF = Open Format)*
- ★★★★ *verwende URIs, um Dinge zu bezeichnen, damit deine Daten verlinkt werden können (URI = Uniform Resource Identifier)*
- ★★★★★ *verlinke deine Daten mit anderen Daten, um Kontexte herzustellen (LD = Linked Data) Quelle: <http://5stardata.info/de/>*

5.2 Anforderungen der vorgesehenen Zielgruppen

In der vorausgehenden Projektarbeit wurden die Anforderungen der Zielgruppen zusammengetragen und formuliert. Da diese Masterarbeit sich an diesen Anforderungen messen muss, werden sie hier nochmals wiedergegeben:

Die Zielgruppe Allgemeinheit (Global Community) möchte schnell herausfinden was es im Archiv gibt und möchte die Musik auch gleich anhören und herunterladen können. Angeboten werden sollen Formate die aktuell übliche PC und Smartphone Betriebssysteme ohne Zusatzsoftware abspielen können.

Die Zielgruppe Musikwissenschaftler, Historiker und Interpretationsforscher hat zuerst mal die selben Anforderungen wie die Allgemeinheit aber zusätzlich die Anforderung an detaillierte Suchmöglichkeiten. So soll gezielt in einer Auswahl von Genres, Zeitab-

schnitten, Urhebern, Interpreten, geographischen Orten etc. gesucht werden können. Es sollen Listen mit allen verfügbaren Werken von einem Urheber oder Interpreten etc. zugänglich sein. Es muss möglich sein, die Daten aus dem Public Domain Projekt mit Daten aus anderen Quellen zu verknüpfen. Dazu ist es nötig, dass die Daten Identifikatoren aus anerkannten Normdateien zur eindeutigen Identifizierung enthalten. Es soll möglich sein, auf die vollständigen AIPs zugreifen zu können.

Die Zielgruppe Suchmaschinen, Metaarchive, Datenanalyseprogramme (Bots) , besteht nicht aus Menschen sondern aus Maschinen und hat entsprechend andere Anforderungen an die Durchsuchbarkeit und Darstellung der Daten. Zu dieser Zielgruppe gehören auch die Abfragesysteme der mit dem Public Domain Projekt verbundenen Archive. Es muss eine entsprechende Softwareschnittstelle (Application programming interface, API) für einfache Suchanfragen bestehen mit dessen Antworten dann Audiowerke heruntergeladen oder die Erhaltungsmetadaten abgerufen werden können. Des weiteren besteht der Wunsch nach einer Möglichkeit komplexe Anfragen zu stellen (Z. B. Liste aller Werke von 1730 bis 1750 von Schülern von Johann Sebastian Bach bei der eine Orgel unter den Instrumenten ist) und dessen Antworten Referenzen auf maschineninterpretierbare Konzepte und Normdaten beinhaltet (Semantic Web, Linked Data). Der aktuelle Stand der Technik für solche Anfragen ist SPARQL⁴⁶, dabei werden die Daten im Format RDF 1.1, serialisiert als RDF/XML, zurückgeliefert. [CHZ16], Seite 40

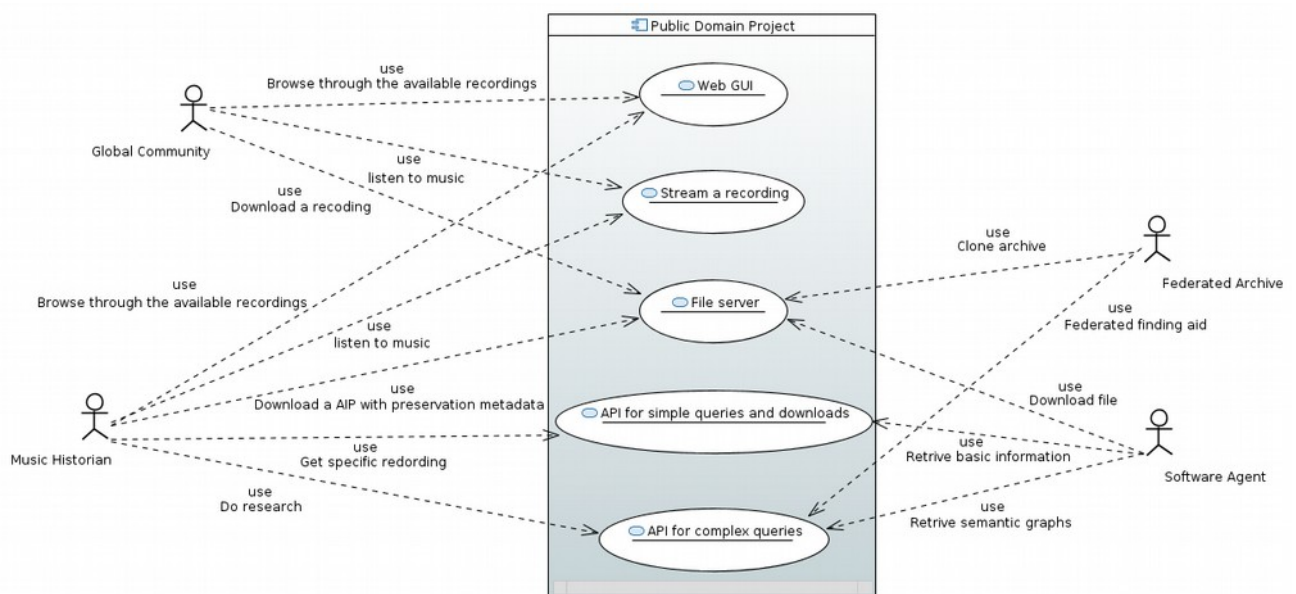


Abbildung 15: UseCase Diagramm der vorgesehenen Zielgruppen, Quelle: [CHZ16], Seite 40

⁴⁶ <https://www.w3.org/TR/sparql11-query/>

5.3 Aktueller Stand der Metadaten im Public Domain Projekt

Das Public Domain Projekt erfasst detaillierte Metadaten zu den Tonträgern und den Liedern darauf. Derzeit sind die Metadaten im Public Domain Projekt für jedes einzelne Werk auf einer eigenen Wikiseite als Tabelle abgelegt (Beispiel im Anhang 14.4.1). Diese Wikiseiten lassen sich zwar per MediaWiki API abrufen, aber müssen danach aufwändig weiterverarbeitet werden, da die Daten zwar als XML ausgeliefert werden aber darin die Syntax von MediaWiki verwendet wird und nicht in einem verbreiteten Format zur Weiterverarbeitung. Ein Beispiel im XML Format ist im Anhang 14.4.2 zu finden.

Entsprechend bewegt sich das Projekt derzeit auf dem Niveau ★★ Open Data.

5.3.1 Semantic MediaWiki

Auf dem Webserver des Public Domain Projektes ist neben MediaWiki auch seit mehreren Jahren die umfangreiche Erweiterung Semantic MediaWiki⁴⁷ installiert. Diese wurde bisher aber aus Mangel an Zeit, Erfahrung bzw. interessierten Freiwilligen nicht genutzt.

Semantic MediaWiki erlaubt es, in einem Wiki enthaltene Daten maschinenlesbar zu machen und diese sowohl intern (auf Wikiseiten) wie extern (als CSV, JSON, RDF/XML oder mit einem SPARQL Endpoint) abzufragen.

Im HTML Header jeder Wikiseite ist ein Link zu den RDF/XML Daten eingebettet: `<link rel="ExportRDF" type="application/rdf+xml" title="Busy-bee-d47-3259" href="/index.php?title=Special:ExportRDF/Busy-bee-d47-3259&xmlmime=rdf"/>` Daher ist jetzt schon für jede Suchmaschine bzw. jeden Open Data Interessierten erkennbar, dass das Public Domain Projekt Daten als RDF/XML zur Verfügung stellen könnte.

Da derzeit aber keine Funktionen von Semantic MediaWiki genutzt werden, sind fast keine Daten im Export. Diese sind so gut wie ausdruckslos und basieren nur auf der Ontologie von Semantic MediaWiki. Ein Beispiel dazu findet sich im Anhang 14.4.3.

⁴⁷ <https://www.semantic-mediawiki.org/>

5.4 Evaluation der einzusetzenden Standards für die Metadaten

Es wurde viel Zeit in die Suche und den Vergleich von Metadatenstandards investiert, um herauszufinden, was andere Archive, Bibliotheken oder andere Branchen zur Langzeiterhaltung oder zur Veröffentlichung von Linked Open Data verwenden.

Es hat sich gezeigt, dass derzeit sehr viele solcher Standards und Ontologien existieren [FUR11], [CHZ16]. Diese sind dann vielfach entweder sehr eingeschränkt, nicht maschineninterpretierbar oder spezifisch für einen Fachbereich. Viele der gefundenen Ontologien sind regional oder auf bestimmte Sprachen/Fachbereiche limitiert. Im Bereich der Archive der Audiovisuellen Medien sind auch verschiedene Standards in Verwendung, ohne dass sich eine Konsolidierung abzeichnet.

Es hat sich gezeigt, dass ein einzelner Metadatenstandard nicht ausreichend ist, um alle Erhaltungsmetadaten zu erfassen, sondern mehrere Standards kombiniert werden müssen, um die Anforderungen zu erfüllen.

Wie im Verlauf dieser Arbeit gelernt wurde, ist dies kein Mangel, sondern die aktuell favorisierte Vorgehensweise. Für Eigenschaften, die oft vorkommen bzw. in vielen Modellen die selbe Aussage besitzen, ist es sinnvoll die selbe Ontologie zu verwenden. So können die semantischen Daten aus verschiedenen Bereichen direkt miteinander verknüpft werden.

Entsprechend sind modernere Ontologien meistens klein, da sie nur die spezifischen Eigenschaften eines Anwendungsfeldes abdecken müssen, die mit bestehenden Ontologien nicht ausgedrückt werden können. Beispiele dazu sind SemWeb Vocab Status ontology (vs) mit gerade mal drei Eigenschaften, um die Version einer Ontologie oder Bestandteilen davon anzugeben, die Timeline Ontologie, um Zeitbereiche zu beschreiben oder Geo, um geographische Koordinaten auszuzeichnen.

5.4.1 Grundlegende Ontologien

In den letzten Jahren hat sich eine kleine Gruppe von grundlegenden Ontologien herausgebildet, die oft verwendet werden und auf denen die Spezialisierungen aufbauen.

Dazu gehören (nicht abschliessend):

- Dublin Core (siehe nächste Seite)
- Schema.org (siehe nächste Seite)
- foaf, Friend of a friend, Ontologie um Personenbeziehungen zu modellieren
- Skos, Simple Knowledge Organization System

5.4.1.1 Dublin Core (Dc, Dcterms)

Das Kernelement Set Dublin Core Metadata Element Set (Dc) ist 1995 aus dem Bedürfnis entstanden, Katalogbestände über verschiedene Institutionen auszutauschen, was nicht einfach ist, weil jede Institution anderen Metadatenvorgaben folgt. Dc ist vor dem Semantic Web und RDF entstanden. Das Core im Namen bezieht sich darauf, dass es sich hier um ein

Set von 15 Basisattributen handelt, die einem minimalen Konsens entsprechen, welche Begleitinformation bei jedem Werk vorhanden sind. Es bildet daher die Basis für heutige Austauschformate. Entsprechend ist klar, dass Dublin Core nicht gross genug ist, um alle Erhaltungsmetadaten aufzunehmen. Mit Dc können Herkunftsinformationen (Provenance Information) und ein Teil der Kontextinformation (Context Information) abgedeckt werden. Es besteht zudem das Problem, dass zwar die 15 Felder definiert sind, aber nicht strikt das Format oder Vokabular des Feldinhalts.

Seit 2012 existiert der erweiterte Standard „DCMI Metadata Terms“⁴⁸ (Dcterms), der die ursprünglichen 15 und ein Set von erweiterten Attributen umfasst. Zusätzlich kamen auch Klassen und Vokabular Definitionen dazu. Dieser Standard enthält auch Anpassungen damit Dc den Semantic Web Standards genügt.

Durch sein Alter und seine Herkunft aus dem Bibliothekswesen ist Dc sehr verbreitet und wird von mehreren Metabibliotheken⁴⁹ als Austauschformat vorgeschrieben.

5.4.1.2 Schema.org (Schema)

Die Ontologie von Schema.org ist das Resultat einer Zusammenarbeit von grossen Suchmaschinenbetreibern und wurde 2011 veröffentlicht. Sie wird seither in einem für alle offenstehenden Prozess weiter entwickelt.

Schema.org is a set of constructs that allow website designers to include structured metadata in their Web pages, to be consumed by the major search engines Bing, Google, and Yahoo! Schema.org is designed to represent resources from a great diversity of domains. It thus duplicates many elements from other element sets, and fails to capture the richness of library data. Quelle:

https://www.w3.org/2005/Incubator/ld/wiki/Vocabulary_and_Dataset

Wie dem Zitat zu entnehmen ist, gibt es eine grössere Überschneidung von Properties zwischen Dc und Schema. Bei den Klassen ist die Überschneidung geringer, da sich Dc eher auf Bibliotheksobjekte (Bücher, Papier, Regelwerk etc.) und Schema eher auf Multimediainhalte auf Webseiten bezieht.

⁴⁸ DCMI steht für „Dublin Core Metadata Initiative“

⁴⁹ Ein Portal mit dem in den Beständen von vielen angeschlossenen Bibliotheken/Archiven gesucht werden kann.

5.4.2 Spezialisierte Ontologien relevant für das Public Domain Projekt

5.4.2.1 Creative Commons Rights Expression Language (CC)

Die Creative Commons Rights Expression Language (CC)⁵⁰ ist ein Vokabular, um Lizenzen und Nutzungsbedingungen zu beschreiben. Es basiert auf dem Creative Commons Lizenzmodell und ist daher weit verbreitet. Es wird oft genutzt zusammen mit Dcterms, Schema.org, edm, WikiData oder ist in den Empfehlungen aufgeführt.

5.4.2.2 Music Ontology (mo)

The Music Ontology provides a vocabulary for publishing and linking a wide range of music-related data on the Web. Music Ontology data can be published by anyone as part of a web site or an API and linked with existing data, therefore creating a music-related web of data. Quelle: <http://musicontology.com/docs/faq.html>

Die Music Ontology⁵¹ ist für den Bereich, in dem sich das Public Domain Projekt bewegt, die relevanteste spezialisierte Ontologie und wird auch von MusicBrainz verwendet. Sie definiert Klassen und Properties, die den gesamten Ablauf beschreiben von einer musikalischen Idee (Komposition) über die Aufführung und deren Aufnahme bis zu einem veröffentlichten Lied mit dazugehöriger Audiodatei. Weitere Details können dem Paper [RAI07] oder dem Tutorial [GYÖ12] entnommen werden.

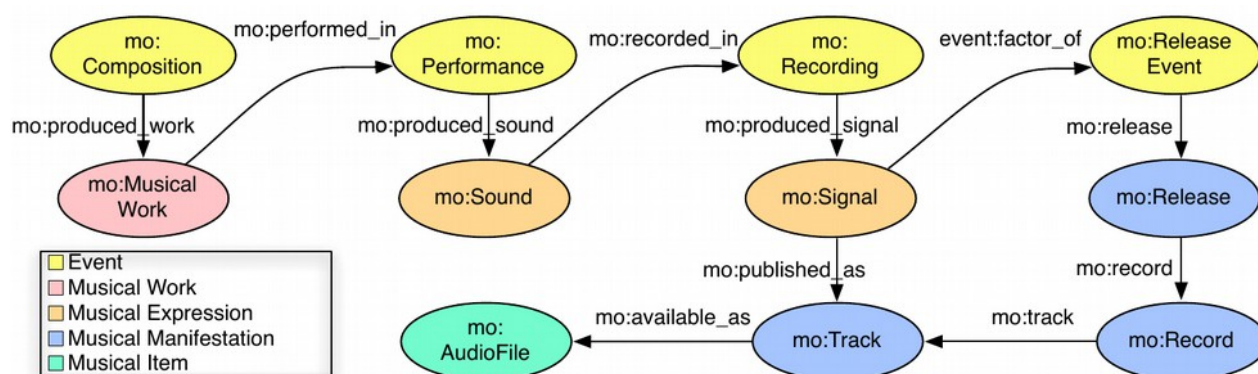


Abbildung 16: Klassen und deren Beziehungen die von der Music Ontology definiert werden. Quelle: [GYÖ12], Seite 98

5.4.2.3 Studio Ontology

Die Studio Ontology⁵² gehört zur Entwicklung der Music Ontology. Sie hat zum Ziel die Produktion von Musik mit den verwendeten Geräten (Mikrofonen, Mischpulten, Effektgeräten etc.) und deren Einstellungen zu verknüpfen. Diese Ontologie wäre für das Public Domain Projekt geeignet gewesen, die bei der Digitalisierung verwendeten Geräte zu dokumentieren.

⁵⁰ <https://creativecommons.org/ns>

⁵¹ <http://musicontology.com/specification>

⁵² <http://isophonics.net/content/studio-ontology>

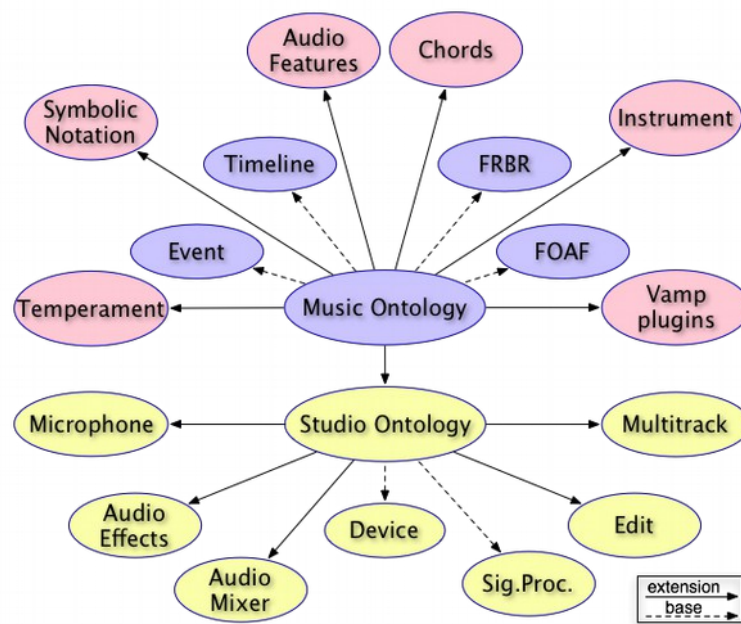


Abbildung 17: Beziehungen zwischen der Music- und Studioontologie
 Quelle: [GYÖ12], Seite 101

Leider konnten keine Details zu dieser Ontologie gefunden werden. Es wurde zwar ein Link⁵³ für eine Spezifikation auf Purl.org⁵⁴ gefunden, dieser zeigt aber leider ins Leere. Weitere Recherchen blieben erfolglos.

5.4.2.4 Logistics Core (LogiCO)

Eine weitere spezialisierte Ontologie ist notwendig um Daten zu erfassen, die der Lagerverwaltung des physischen Archivs der Schweizerischen Stiftung Public Domain dienen. Dies umfasst alle Angaben, die nötig sind, um einen Tonträger innerhalb des Archivs wieder zu finden. Dazu gehören Angaben zu Behältern, deren Abmessungen, Gewicht, Standorten und Lagerplätzen.

Diesen Bereich (und noch viel mehr) deckt die Logistics Core Ontology⁵⁵ ab. Recherchen zur Verbreitung dieser Ontologie oder zu gebräuchlicheren Alternativen blieben erfolglos. Es blieb unklar, ob die Logistik sich noch nicht mit Semantic Web beschäftigt oder ob diese Informationen nur in Fachpublikationen zu finden wären.

Für die Schweizerische Stiftung Public Domain spielt die Verbreitung von LogiCO keine Rolle, da die Daten zur Lagerverwaltung nur einen kleinen Wiederverwendungswert ausserhalb des Public Domain Projekts haben.

⁵³ <http://motools.sourceforge.net/studio/>

⁵⁴ <https://archive.org/services/purl/purl/ontology/studio/>

⁵⁵ <http://ontology.tno.nl/logico/>

5.4.3 Verwendete Ontologien bei verwandten Projekten

Um aus der grossen Anzahl Ontologien die passenden für das Public Domain Projekt auszuwählen, die auch von den Zielgruppen verstanden werden, wurden verschiedene Projekte untersucht, die Ähnlichkeit mit dem Public Domain Projekt aufweisen oder von den Nutzern her vergleichbar sind. Untersucht wurde MusicBrainz⁵⁶, Dbpedia⁵⁷, WikiData⁵⁸, europeana⁵⁹ und Memobase⁶⁰ (Dokumentation per E-Mail erhalten). Die Listen dazu sind im Anhang 14.2 zu finden.

Es hat sich gezeigt, dass keine Ontologie in allen Projekten genutzt wird. Wie erwartet, wird von fast allen Dcterms verwendet, ausser WikiData, welches als einziges Projekt Schema einsetzt. Von vielen wird Geo, Foaf und Skos verwendet. MusicBrainz setzt Mo ein. Memobase dagegen setzt die Ontology for Media Resources (ma-ont) ein. Diese wird nur von EBUcore referenziert und ist ausserhalb von Rundfunkanstalten kaum verbreitet, aber Schema hat daraus seine Properties übernommen. Europeana setzt das als MARC Nachfolger konzipierte RDA⁶¹ ein.

Alle untersuchten Projekte definieren auch projektspezifische Ontologien, WikiData teilt diese z. B. in 13 kleinere WikiData Ontologien auf. Die Dbpedia Ontologie ist sehr umfangreich. Falls dies möglich ist, werden die Dbpedia Properties durch owl:isNarrowerOf/owl:isBroaderOf, owl:sameAs oder owl:similar mit anderen Ontologien in Beziehung gesetzt. Da Dbpedia eine der am häufigsten abgefragten Semantic Web Datenquelle ist, ist deren Ontologie fast schon zu den grundlegenden Ontologien aus Kapitel 5.4.1 dazu zu zählen.

5.4.4 Nicht weiterverfolgte Metadatenstandards

Im Verlauf der vorausgegangenen Projektarbeit [CHZ16] wurden unter anderem folgende Standards evaluiert und danach nicht mehr weiterverfolgt:

- Library of Congress Subject Headings (LCSH), nicht maschinenlesbar
- Mets, [...] *providing an encoding format for descriptive, administrative, and structural metadata for textual and image-based works.*
<http://www.loc.gov/standards/mets/METSOOverview.v2.html>
- MARC21, MACHine Readable Cataloguing, nicht allgemein menschenlesbar, da die Kategorisierung nur aus Nummern zusammengesetzt wird
- foaf, ist in PREMIS inkludiert
- Pbcore, kleine Verbreitung, nicht kompatibel mit RDF/LinkedData
- bbcore, Verwendung innerhalb der British Broadcast Corporation (BBC)
- EBUCore, verwendet von Memoria, mehr orientiert an Produktionsabläufen von Radio und Fernsehstudios als Langzeiterhaltung, darum zu überladen für das Public Domain Projekt

⁵⁶ <https://github.com/LinkedBrainz/MusicBrainz-R2RML>

⁵⁷ <http://dbpedia.org/ontology/>

⁵⁸ <https://www.wikidata.org>

⁵⁹ <http://pro.europeana.eu/share-your-data/data-guidelines/edm-documentation>

⁶⁰ <http://memobase.ch>

⁶¹ In [FUR11] sind Details zu MARC (Seite 18) und RDA (Seite 98) nach zu lesen.

- SKOS, ist in PREMIS inkludiert
- Functional Requirements for Bibliographic Records (FRBR), fokussiert auf Bücher in Bibliothekskatalogen, foaf ist zu einem Teil äquivalent mit den FRBR Metadaten
- ISAD(G), [...] *does not itself provide a machine-readable binding* <http://www.u-koln.ac.uk/metadata/dcmi/collection-provenance/>, fokussiert auf Aktenarchive
- Data Catalog Vocabulary (DCAT), Austauschformat für Datenverzeichnisse, nicht passend
- XMP ist nur für Bild- und PDF-Dateien geeignet
- IPTC ist nur für Bilddateien geeignet

5.5 Zuweisung von Strings anstelle von URIs

Wie der Dokumentation von Schema.org zu entnehmen ist (Dublin Core verweist auch auf diesen Abschnitt), ist es legitim, Properties auch einen String zu zu weisen, wenn eigentlich eine URI erwartet wird:

We also expect that often, where we expect a property value of type Person, Place, Organization or some other subclassOf Thing, we will get a text string, even if our schemas don't formally document that expectation. In the spirit of "some data is better than none", search engines will often accept this markup and do the best we can. Similarly, some types such as [Role](#) and [URL](#) can be used with all properties, and we encourage this kind of experimentation amongst data consumers. Quelle: <https://schema.org/docs/datamodel.html>

Ein String anstatt einer URI (und der Information, die von da abgerufen werden kann) bietet natürlich eine kleinere Informationsmenge und keine Eindeutigkeit, ist also eine starke Einschränkung der Aussagekraft. Auf der anderen Seite erlaubt diese Lockerung ein schrittweises Einführen semantischer Technologien. In einem ersten Schritt werden vorhandene Daten, die meistens nur als Strings vorliegen, so wie sie sind den Properties zugewiesen. In weiteren Schritten können dann diese Daten so aufbereitet werden, dass nach und nach den Properties URIs zugewiesen werden können. Der Vorteil liegt darin, dass schon im ersten Schritt weit verbreitete Ontologien verwendet werden können und die semantischen Daten von Anfang an, in einer reduzierten Form, weiterverwendet und verknüpft werden können.

5.6 Implementation eines Prototyps auf Basis von LogiCO

Ab diesem Kapitel gilt für die Darstellung:

- Properties (Attribute) sind monospace. Zum Beispiel: `Dcterms:hasPart`
- Templates (MediaWiki Vorlagen) sind kursiv. Zum Beispiel: *Audio track*

Um den Umgang mit Semantic MediaWiki und der für Formulare zuständigen Extension Page Forms⁶² zu lernen, wurde entschieden, den Bereich Archivlogistik vor den Teilen die die Tonträger betreffen, umzusetzen. Der Bereich Archivlogistik eignete sich dafür gut, da bisher ein Tonträger lediglich einer Kategorie hinzugefügt wurde, deren Name der ID einer Lagerkiste entspricht. Es musste also nicht auf bisherige Arbeitsabläufe oder bestehende Templates Rücksicht genommen werden. Wichtige Hinweise und Gedankenanstöße enthält das Paper „Semantic MediaWiki with Property Clusters“⁶³

Ausgewählte Quellcodes zu diesen Prototypen sind als Beispiele für die Syntax von MediaWiki, Semantic MediaWiki und semantischen Abfragen im Anhang 14.3 wiedergegeben.

Legende der spezialisierten Ontologien:

- LogiCO: Logistics Core Ontology
- PD: Properties die spezifisch für das Public Domain Projekt definiert wurden

5.6.1 Storage box Template

Zentral für die Logistik ist der Lagerbehälter, in dem ein Tonträger gelagert wird. Dieser besitzt verschiedene Eigenschaften, die mit LogiCO modelliert werden:

Ein Lagerbehälter ist je nach Typ ein Objekt der Klasse `LogiCO:Carton` oder `LogiCO:Box`. Dies sind Unterklassen von `LogiCO:MoveableResource`. Die Zuweisung eines Objekts zu einer Klasse erfolgt in Semantic MediaWiki durch das Zuweisen einer Kategorie. Eine solche Kategorie ist standardmässig eine Klasse der projektspezifischen Ontologie. Um beispielsweise die Kategorie `Carton` als Klasse `LogiCO:Carton` zu deklarieren wird das spezielle Property `[[imported from::LogiCO:Carton]]` auf der Wikiseite der Kategorie benötigt.

Ein Tonträger, der in diesem Lagerbehälter untergebracht ist, muss das Property `LogiCO:isPackagedIn` (mit dem Datentyp `Page`) auf den Seitennamen des Lagerbehälters setzen. Z. B.: `[[LogiCO:isPackagedIn::EURO-SD25-00008 demo]]`

Damit ist der Tonträger mit dem Lagerbehälter verknüpft und der Inhalt eines Lagerbehälters kann mit einer einfachen semantischen Abfrage aufgelistet werden:

```
{#{ask: [[LogiCO:isPackagedIn::{{FULLPAGENAME}}]]
| link=subject
| format=ul
| order=asc
| default=This box is empty or has unknown content
}}
```

⁶² https://www.mediawiki.org/wiki/Extension:Page_Forms

⁶³ http://semeb.com/dpldemo/index.php?title=Semantic_MediaWiki_with_Property_Clusters_%28print_version%29

Metadatenfeld	LogiCO Property	Datentyp
Box ID	LogiCO:id	Text (Wird automatisch auf den Seitennamen gesetzt)
Box type	PD:IsStorageBoxType	Text
Box size	LogiCO:hasLengthValue, LogiCO:hasWidthValue, LogiCO:hasHeightValue	Benutzerspezifischer Typ: Length
Box weight	LogiCO:hasWeightValue	Benutzerspezifischer Typ: Weight
Archive	LogiCO:hasLocation	Page (Link zu Wikiseite, wird im RDF als URL exportiert)
Shelve	PD:InShelve	Text
Level	PD:OnLevel	Number
Is it full?	LogiCO:isFull	Boolean

Tabelle 3: Zuweisung der Felder des *Storage box* Template zu den Ontologien

Die Auflistung des Inhalts übernimmt auch das Template *Storage box*. Das folgende Bild zeigt die menschenlesbare Darstellung der Lagerbehälterinformation:

Storage box location

Box ID	EURO-SD25-00008 demo
Box type	Carton 40cm
Box size	40 cm x 40 cm x 40 cm (Length x Width x Height)
Box weight	35 kg
Archive	Archive Rüti
Shelve	5
Level	2
Is it full?	Yes

Records stored in this box

- [Polydor 67575 demo](#)
- [Polydor 67575-67577 demo](#)

Abbildung 18: Ausgabe des Templates *Storage box*

Semantic MediaWiki bietet im Menü die Funktion *Browse Properties* um die von einer Wikiseite gesetzten Properties und deren Werte in einer menschenlesbaren Form anzeigen zu lassen. Dabei werden auch gleich die Wikiseiten (Subjekt) angezeigt, die Properties besitzen, die die angezeigte Seite als Ziel (Objekt) gesetzt haben, ohne dass eine Query nötig ist.

Der dazugehörige RDF/XML Export ist in Anhang 14.4.5 zu finden.

EURO-SD25-00008 demo	
InShelve	5 + 🔍
IsStorageBoxType	Carton 40cm + 🔍
LogiCO:hasHeightValue	0.4 m (4.0e-4 km, 4 dm, 40 cm, 400 mm) + 🔍
LogiCO:hasLengthValue	0.4 m (4.0e-4 km, 4 dm, 40 cm, 400 mm) + 🔍
LogiCO:hasLocation	Archive Rüti + 📍
LogiCO:hasWeightValue	35 kg (35,000 g, 0.035 T) + 🔍
LogiCO:hasWidthValue	0.4 m (4.0e-4 km, 4 dm, 40 cm, 400 mm) + 🔍
LogiCO:id	EURO-SD25-00008 demo + 🔍
LogiCO:isFull	true + 🔍
OnLevel	2 + 🔍
Has query	EURO-SD25-00008 demo + 📍
Categories	Carton
Modification date	17:28:45, 6 January 2017 + 🔍
hide properties that link here	
Polydor 67575 demo + 📍 , Polydor 67575-67577 demo + 📍 LogiCO:isPackagedIn	

Abbildung 19: Properties und deren Werte von einem Lagerbehälter (Beispiel)

5.6.2 Storage box Formular

Nachdem das Template erstellt war, konnte das dazugehörige Formular entwickelt werden, das zur Eingabe und zum Ändern eines Lagerbehälters dient.

Es wurde darauf geachtet, dass eine Eingabe wenn möglich auf die erlaubten Werte eingeschränkt wird. Bei Properties, die nur eine kurze vorgegebene Liste von Werten annehmen können, ist dies besonders einfach mit dem Eingabetyp Radiobutton zu erreichen. Diesem Eingabetyp wird der Name des Properties übergeben, das als Werteinschränkung dienen soll. Mit dieser Methode wird in diesem Template der Standort und der Typ des Lagerbehälters ausgewählt.

Das Eingabefeld Level wird durch einen regulären Ausdruck auf den Wertebereich 1 bis 9 oder leer eingeschränkt. Ein Namensschema für die Regale existiert noch nicht, darum ist das Feld Shelve ein unbeschränktes Textfeld.

Box type: RAKO EURO Carton 40cm Open shelf 1m
Archive: Archive Rüti Archive Küssnacht Office Zürich
Shelve:
Level:
IsFull: Yes No

Free text:

Summary:

This is a minor edit Watch this page

[Cancel](#)

Abbildung 20: Eingabemaske des Formulars Storage box

Damit auf der Wikiseite eines Lagerbehälters der Knopf „Edit with form“ erscheint, muss in den dazu passenden Kategorien das Formular mit der Parserfunktion `{{#default_form:<Formularname>}}` angegeben werden. Für die Lagerbehälter wurden entsprechend die Kategorien Box, Carton, und Package erweitert um die Zeile: `{{#default_form:Storage box}}`

5.7 Implementation der ausgewählten Ontologien für bestehende Daten

Nachdem im vorausgehenden Kapitel eine Auswahl geeigneter Ontologien erstellt wurde, wurden die bestehenden Templates, die für die Erfassung von Tonträgern und Liedern verwendet werden, um Properties erweitert. Damit wird erreicht, dass die bisher erfassten Daten maschinenlesbar sind und für semantische Abfragen (Queries) genutzt werden können.

Jedem Datenfeld soll mindestens ein Property zugewiesen werden, auch wenn in den evaluierten Ontologien kein passendes enthalten ist. In diesem Fall muss ein projektspezifisches Property definiert werden. So wird erreicht, dass alle Daten maschinenlesbar sind. Zudem wird erreicht, dass sie gleich behandelt werden können, wie Daten in einer Datenbank. Dies erlaubt es dem Public Domain Projekt, in Zukunft serverseitige Hilfsprogramme (Bots) mit weniger Aufwand als bisher zu implementieren. Die Daten lassen sich so auch leichter in andere Datenmodelle migrieren.

5.7.1 Mapping zwischen Metadatenfeldern und Ontologien

In den folgenden Tabellen sind von einem Template jeweils die vorhandenen Datenfelder aufgeführt und die dazu passenden Properties aus den ausgewählten Ontologien. Es wurde jeweils versucht, das genauest mögliche Property zu finden. So unterscheidet Dbpedia z. B. zwischen einem Komponisten (Dbo:composer) und einem Liedtexter (Dbo:lyrics). In Dublin Core oder Schema sind beides Urheber (Creator).

Die projektspezifischen Properties sind auch in diesen Tabellen definiert.

Legende der spezialisierten Ontologien:

- Dbo: Ontologie von Dbpedia
- FOAF: Friend of a Friend
- CC: Creative Commons Rights Expression Language
- LogiCO: Logistics Core Ontology
- PD: Properties die spezifisch für das Public Domain Projekt definiert wurden

5.7.1.1 Audio file Template

Metadatenfeld	DC bzw. Dcterms	Schema.org	Spezialisierte Ontologie
Image(s)		Image ThumbnailUrl	FOAF:Depiction
Label	Publisher	Publisher	
Cat. no.		CatalogNumber	
Order number			PD:OrderNumber
Matrix/StamperID			PD:StamperID
1st release date	Issued	DatePublished	
1st recording date		DateCreated	Dbp:recordingDate
Coupling date			PD:CouplingDate
Cutout date			PD:CutoutDate
Place of recording		LocationCreated	
Description	Description	Description	
Author(s)/Composer(s)	Creator	Creator	Dbp:composer
Lyricist(s)	Creator	Creator	Dbp:lyrics
Music arranger(s)	Contributor	Editor	
Conductor(s)	Contributor	Contributor	PD:Conductor
Performer(s)	Contributor	Contributor	
Vocal range			PD:VocalRange
Title/Work	Title	Name	
Content	Description	Description	
Genre(s)		Genre	
FLAC	hasFormat	Audio FileFormat	PD:FlacFileLink
Ogg (Vorbis/Theora)	hasFormat	Audio FileFormat	PD:OggFileLink
PD CH			PD:Pdch
PD EU			PD:Pdeu
PD USA			PD:Pdusa
PD INT			PD:Pdint
Other version	hasVersion		

Tabelle 4: Zuweisung der Felder des Audio file Template zu den Ontologien

5.7.1.2 Audio storage description Template

Metadatenfeld	DC bzw. Dcterms	Schema.org	Spezialisierte Ontologie
Image(s)		Image, ThumbnailUrl	FOAF:Depiction
Medium	Format		
Label	Publisher	Publisher	
Cat. no.		CatalogNumber	
Order number			PD:OrderNumber
Matrix/StamperID			PD:StamperID
1st release date	Issued	DatePublished	
1st recording date		DateCreated	
Coupling date			PD:CouplingDate
Cutout date			PD:CutoutDate
Place of recording		LocationCreated	
References and notes	References	Citation	
Content of the record	Description	Description	
Description/Notes	Description	Description	PD:Notes
Condition			PD:MediumConditionReport

Tabelle 5: Zuweisung der Felder des Audio storage description Template zu den Ontologien

5.7.1.3 Phonograph cylinder description Template

Metadatenfeld	DC bzw. Dcterms	Schema.org	Spezialisierte Ontologie
Image(s)		Image, ThumbnailUrl	FOAF:Depiction
Medium	Format		
Label	Publisher	Publisher	
Cylinder no.		CatalogueNumber	
1st release date	Issued	DatePublished	
1st recording date		DateCreated	
Cutout date			PD:CutoutDate
Place of recording		LocationCreated	
References	References	Citation	
Content of the record	Description	Description	
Notes			PD:Notes
Condition			PD:MediumConditionReport

Tabelle 6: Zuweisung der Felder des Phonograph cylinder description Template zu den Ontologien

5.7.1.4 Storage media location report Template

Metadatenfeld	DC bzw. Dcterms	Schema.org	Spezialisierte Ontologie
Image(s)		Image, ThumbnailUrl	FOAF:Depiction
Medium	Format		
Label	Publisher	Publisher	
Cat. no.		CatalogueNumber	
Order number			PD:OrderNumber
Matrix/StamperID			PD:StamperID
Collection		Funder	
Cleaning process			PD:AppliedCleaningProcess
Digitizing process			PD:AppliedDigitizingProcess
Location			LogiCO:Location
Archive			Wird derzeit nicht verwendet
Shelve			PD:InShelve
Level			PD:OnLevel
Container			LogiCO:IsPackagedIn

Tabelle 7: Zuweisung der Felder des *Storage media location report* Template zu den Ontologien

5.7.1.5 License tags

Eine ganze Gruppe an Templates sind die sogenannten *License tags*. Diese werden auf Wiki-seiten von Liedern, Bildern etc. eingesetzt, um deren Urheberrechtstatus wiederzugeben.

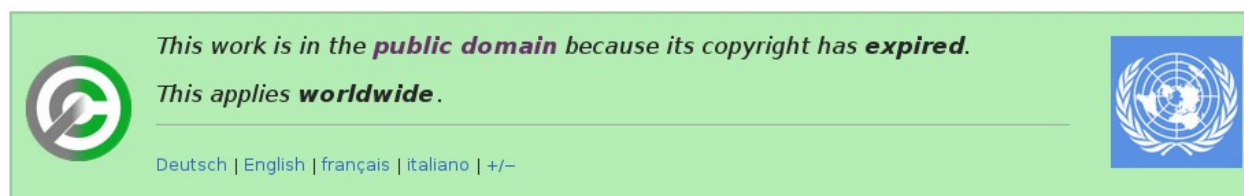


Abbildung 21: Der License tag *PD-INT* beschreibt, dass ein Werk weltweit gemeinfrei ist

Die License tags besitzen keine Eingabefelder sondern es wird auf einer Wikiseite der passende License tag eingefügt. Entsprechend gibt es für die verschiedenen Status bzw. für verschiedene Regionen separate License tags, welche auch um Properties erweitert werden, da der rechtliche Status eine sehr wichtige Information für alle Zielgruppen ist.

Dafür wird die Creative Commons Rights Expression Language (CC) eingesetzt. Sie definiert nicht nur die Properties, sondern auch die zugelassenen Werte. Jedes License tag setzt die Properties fix auf die korrekten Werte.

Das Hilfstemplate *Smw license* erleichtert diese Aufgabe.

Metadatenfeld	DC bzw. Dc-terms	Creative Commons	Wertbereich
License	License		URL zu Legalcode bei Creative Commons
Permissions		Cc:permits	Reproduction, Distribution, Derivative Works, Sharing
Prohibitions		Cc:prohibits	Commercial Use, High Income Nation Use
Jurisdiction		Cc:jurisdiction	Ländercode oder URI zu einem CC:Jurisdiction Objekt

Tabelle 8: Zuweisung der Felder der *License tag Templates* zu den Ontologien

Das Template *PD-INT* setzt beispielsweise das Property Cc:permits auf alle der definierten Werte und setzt kein Cc:jurisdiction Property.

Das Template *RR-EU* hingegen setzt das Property Cc:permits auf die Werte „Reproduction“, „Distribution“ und setzt das Property Cc:jurisdiction auf den Wert „EU“.



Abbildung 22: Der License tag *RR-EU* beschreibt, dass ein Werk frei kopiert, aber nicht verändert werden darf, da die verwandten Schutzrechte in der EU abgelaufen sind

5.7.2 Nicht erweiterte Templates

Das Template *Audio file information (cylinder records)* wurde nicht erweitert, weil es bisher nur von zwölf Seiten benutzt wird. Es ist ein vertretbarer Aufwand, diese auf die neuen Templates umzustellen und dieses Template danach zu löschen.

5.7.3 Bemerkungen zur Implementierung

Die Formatierung der Bilder wird bisher nicht in den Templates gemacht, was bedeutet, dass das Datenfeld Wikisyntax enthält. Beim Erweitern wurde dies umgestellt, damit die Properties korrekt gesetzt werden können. Entsprechend müssen alle Wikiseiten, die diese Templates nutzen, angepasst werden. Die dazu benötigten regulären Ausdrücke sind in Anhang 14.6 aufgelistet.

Das Datenfeld FLAC enthält auch Wikisyntax, dazu wurde das Template *ExtractFlacFileUrl* entwickelt, das daraus die URL extrahiert und die Properties korrekt setzt.

5.8 Definition eines Datenmodells für das Public Domain Projekt

5.8.1 Probleme der bestehenden Templates

Es war vorgängig bekannt, dass mit den aktuell eingesetzten Templates Daten mehrfach erfasst werden müssen. So wird z. B. die Katalognummer und Matrizenummer auf mehreren Wikiseiten erfasst und teilweise sogar auf der selben Wikiseite wiederholt. Dies führt zu einer unnötigen Mehrarbeit und es ist klar, dass früher oder später die Daten dadurch inkonsistent sind.

Durch den Einsatz von Semantic MediaWiki können auf einer Wikiseite beliebige Daten abgefragt und angezeigt werden, so dass diese nur an einem Ort erfasst werden müssen und trotzdem alle für den Benutzer auf einer Wikiseite interessanten Daten angezeigt werden können. Dieses Potenzial soll mit einem neuen Datenmodell genutzt werden. In Semantic MediaWiki stehen dazu die Parserfunktionen `#ask` und `#show` zur Verfügung.

Beim Erweitern der aktuell eingesetzten Templates wurde festgestellt, dass nicht nur die doppelte Erfassung ein Problem ist, sondern auch der Zweck eines Templates nicht eindeutig ist. So wird das Template *Audio file* wie beschrieben sowohl zur Beschreibung von einzelnen Liedern wie auch ganzer Alben verwendet und es beschreibt Eigenschaften des Lieds, des Tonträgers, der Digitalisate und des Urheberrechtsstatus. Ein anderes Beispiel ist das Template *Media location report*, dieses beschreibt den Lagerort einer Platte und den für die Lieder verwendeten Reinigungs-/Digitalisierungsprozess. Frei nach den Prinzipien von Clean Code⁶⁴ muss ein Template aufgeteilt werden, wenn zur Beschreibung des Zwecks unvermeidbar das Wort „und“ benutzt werden muss. Das Verwenden von Formularen unterstützt dieses Vorgehen, da mit einem einzelnen Formular Wikiseiten, die aus mehreren Templates bestehen, erzeugt und bearbeitet werden können. So kann ein Formular einen Erfassungsschritt abbilden und die einzelnen Templates können zusammengehörende Daten gruppieren. Damit wird auch das Formular übersichtlicher in der Anwendung.

Bei der Arbeit mit den bisher erfassten Werken ist aufgefallen, dass die jetzigen Templates nicht geeignet sind, um mehrfach vorhandene Objekte zu erfassen. Dies kommt vor, wenn Digitalisate aus unterschiedlichen Prozessen vorliegen oder Tonträger physisch mehrfach vorhanden sind.

5.8.2 Neu zu erfassende Daten

Aus der Analyse der Anforderungen von Memoria.v und europeana war bekannt, dass in den bisherigen Daten die Sprache des Inhalts fehlt und neu mit erfasst werden muss. Des Weiteren zeigte die Analyse der verbreiteten Ontologien weitere Datenfelder, die erfasst werden sollten.

Die folgenden Datenfelder sollen deshalb neu erfasst werden:

- Sprache der Inhaltsinformation, also die gesprochene Sprache oder des Gesangs, wenn vorhanden (Dc:language)

⁶⁴ Eine Methode/Klasse soll genau eine Aufgabe erfüllen, sonst muss sie aufgeteilt werden

- Die Beziehung der Objekte zueinander (Dcterms:isPartOf/hasPart, Mo:record/Mo:track)
- Position eines Liedes innerhalb eines Tonträgers (Mo:record side, Mo:track_number)
- Urheber des Covers (Dbo:coverArtist)
- Separates Feld für die Sammlung bzw. den Sammler (Schema:funder)
- MusicBrainz ID für Lied, Tonträger und Sets (Mo:musicBrainz, Mo:musicbrainz_guid)
- Getrennte Notizen/Bemerkungen über den Zustand, zur Digitalisierung und der Lizenzabklärung (PD:ItemNotes, PD:DigitizingNotes, PD:LicenseNotes)

5.8.3 Datenmodell

Das Datenmodell besteht aus den verschiedenen Objekten, die vom Public Domain Projekt erfasst oder erzeugt werden. Es definiert, zu welchen Klassen diese Objekte gehören und welche Properties (semantische Attribute) diese besitzen.

Klassenname	Deutschsprachiger Begriff	Beschreibung	Schema.org	Music Ontology
Audio track	Lied	Einzelnes Lied (Track) auf einem Tonträger oder aus einem digitalen Angebot. Nicht zu verwechseln mit dem abstrakteren Begriff des komponierten Werkes. Kann Musik sein, aber auch Reden, Naturaufnahmen etc.	MusicRecording	Track
Audio record	Tonträger	Veröffentlichter Tonträger (CD, Schallplatte, Zip Datei etc.). Es können Tonträger aus verschiedenen Ländern oder Auflagen separat existieren.	MusicRelease	Record
Audio release set	Serie, Set	Gruppe von Tonträgern, die gemeinsam veröffentlicht wurden. Ein Album auf einer CD wird beispielsweise auf zwei Vinylplatten veröffentlicht.	MusicRelease	Release

Tabelle 9: Im Datenmodell definierte Klassen

Es wurde darauf verzichtet, Objekte für Komponisten, Mitwirkende, Plattenlabels etc. anzulegen. Detaillierte Informationen dazu sollen nicht innerhalb des Public Domain Projekts erstellt werden, sondern es soll, so wie es vom Semantic Web gedacht ist, auf Ressourcen verwiesen werden, die diese Information anbieten. Dazu gibt es im Datenmodell die Möglichkeit, URIs zu solchen externen Objekten zu setzen.

Weiterhin wurde darauf verzichtet, eine zusätzliche Hierarchiestufe einzufügen für das digitalisierte Lied (Digitized audio track). Dies ist ein Kompromiss zu Gunsten der Übersichtlichkeit im MediaWiki und der etwas vereinfachten Erfassung. In Semantic MediaWiki ist jede Wikiseite ein Objekt, es ist nicht möglich, auf einer Seite zwei Objekte zu definieren. Somit müsste für ein digitalisiertes Lied eine weitere Wikiseite angelegt werden und der Benutzer müsste nochmals weiterklicken, um an die gewünschte Information zu kommen. Der Vorteil wäre, dass ein separates Objekt es erlauben würde, Properties zu den technischen Eigenschaften des Digitalisats zu veröffentlichen (Music Ontology und PREMIS bieten sich dazu an). Dieser Mangel kann behoben werden durch die in Kapitel 3.3 vorgeschlagene Struktur des Archivspeichers und ein AIP auf Basis von Matroska.

Auch verzichtet wurde auf die Erzeugung von separaten Objekten für jeden einzelnen Tonträger, der im physischen Archiv gelagert wird. Dies auch zur Vereinfachung der Dateneingabe und Pflege. Der Zusatznutzen dabei wäre auch gering, da ein solches Objekt nur wenige Properties, besitzen würde.

In der graphischen Darstellung wurde darauf verzichtet, jedes einzelne Property darzustellen, da die Graphik sonst sehr gross werden würde, ohne den Nutzen zu erhöhen. Relevant sind hier vor allem die Beziehungen untereinander. Die Properties der Objekte sind den Tabellen aus Kapitel 5.7.1 und der Liste aus dem vorhergehenden Absatz zu entnehmen. Das Datenmodell gibt auch vor, welche Daten bei welchem Objekt erfasst werden und diese werden nur an einer Stelle erfasst. Eine Ausnahme ist der Lagerplatz eines Tonträgers und des eventuell vorhandenen Sets. Das Datenfeld Label (Publisher) und Veröffentlichungsdatum können für ein Lied und einen Tonträger unterschiedlich sein (z. B. bei einer Compilation), darum besitzen sie für beide Objekte eine andere Aussage.

Der Urheberrechtsstatus kann in diesem Model nur für Lieder und Bilder definiert werden, aber nicht für zusammengesetzte Objekte wie einen Tonträger.

Nur ein physisches Objekt wie ein Tonträger kann einen Lagerplatz besitzen. Ein Lied ist kein physisches Objekt.

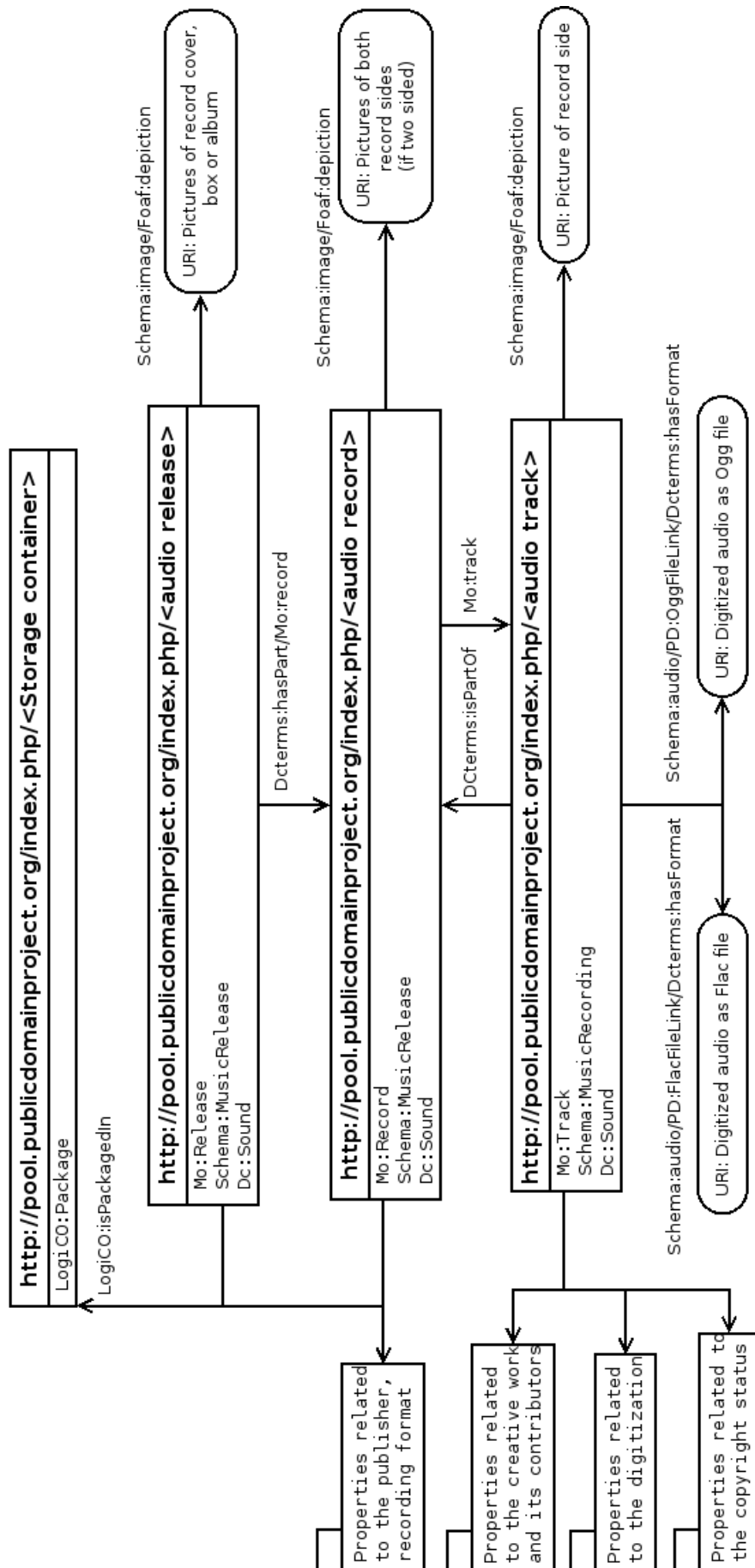


Abbildung 23: Datenmodell das für das Public Domain Projekt entwickelt wurde.

5.9 Entwicklung der neuen Templates auf Basis des Datenmodells

5.9.1 Vorgehen bei der Entwicklung der neuen Templates

Das Datenmodell aus Kapitel 5.8.3 gibt Struktur und Namen für die neuen Templates vor.

Datenfelder, die als optional betrachtet werden oder bei denen nicht erwartet wird, dass alle erfassten Objekte diese Eigenschaft besitzen, werden nur angezeigt, wenn Daten vorliegen (Mit der Parserfunktion `#if:` realisiert). Dies entschlackt die Wikiseiten für den Betrachter und fördert den Lesefluss. Ein weiterer Effekt ist, dass nicht versucht wird Platzhalter wie z. B. „none“, „unknown“ oder „?“ einzufügen.

Es wurde darauf geachtet, dass in den Datenfeldern kein Wikisyntax nötig ist für eine Korrekte Darstellung. Dies erlaubt einen besseren Datenexport und reduziert auch den Schulaufwand bei neuen Helfern.

Die allgemeine Zielgruppe möchte die digitalisierten Lieder mit kleinem Aufwand finden und herunterladen. Zu diesem Zweck werden auf den Wikiseiten von Tonträgern und Sets download Links zu allen verfügbaren Digitalisaten eingeblendet.

Die für die Zielgruppen am ehesten relevanten Informationen sollen zuoberst dargestellt werden. Es wird davon ausgegangen, dass dies Titel, Beschreibung, Bilder und Download Links sind.

Eigenschaften, die bisher üblicherweise in Datenfeldern wie auch zur Kategorisierung verwendet wurden, sollen gleich vom Template genutzt werden, um die Links zu den Kategorien zu setzen. Dies reduziert wiederum den Erfassungsaufwand und ist eine weitere Massnahme, um die Werte von Datenfeldern zu harmonisieren. Dies wird vor allem für die Felder „Label“, „Composer“, „Genre“ und „Medium“ (Format des Tonträgers) angewendet.

5.9.2 Implementation der neuen Templates

Die Templates, die im Kapitel Implementation eines Prototyps auf Basis von LogiCO entwickelt wurden, sind integraler Bestandteil des Datenmodells.

5.9.2.1 Hilfstemplates für beteiligte Personen

Es wurde eine Reihe von Templates erstellt, um die Informationen zu beteiligten Personen zu erfassen, sie darzustellen und die passenden Properties zu setzen. Dies sind: *Render contributor information*, *Smw composer*, *Smw conductor*, *Smw coverArtist*, *Smw List record content*, *Smw lyricist*, *Smw music arranger*, *Smw performer*.

Dies erlaubt es, die bisherige, detaillierte Erfassung mit Lebensdaten weiterzuführen und trotzdem die kompakte Darstellung beizubehalten. Zusätzlich können nun Links zu weiterführenden Informationen und MusicBrainz IDs erfasst werden. Wenn eine MusicBrainz ID erfasst wurde, wird ein Icon mit einem Link zur MusicBrainz Webseite eingeblendet. Zu sehen z. B. in Abbildung 26.

5.9.2.2 Template Audio record

Das Template *Audio record* dient zur Erfassung der Daten, die spezifisch für einen Tonträger sind. Dies sind die Katalognummer/Bestellnummer, das Veröffentlichungsdatum und das Tonträgerformat.

Das Template setzt einen Link zur Kategorie Audio record und deklariert so, dass es sich hier um ein Objekt der Klassen `Mo:record`, `Schema:MusicRelease`, `Dc:Sound` handelt.


Image(s)	
Content of this record	<ul style="list-style-type: none">• Side A: Symphony No. 8 in B minor, 1st movement: Allegro moderato by Franz Schubert, Audio download
Part of	Polydor 67575-67577 demo
Medium	78 rpm records (double sided, 12")
Label	Polydor Records
Cat. no.	67575
1st release date	1941
Coupling date	1940
References	Coupling date printed on record, CHARM

Abbildung 24: Ausgabe des Templates *Audio record*

Alles was im Feld „Content of this record“ ausgegeben wird, stammt von den entsprechenden *Audio track* Wikiseiten, die per semantischer Abfrage eingeblendet werden. Dazu dient das Property `PD:IsOnRecord`. Über eine semantische Abfrage (zusammen mit einer Variable und einem `#arraymap` Aufruf) wird auch das eigentlich gewünschte Property `Mo:track` gesetzt. Dies ist etwas umständlich, da SMW zwar nach inversen Properties (Subjekt und Objekt werden vertauscht) suchen kann, aber es nicht möglich ist ein Property so zu setzen.

5.9.2.3 Template Archived item

Das Template *Archived item* dient zur Erfassung der Daten, die spezifisch für einen physisch in der Sammlung der Stiftung vorhandenen Tonträger sind. Dies sind der Lagerbehälter, in dem er aufbewahrt wird, zu welcher Sammlung er gehört und optional vorhandene Bilder und Beschreibungen zum Zustand des Tonträgers (z. B. bei starker Abnutzung, Defekten, Schimmelpilzbefall etc.)

Die Angaben zum Lagerplatz werden durch entsprechende Queries (mit #show) vom Lagerbehälter abgefragt und angezeigt.

Container	EURO-SD25-00008 demo
Archive	Archive Rüti
Shelve	5
Level	2
Collection	Carl Flisch collection
Condition report(s)	

Abbildung 25: Ausgabe des Templates *Archived item*

5.9.2.4 Template Audio track

Das Template *Audio track* dient zur Erfassung der Daten, die spezifisch für ein Lied sind. Wie dem Datenmodell zu entnehmen ist, ist das der grösste Teil der vom Public Domain Projekt erfassten Daten. Dies umfasst insbesondere die Angaben zu Titel, Inhalt, beteiligte Personen, Label, Veröffentlichungsdatum und Lokalisierung innerhalb des Tonträgers.

Das Template setzt einen Link zur Kategorie Audio track und deklariert so, dass es sich hier um ein Objekt der Klassen `Mo:track`, `Schema:MusicRecording`, `Dc:Sound` handelt.

Die beteiligten Personen werden durch die schon beschriebenen Hilfstemplates erfasst.

Die Zugehörigkeit zu einem Tonträger wird durch das Datenfeld „Part of record“ und die dazugehörigen Properties `PD:IsOnRecord` und `Dcterms;isPartOf` deklariert.

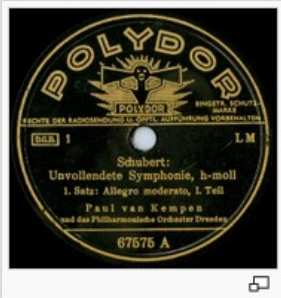


Title/Work	Symphony No. 8 in B minor, 1st movement: Allegro moderato wikipedia:Symphony No. 8 (Schubert)
Author(s)/Composer(s)	Franz Schubert (1797 - 1828) 
Image(s)	
Genre(s)	Symphony
Conductor(s)	Paul van Kempen (1893 - 1955) 
Performer(s)	Dresden Philharmonic (Orchestra, 1915 -) 
Label	Polydor Records
Part of record	Polydor 67575 demo
Record side	A
Matrix/StamperID	67575 A-M, 859 ² GE 9
Place of recording	Dresden (German Reich)
1st recording date	1940/05/25
1st release date	1941
Description	Part 1: 1st movement: Allegro moderato (part 1), Instrumental
References	Coupling date is printed on record, CHARM 

Abbildung 26: Ausgabe des Templates *Audio track*. Wenn die entsprechenden Metadaten vorhanden sind, werden weitere Zeilen eingeblendet

5.9.2.5 Template Digitized audio track

Das Template *Audio track* dient zur Erfassung der Daten, die spezifisch für ein digitalisiertes Lied sind. Dies sind die Links zu den Digitalisaten, der verwendete Reinigungs- und Digitalisierungsprozess und welcher physische Tonträger digitalisiert wurde (= aus welcher Sammlung wurde der Tonträger entnommen).

Der Link zur datenreduzierten Ogg Datei ist optional.

Ein schon vor mehreren Jahren geäußertes Wunsch von Anwendern war, eine Möglichkeit zu haben, Probleme von Digitalisaten zu melden. Zu diesem Zweck blendet dieses Template einen Link zu einem Kontaktformular ein.




High resolution audio (Flac)	Download high resolution audio file
Compressed audio (Ogg Vorbis)	Download compressed audio file
Collection	Carl Flisch collection
Cleaning process	<p>This file has no description of the used Cleaning process. Please add here the Cleaning process if available.</p>  <p>Notify the uploader with: <code>{{subst:add-desc- 1=Polydor-67575a-859ge9 demo}}</code></p>
Digitizing process	<p>This file has no description of the used Digitizing process. Please add here the Digitizing process if available.</p>  <p>Notify the uploader with: <code>{{subst:add-desc- 1=Polydor-67575a-859ge9 demo}}</code></p>
Report problems with this file	

Abbildung 27: Ausgabe des Templates *Digitized audio track*

5.9.2.6 Template License investigation


Das Template *License investigation* dient zur Erfassung der Daten, die spezifisch für den Urheberrechtsstatus sind. Dabei wird wie bisher schon unterschieden zwischen den Regionen Schweiz, Europäische Union (EU), Vereinigte Staaten von Amerika (USA) und weltweit.

Licensing [\[edit\]](#)



*This work is in the **public domain** because its copyright has **expired**.*

*This applies **worldwide**.*



[Deutsch](#) | [English](#) | [français](#) | [italiano](#) | +/-

PD CH	1992/01/01
PD EU	2012/01/01
PD USA	1992/01/01
PD INT	2012/01/01

Abbildung 28: Ausgabe des Templates *License investigation*. Die grüne Box wird „license tag“ genannt.

Das Resultat der Urheberrechtsabklärung kann sein, dass ein Lied in einer Region gemeinfrei ist und in einer anderen noch nicht. Darum muss das Feld für die Lizenz eine Liste von License-tag Templates aufnehmen können.

Leider haben diverse License-tag Templates einen Defekt. Dieser wirkt sich so aus, dass alles, was auf einer Wikiseite nach einem solchen License-tag kommt, innerhalb dessen dargestellt wird (Die Tabelle im Tag wird nicht korrekt geschlossen). Die Templates *PD-INT* und *CI* (Copyright Investigation) wurden für diese Arbeit korrigiert, die restlichen müssen noch überprüft bzw. korrigiert werden.

Wird keine Lizenz angegeben, was bei einer Neuerfassung üblich ist, wird automatisch das Template *CI* eingeblendet. Wie schon im bisherigen Arbeitsablauf im Public Domain Projekt, setzt dies die Zugehörigkeit zur Kategorie „Copyright investigation“. Diese Kategorie wird genutzt, um noch ausstehende Lizenzabklärungen abzuarbeiten.

5.9.2.7 Template Audio release set

Das Template *Audio release set* dient zur Erfassung der Daten, die spezifisch für eine veröffentlichte Gruppe von Tonträgern (Set) sind. Dies sind die dazu gehörigen Tonträger, Katalognummer/Bestellnummer, das Veröffentlichungsdatum, Label (Publisher) und Gestalter des Coverart (Cover artist).



Das Template setzt einen Link zur Kategorie Audio release set und deklariert so, dass es sich hier um ein Objekt der Klassen `Mo:release`, `Schema:MusicRelease`, `Dc:Sound` handelt.

Dieses Template wird eingesetzt, wenn mehrere Tonträger zusammengehören. In der Zeit von Schellackplatten war die Spielzeit pro Plattenseite auf ca. 4 min beschränkt, so musste ein Werk der klassischen Musik auf viele Platten aufgeteilt werden. Diese Platten wurden einzeln oder als Sets, verpackt in an Bücher erinnernde Alben, verkauft. Auch in modernerer Zeit gibt es solche Sets, die aus mehreren Vinyl Platten (LPs) oder CDs bestehen.

Da nicht jeder Tonträger einem Set angehört, wird die Information über die Zugehörigkeit zu einem Set im Template *Audio release set* gepflegt (im Gegensatz zum *Audio record*, der die Bestandteile abfragt).

Es ist denkbar dieses Template in Zukunft zu erweitern, um weitere Gegenstände zu erfassen, die zusammen mit den Tonträgern veröffentlichten wurden, wie Begleitbroschüren etc.

Das Template listet alle im Set enthaltenen Tonträger und deren Lieder auf. Zu jedem Lied, zu dem ein Digitalisat vorliegt, wird ein entsprechender Download Link eingeblendet. Hierzu werden die Hilfstemplates *Smw Query release set record content*, *Smw record content table intro*, *Smw record content table outro*, *Smw record content table row* benötigt.

Title/Work	Symphony No. 8 in B minor "Unfinished Symphony"
Image(s)	
Label	Polydor Records
Cat. no.	67575-67577
1st release date	1941
MusicBrainz ID	0987b415-dc44-37b0-a430-bbf71a1e3eb2 

Content of **Polydor 67575 demo**

Record side	Track number	Title	Composer	High resolution audio (Flac)
A		Symphony No. 8 in B minor, 1st movement: Allegro moderato	Franz Schubert	Audio download

No content of **Polydor 67576 demo**



This audio medium is missing.

Please create it with the form [Audio record](#) if available.

No content of **Polydor 67577 demo**



This audio medium is missing.

Please create it with the form [Audio record](#) if available.

Abbildung 29: Ausgabe des Templates *Audio release set*. Alles, was unterhalb der ersten Tabelle zu sehen ist, wird per semantischer Abfrage von anderen Wikiseiten ausgelesen.

5.10 Entwicklung von Formularen zum Datenmodell

Die MediaWiki Erweiterung Page Forms⁶⁵ ermöglicht die Entwicklung von Formularen, die zur Erzeugung und Bearbeitung von Wikiseiten oder Teilen davon dienen. Damit ein Formular genutzt werden kann, muss die Wikiseite Templates nutzen, denen Werte übergeben werden. Eine solche Wikiseite kann mehrere Templates in verschiedenen Abschnitten nutzen, die alle mit dem selben Formular bearbeitet werden können.

Bilder, die mit einer solchen Wikiseite angezeigt werden sollen, können im selben Arbeitsschritt mit dem Formular hochgeladen werden.

Page Forms unterstützt verschiedene Eingabemethoden, wie sie z. B. von GUI Toolkits bekannt sind, wie Check Boxen, Radio Buttons, Pull-Down Menüs, Datum Auswahl und Textfelder. Bei Textfeldern kann auch eine Autovervollständigung aktiviert werden.

5.10.1 Einschränken der möglichen Werte

Die Eingabemethoden, die nur eine begrenzte Menge an Werten dem Benutzer anbieten oder Textfelder mit Autovervollständigung benötigen als Parameter eine Quelle dieser Werte. Dies kann ein Property von Semantic MediaWiki sein, das mit `[[Allows value:<erlaubter Wert>]]` alle erlaubten Werte aufzählt oder im Formular werden die Werte aufgelistet. Pull-Down Menüs und Textfelder können auch andere Wikiseiten auflisten, diese können auf einen Namensraum (Name space), eine Kategorie oder ein Konzept eingeschränkt werden. Dabei kann jeweils angegeben werden, ob nur ein einzelner Wert oder eine Liste von Werten ausgewählt werden darf.

Eine andere Möglichkeit, bei Textfeldern die Werte einzuschränken, ist die Eingabemethode „regex“, wobei ein regulärer Ausdruck als zusätzlicher Parameter übergeben wird. Vor dem Speichern wird dieser Ausdruck auf den Feldinhalt angewendet und wenn dieser fehlschlägt (no match), wird eine definierbare Fehlermeldung ausgegeben.

Für die MusicBrainz ID (MBID) z. B. wird ein regulärer Ausdruck verwendet. Die MBID ist ein universally unique identifier (UUID)⁶⁶ in der üblichen 8-4-4-4-12 Schreibform. Um die MBID im Formular zu prüfen, wird ein regulärer Ausdruck in JavaScript kompatibler Notation⁶⁷ benötigt:

```
^([0-9a-f]{8}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{12}){0,1}$
```

5.10.2 Konzepte (Semantic MediaWiki concepts)

Ein Konzept ist eine Sammlung von Wikiseiten, die mit einer semantischen Anfrage erzeugt wird. So können z. B. Seiten aufgelistet werden, die in einer Kategorie oder einer Unterkategorie davon sind oder alle Opern die zwischen 1939 und 1945 veröffentlicht wurden. Konzepte werden auch als dynamische Kategorien bezeichnet.

⁶⁵ https://www.mediawiki.org/wiki/Extension:Page_Forms

⁶⁶ https://en.wikipedia.org/wiki/Universally_unique_identifier

⁶⁷ https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide/Regular_Expressions

5.10.3 Formular Audio record

Audio record information (audio medium) [\[edit\]](#)

Image:	<input type="text"/>	Upload file
Image1:	<input type="text"/>	Upload file
Image2:	<input type="text"/>	Upload file
Image3:	<input type="text"/>	Upload file
Description: ⓘ	<input type="text"/>	
Medium: ⓘ	<input type="text"/>	
Label: ⓘ	<input type="text"/>	
Catalogue number:	<input type="text"/>	
Order number:	<input type="text"/>	
Cover artist:	<input type="button" value="Add another cover artist"/>	
1st release date:	<input type="text"/>	<input type="text"/>
Coupling date:	<input type="text"/>	<input type="text"/>
Cutout date:	<input type="text"/>	<input type="text"/>
Reference: ⓘ	<input type="text"/>	
MusicBrainzID:	<input type="text"/>	

Description of the item(s) in our collection [\[edit\]](#)

This item is packaged in this container:	<input type="text" value="EURO-SD25-00008 demo"/>
This item is part of this collection:	<input type="text" value="Carl Flisch collection"/>
Condition report 1:	<input type="text"/>
Condition report image 1:	<input type="text"/> Upload file
Condition report 2:	<input type="text"/>
Condition report image 2:	<input type="text"/> Upload file
Condition report image 3:	<input type="text"/>
Condition report image 3:	<input type="text"/> Upload file
Condition report 4:	<input type="text"/>
Condition report image 4:	<input type="text"/> Upload file
Notes:	<input type="text"/>

Free text:

```
[[Category:Recorded music/P]] <!-- Recorded music -->
[[Category:Siemens-Polydor album 67575-67577]] <!-- record album -->
```

Summary:

This is a minor edit Watch this page

Abbildung 30: Eingabemaske des Formulars Audio record.

Der grau hinterlegte Formularteil bezeichnet ein Template (in diesem Fall *Archived item*), das nicht, ein- oder mehrfach verwendet werden darf. Mit den Knöpfen rechts kann ein Template gelöscht, verschoben oder hinzugefügt werden. Dies wird im nächsten Formular (Audio track) intensiv eingesetzt, um alle beteiligten Personen zu erfassen.

5.10.4 Formular Audio track

Title:

Title link:

Composer:

Name:

Year Born:

Year Died:

Wikilink:

MusicBrainzID:

Image:

Image1:

Image2:

Image3:

Genre:

Language:

Lyricist:

Music arranger:

Conductor:

Name:

Year Born:

Year Died:

Wikilink:

MusicBrainzID:

Performer:

Name:

Function:

Year Born:

Year Died:

Wikilink:

MusicBrainzID:

Label:

Part of:

Record side: None A B

Track number:

Matrix/StamperID:

Place of recording:

Place link:

Recording date:

1st release date:

Description:

Reference:

MusicBrainzID:

Other versions:

Abbildung 31: Eingabemaske des Formulars Audio track, oberer Teil.

Digitized audio track [\[edit\]](#)

Flac file url:

Ogg file url:

Collection:

Cleaning process:

Digitizing process:

Notes:

Licensing [\[edit\]](#)

License(s):

Public domain in CH:

Public domain in EU:

Public domain in USA:

Public domain international:

Notes:

Free text:

```
[[Category:Polydor Records]] <!-- Label information -->
[[Category:Siemens-Polydor album 67575-67577]] <!-- record album -->
[[Category:1940s music]] <!-- Music decade (1st release date is important) -->
[[Category:Symphonies]] <!-- Genre -->
[[Category:Franz Schubert]] <!-- Composer -->
[[Category:Paul van Kempen]] <!-- Conductor -->
[[Category:Dresden Philharmonic]] <!-- Performer (Orchestra) -->
[[Category:Symphony No. 8 (Schubert)]] <!-- Musical composition/Oratorio -->
[[Category:PD CH 1992]] <!-- Public Domain -->
[[Category:PD EU 2012]] <!-- Public Domain -->
[[Category:PD USA 1992]] <!-- Public Domain -->
[[Category:PD INT 2012]] <!-- Public Domain -->
```

Summary:

This is a minor edit Watch this page

Abbildung 32: Ausgabe des Formulars Audio track, unterer Teil.

Um einem Formularfeld Informationen zu dessen Verwendung oder den erwarteten Daten anzufügen, eignet sich die Parser Funktion `#info`, der ein Hinweistext übergeben wird⁶⁸. Diese Funktion zeigt ein blaues Icon mit einem Fragezeichen darin an. Wenn der Mauszeiger auf dieses Icon zeigt, erscheint der Hinweistext in einer schwebenden Box.

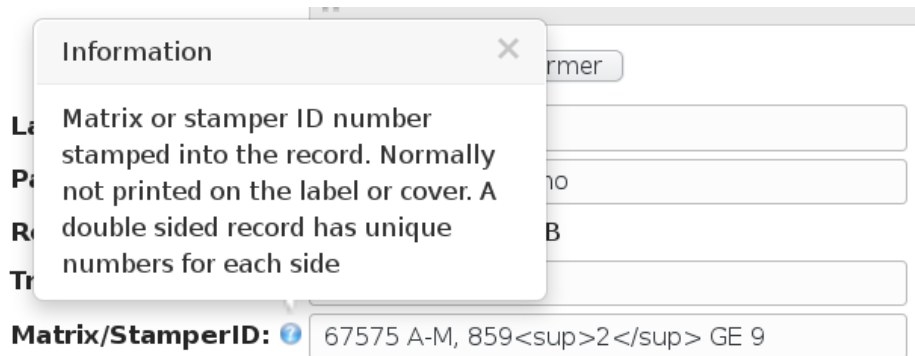


Abbildung 33: Hinweistext zum Feld Matrix/Stamper ID

⁶⁸ `#info` ist eine Funktion die von Semantic MediaWiki bereitgestellt wird

5.10.5 Formular Audio release set

Audio release set information [\[edit\]](#)

Title:	<input type="text"/>
Title link: 🔗	<input type="text"/>
Description:	<input type="text"/>
Content: 🔗	<input type="text"/>
Image:	<input type="text"/> Upload file
Image1:	<input type="text"/> Upload file
Image2:	<input type="text"/> Upload file
Image3:	<input type="text"/> Upload file
Label: 🔗	<input type="text"/>
Catalogue number:	<input type="text"/>
Order number:	<input type="text"/>
1st release date:	<input type="text"/> <input type="text"/> <input type="text"/>
Cover artist:	<input type="text"/> Add another cover artist
Reference: 🔗	<input type="text"/>
MusicBrainzID:	<input type="text"/>
Other versions: 🔗	<input type="text"/>

Description of the item(s) in our collection [\[edit\]](#)

This item is packaged in this container:	<input type="text" value="EURO-SD25-00008 demo"/>
This item is part of this collection:	<input type="text" value="Carl Flisch collection"/>
Condition report 1:	<input type="text"/>
Condition report image 1:	<input type="text"/> Upload file
Condition report 2:	<input type="text"/>
Condition report image 2:	<input type="text"/> Upload file
Condition report image 3:	<input type="text"/>
Condition report image 3:	<input type="text"/> Upload file
Condition report 4:	<input type="text"/>
Condition report image 4:	<input type="text"/> Upload file
Notes:	<input type="text"/>

[Add another archived item](#)

Free text:

```
[[Category:Siemens-Polydor album 67575-67577]] <!-- record album -->
[[Category:1940s music]] <!-- Music decade (1st release date is important) -->
[[Category:Symphonies]] <!-- Genre -->
[[Category:Symphonies by Franz Schubert]]
[[Category:Franz Schubert]] <!-- Composer -->
[[Category:Paul van Kempen]] <!-- Conductor -->
[[Category:Dresden Philharmonic]] <!-- Performer (Orchestra) -->
[[Category:Symphony No. 8 (Schubert)]] <!-- Musical composition/Oratorio -->
[[Category:PD CH 1992]] <!-- Public Domain -->
[[Category:PD EU 2012]] <!-- Public Domain -->
[[Category:PD USA 1992]] <!-- Public Domain -->
[[Category:PD INT 2012]] <!-- Public Domain -->
```

Summary:

This is a minor edit Watch this page

[Save page](#) [Show preview](#) [Show changes](#) [Cancel](#)

Abbildung 34: Eingabemaske des Formulars Audio release set.

5.10.6 Arbeitsablauf mit den neuen Formularen

1. Ein Lagerbehälter wird vom Archiv ins Büro transportiert, wo die Erfassung stattfindet.
2. Zuerst wird der Lagerbehälter erfasst, in dem die Platten oder Zylinder verpackt sind. Dazu dient das Formular: **Storage box**
3. Fotografisches Erfassen des Tonträgers, dies umfasst die Labels in der Mitte der Platte inkl. der Matrizennummer von beiden Seiten und, wenn vorhanden, die Plattenhülle (Cover) oder der Köcher eines Zylinders. Bei Plattensets (Schellackalben) wird die Front- und Stirnseite der Umverpackung fotografiert. Beschädigungen der Tonträger werden gezielt fotografiert.
4. Umbenennen/Abspeichern der Fotografien gemäss dem schon bestehenden Namensschema.
5. Erfassen des Tonträgers
Dazu dient das Formular: **Audio medium**
6. Erfassen der einzelnen Lieder
Dazu dient das Formular: **Audio track**
7. Erfassen des Plattensets
Dazu dient das Formular: **Audio release set**

Diese sechs Schritte ermöglichen das vollständige Katalogisieren der Sammlung der Schweizerischen Stiftung Public Domain.

8. Abklären des Urheberrechtsstatus der einzelnen Lieder und entsprechendes Setzen der Lizenzinformation.
9. Waschen und Digitalisieren der gemeinfreien Tonträger. Dazu existiert bereits eine separate Arbeitsanleitung.
10. Flac Dateien per FTP auf den Archivserver hochladen.
11. Die schon erfassten Lieder mit dem Digitalisat ergänzen. Dies geschieht über das selbe Formular wie schon die erste Erfassung: **Audio track**

5.11 Weiterführende Arbeiten

5.11.1 Anpassen/Automatisieren des Uploads in Wikimedia Commons

Mit den neu vorhandenen maschinenlesbaren Metadaten kann der aufwändige manuelle Prozess der Integration der Digitalisate in Wikimedia Commons wesentlich vereinfacht oder sogar automatisiert werden.

Dazu muss im wesentlichen das bestehende Metadatenmapping für das GLAM Wiki Toolset (GWT)⁶⁹ angepasst werden.

5.11.2 Automatisiertes Erstellen von Ogg/Vorbis Dateien

Mit Hilfe eines Bots können neu hinzugefügte Flac Dateien erkannt werden, um dann daraus datenreduzierte Ogg/Vorbis Dateien zu erstellen und diese in der Wikiseite zu verlinken.

Dies ist die Voraussetzung, damit direkt im Webbrowser die digitalisierten Lieder angehört werden können, denn Flac wird von keinem verbreiteten Webbrowser unterstützt⁷⁰.

5.11.3 Bereitstellen eines SPARQL Endpoints

Die Metadaten, die durch diese Masterthesis in maschinenlesbarer Form zur Verfügung gestellt werden, können von Suchmaschinen direkt genutzt werden. Semantic MediaWiki stellt auch eine API zur Verfügung, um semantische Anfragen zu stellen. Dies erfüllt die meisten Anforderungen, die von den vorgesehenen Zielgruppen Musikwissenschaftler und Bots gefordert werden. Um alle Anforderungen zu erfüllen, also um komplexere Anfragen zu stellen und um Anfragen in einer verbreiteten Sprache zu stellen, ist ein SPARQL Endpoint nötig.

Dazu muss neben der bestehenden MySQL Datenbank auch ein Triple Store installiert werden. Informationen dazu sind auf der Webseite von Semantic MediaWiki⁷¹ und im Vergleich verschiedener Triple Stores unter Open Source Lizenz⁷² zu finden.

5.11.4 Ausbau der Metadaten von Open Data zu Linked Open Data

Um die Metadaten auf Niveau Linked Open Data (5-Sterne Daten, siehe 5.1). zu bringen, muss jedes Stringliteral durch einen dereferenzierbaren URI ersetzt werden. Entsprechend muss für jedes Feld, das so ausgezeichnet werden soll, eine Linked Data Quelle vorhanden und definiert sein.

Der Wert einer solchen Aufarbeitung besteht vor allem in der gesteigerten Vergleichbarkeit der Daten mit anderen Quellen und der Eindeutigkeit der beteiligten Personen. Es ist zu erwarten, dass dies mit viel manueller Arbeit verbundenen ist und dass es nötig sein wird, sich bei den Linked Data Quellen aktiv einzubringen, um Lücken darin schliessen zu können, die sonst ein vollständiges Auszeichnen der Daten im Public Domain Projekt verhindern würden.

⁶⁹ https://commons.wikimedia.org/wiki/Commons:GLAMwiki_Toolset_Project

⁷⁰ Informationen zu HTML5 Audioformaten und Browsertest: <http://hpr.dogphilosophy.net/test/index.php>

⁷¹ https://www.semantic-mediawiki.org/wiki/Help:Using_SPARQL_and_RDF_stores

⁷² <http://db-engines.com/en/system/Blazegraph%3BRDF4J%3BVirtuoso>

Für die verschiedenen Metadatenfelder bieten sich folgende Quellen an:

- Beteiligte Personen: MusicBrainz, VIAF⁷³, WikiData
- Geographische Orte: Getty, Geonames⁷⁴, WikiData
- Genre: Library of Congress, Dbpedia
- Instrument und Stimmlage: MusicBrainz Instrument Ontology
- Musikalisches Werk: MusicBrainz,

5.11.5 Extraktion und Bereitstellung von Audiofeatures

Noch weiter gehen kann man durch die Analyse des Inhalts der Digitalisate durch Extraktion von Audiofeatures, Audiocharakterisierung oder Audiofingerprinting (z. B. mit Acoustic-Brainz⁷⁵). Die Resultate davon können wieder als Semantic Web Daten zur Verfügung gestellt werden. Dabei können Ontologien wie die Audio Features Ontology benutzt werden:

*The Audio Features Ontology allows for the description of signal processing results on the Semantic Web. This ontology expresses some common concepts to represent features of audio signals.*⁷⁶

Dies würde es ermöglichen, sehr viel weitergehende Suchanfragen zu stellen oder Werke aufgrund von ihren musikalischen Eigenschaften in Beziehung zu setzen z. B. aufgrund von Rhythmus, Tonalität, Tempo, Obertongehalt.

Mit dem Werkzeug Sonic Annotator⁷⁷ aus dem Vamp-plugins Projekt können solche Analysen automatisiert durchgeführt werden:

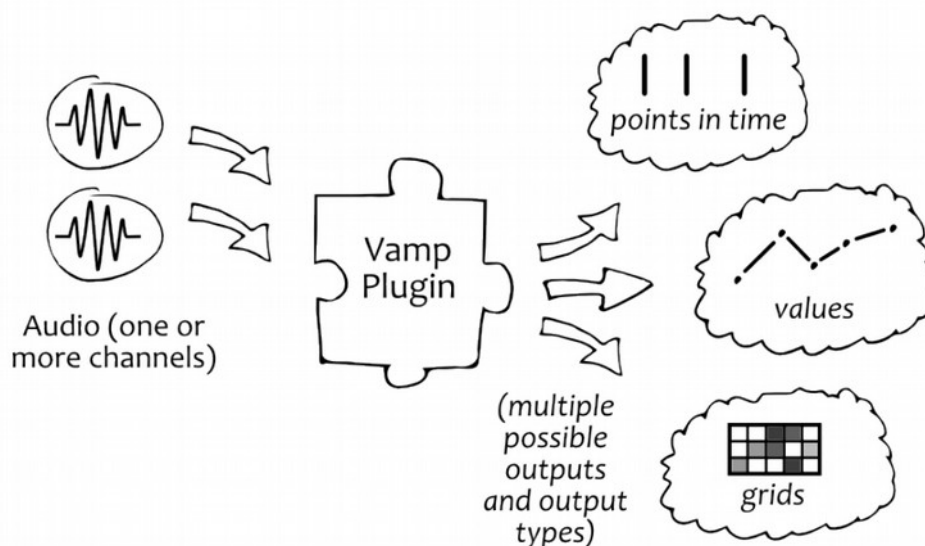


Abbildung 35: Signalflussdiagramm der Audioanalyse mit Sonic Annotator und Vamp Plugins

⁷³ <https://viaf.org/> Eine Beschreibung zu VIAF findet sich in [CHZ16], Seite 46

⁷⁴ <http://www.geonames.org>

⁷⁵ <https://acousticbrainz.org/>

⁷⁶ <http://purl.org/ontology/af/>

⁷⁷ <http://www.vamp-plugins.org/sonic-annotator/>

Sonic Annotator is a command line Vamp plugin host that outputs RDF

Key features:

- *A program for analysing large collections available locally, or on the Web.*
- *It can read a very wide range of audio file formats.*
- *Reads Vamp plugin configuration in RDF*
- *Returns the features in RDF linked with the configuration and editorial data (if available)*
- *Vamp Plugin Ontology:*

Links the results with a plugin and the enclosed algorithm that computed them.

- *Vamp Transform Ontology:*

Allows to express the parameters (e.g. Window size) that were used to obtain a particular set of results. Quelle: [CAN10]

6 Dokumentenverwaltungssystem (DMS)

Die Aufgabe war, ein Dokumentenverwaltungssystem (Document management system, DMS) zu evaluieren und zu implementieren. Damit sollen die internen Daten der Schweizerischen Stiftung Public Domain, wie Protokolle, Gesuche, Verwaltungsdokumente etc. an einem gemeinsamen Ort abgelegt werden und per Internet den Stiftungsratsmitgliedern bzw. den Personen, die sich um Fundraising kümmern, zur Verfügung gestellt werden.

6.1 Cloud Lösung

Der Wunsch des Stiftungsrats war, die Open Source Lösung NextCloud⁷⁸ für diesen Zweck einzusetzen. Für diese gibt es automatische Synchronisationsprogramme für die heute gängigen Desktop- und Smartphonebetriebssysteme.

6.2 Installation

Um die internen von den öffentlichen Daten zu trennen, wurde entschieden, dass für das DMS eine weitere virtuelle Maschine (VM) aufgesetzt wird. So sind sowohl die Zugangsdaten wie auch die Datenbanken von MediaWiki und dem DMS effektiv separiert und die Angriffsfläche auf die internen Daten wird stark reduziert. Vor allem, da beim öffentlichen MediaWiki viele Erweiterungen installiert sind, die weniger gut auf Sicherheitslücken geprüft sind als der MediaWiki Kern oder NextCloud selbst. Für diese VM wird wie bei den bisherigen VMs Debian GNU/Linux (Stable) als Betriebssystem eingesetzt.

Im Netzwerksegment, das für das Public Domain Projekt zur Verfügung steht, war noch eine einzelne IPv4 Adresse frei, die für die neue VM genutzt werden konnte. Für den DNS Eintrag standen mehrere Möglichkeiten offen:

- Neue Subdomain auf publicdomainproject.org
- Aktivieren der Domain pdproject.org, die bisher nur für E-Mails genutzt wird
- Subdomain auf publicdomain.ch
- Umzug der Webseite und des DNS Eintrags von publicdomain.ch

Da die Webseite publicdomain.ch der offizielle Internetauftritt der Stiftung ist und das DMS Daten der Stiftung verwaltet, wurde entschieden, die Webseite und den DNS Eintrag auf die neue VM umzuziehen.

Die VM ist im wesentlichen gleich eingerichtet wie die VM für das MediaWiki und die Konfiguration konnte von dort abgeschaut werden. Als Webserver ist Apache 2.4 installiert mit aktiviertem chroot, PHP 5.6 und APCu als Objekt Cache. Um E-Mails zu senden, ist Exim4 so konfiguriert, dass Mails an den Smarthost pdproject.org weitergeleitet werden.

Die verwalteten Dokumente werden auf dem Archivserver abgelegt. Dazu wurde ein neues Logical Volume mit 50 GiB und ein neuer NFS Export angelegt.

⁷⁸ <https://nextcloud.com>

NextCloud benötigt Zugriff auf eine Datenbank, dazu wurde PostgreSQL installiert. Von der Datenbank wird täglich durch einen Cronjob ein Backup erstellt. Backups, die älter als 14 Tage sind, werden gelöscht. Auszug dazu aus der Datei /etc/crontab/:

```
10 23 * * * root cd /tmp/ && (sudo -u postgres pg_dumpall | gzip >
swissfoundationpublicdomain-$(date '+\%d-\%m-\%Y').sql.gz) && mv /tm
p/swissfoundationpublicdomain* /media/publicdomain/backup/db_backup/
# ...and remove backups older than 14 days
18 0 * * * root find /media/publicdomain/backup/db_backup -name swiss-
foundationpublicdomain*.gz -ctime +14 -delete
```

Um Personen mit unterschiedlichen Aufgaben Zugriff auf die dazu nötigen Dokumente zu geben, wurden folgende Benutzergruppen erstellt:

- Stiftungsrat
- Marketing
- Serveradministratoren

Alle Stiftungsratsmitglieder, Serveradministratoren und zwei weitere beteiligte Personen haben einen Benutzeraccount erhalten und eine kurze Einführung in das neue System.

Auf den Computern im Büro des Public Domain Projekts und dem Laptop des Stiftungsratspräsidenten wurde die Desktopsoftware von NextCloud installiert. Die auf den Bürocomputern vorhandenen Daten wurden in das DMS migriert.

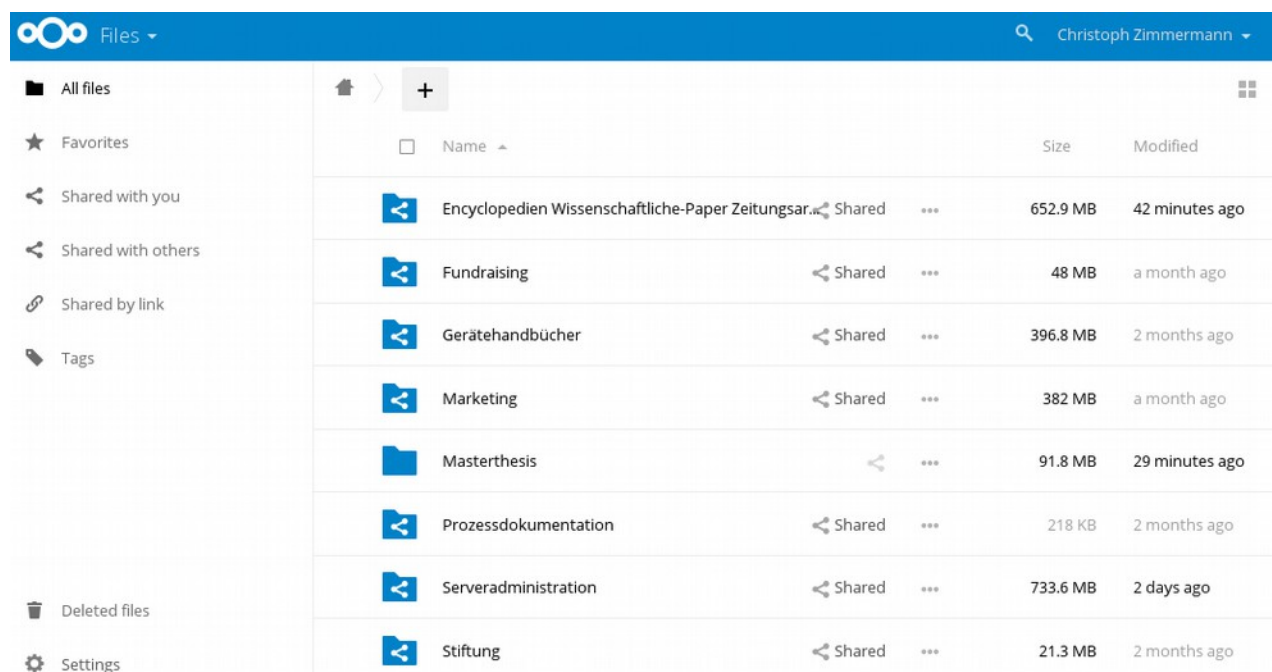


Abbildung 36: Ansicht der Ordner in der NextCloud nach der Datenmigration

6.3 Verschlüsselten Zugriff einrichten

Wenn interne Daten einer Stiftung über das Internet übertragen werden, ist klar, dass diese dazu verschlüsselt werden müssen. Darum muss für den Webserver von publicdomain.ch die Verschlüsselung mit HTTPS/TLS und entsprechenden Zertifikaten eingerichtet werden.

6.3.1 Zertifikate von Let's encrypt

Vor dem Frühling 2016 gab es zwei Möglichkeiten, um Zertifikate für die Internetübertragung zu erzeugen: Selbst erzeugte, nicht unterzeichnete Zertifikate, bei denen die Authentizität vom Benutzer selber überprüft werden muss oder gekaufte Zertifikate von einer zugelassenen Zertifizierungsstelle (CA). Dies war für viele kleinere Webseiten ein Hinderungsgrund, Verschlüsselung bei Webservern einzusetzen. Dieses Problem wurde nun vom Projekt Let's Encrypt gelöst:

Let's Encrypt is a free, automated, and open certificate authority (CA), run for the public's benefit. It is a service provided by the Internet Security Research Group (ISRG). We give people the digital certificates they need in order to enable HTTPS (SSL/TLS) for websites, for free, in the most user-friendly way we can. We do this because we want to create a more secure and privacy-respecting Web. Quelle:

<https://letsencrypt.org/about/>

Zertifikate von Let's encrypt sind nur für drei Monate gültig. Sie werden nur ausgestellt, nachdem automatisiert geprüft wurde, dass der Antragssteller die Kontrolle besitzt über den Webserver, für dessen Domainnamen (z. B. publicdomain.ch) er ein Zertifikat beantragt.

Zur Zertifikatsverwaltung und der regelmässigen Erneuerung wird das Shellscript ACME.sh⁷⁹ gemäss Anleitung eingesetzt.

6.3.2 Konfiguration des Webservers

Um die Verschlüsselung zu aktivieren, muss das Modul „ssl“ geladen werden und dessen Konfiguration angepasst werden in der Datei /etc/apache2/mods-enabled/ssl.conf. In dieser Datei müssen die von ACME.sh verwalteten Zertifikate eingetragen werden. Für die Auswahl der aktivierten Verschlüsselungsalgorithmen und deren Prioritäten wurden die Hinweise der Mozilla Foundation⁸⁰ genutzt. Der Server ist so konfiguriert, dass der Server dem Client das zu verwendende Verfahren vorschreiben kann. Natürlich nur solche, die beide Seiten beherrschen. Ein gewisser Mindeststandard (Intermediate compatibility) wird dabei verlangt, was veraltete Clients unbenutzbar macht.

Die vollständige ssl.conf ist im Anhang 14.8 zu finden.

Nach diesem Schritt wird die Vhost Konfiguration von Apache angepasst, diese ist in der Datei /etc/apache2/sites-enabled/000-default.conf. Der bestehende Block für den Port 80 (HTTP) wird kopiert und als zweiter Block für den Port 443 (HTTPS) eingefügt. Beim

⁷⁹ <https://github.com/Neilpang/acme.sh>

⁸⁰ https://wiki.mozilla.org/Security/Server_Side_TLS#Recommended_configurations Intermediate compatibility : For services that don't need compatibility with legacy clients (mostly WinXP), but still need to support a wide range of clients

ersten wird nun SSLengine off gesetzt und beim zweiten entsprechend SSLengine on. Nach einem Neustart des Webservers kann die Webseite nun unter <https://publicdomain.ch> verschlüsselt erreicht werden.

Eine weitere getroffene Massnahme ist das Aktivieren von HTTP Strict Transport Security (HSTS). Das dazu genutzte Tutorial⁸¹ erklärt dieses Feature so:

The HTTP Strict Transport Security feature lets a web site inform the browser that it should never load the site using HTTP, and should automatically convert all attempts to access the site using HTTP to HTTPS requests instead.

Das bedeutet, dass ein Webbrowser, nachdem er die Webseite einmal verschlüsselt erreicht hat, in Zukunft alle unverschlüsselten Anfragen automatisch ändert in verschlüsselte. So wird verhindert, dass ein Angreifer die Verschlüsselung unterdrücken kann. Für dieses Feature muss sichergestellt sein, dass das Apache Modul „headers“ aktiviert ist.

Die vollständige Vhost Konfiguration ist im Anhang 14.8 zu finden.

NextCloud bringt von sich aus keine Weiterleitung von HTTP auf HTTPS mit, dies muss manuell ergänzt werden in der Datei /var/www/html/cloud/.htaccess mit:

```
<IfModule mod_rewrite.c>
  RewriteCond %{HTTPS} off
  RewriteRule (.*) https://%{HTTP_HOST}%{REQUEST_URI}
</IfModule>
```

6.3.3 Prüfung

Zur Überprüfung der getroffenen Massnahmen wurde das SSL Testwerkzeug von Qualys eingesetzt:

- Der Webserver unter publicdomain.ch hat bei der Prüfung ein A+⁸² erreicht.

Der detaillierte Report ist im Anhang 14.9 zu finden.

6.3.4 Erhalt der Sicherheit (Security)

Sicherheit (Security) ist kein fester Zustand, der einmal hergestellt für immer Bestand hat. Vielmehr ist dazu ein kontinuierliches Überwachen und Aktualisieren nötig, damit neu bekanntgewordene Sicherheitslücken geschlossen werden.

Dafür werden im Public Domain Projekt folgende Strategien eingesetzt:

- Abonnement der Mailingliste für Debian Sicherheitsaktualisierungen⁸³
- Abonnement der Mailingliste für Ankündigungen über MediaWiki Aktualisierungen⁸⁴
- NextCloud ist so konfiguriert, dass Administratoren über verfügbare Updates informiert werden

⁸¹ https://raymii.org/s/tutorials/HTTP_Strict_Transport_Security_for_Apache_NGINX_and_Lighttpd.html

⁸² <https://www.ssllabs.com/ssltest/analyze.html?d=publicdomain.ch>

⁸³ <https://lists.debian.org/debian-security-announce/>

⁸⁴ <https://lists.wikimedia.org/mailman/listinfo/mediawiki-announce>

Damit die Updates nicht für jeden einzelnen Server separat installiert werden müssen (Einloggen, Updates abrufen und installieren, ausloggen), bietet sich der Einsatz eines Softwarewerkzeugs zur Updateverwaltung von Serverclustern an.

Im Projekt wird dazu bereits `apt-dater`⁸⁵ eingesetzt. Sowohl das Verwaltungswerkzeug wie auch der Updateclient `apt-dater-host` sind in der Debian Paketverwaltung enthalten. Die neu erstellte VM wurde entsprechend auch damit ausgerüstet.

7 Zusätzlich ausgeführte Arbeiten

7.1 Authentizität des Zugriffs sicherstellen

Der MediaWiki Server wurde umkonfiguriert, damit dieser verschlüsselt erreichbar ist.

Wieso ist das für öffentliche und frei nutzbare Daten wichtig?

Beim Benutzerzugriff auf Daten aus einem Langzeitarchiv muss nicht nur sichergestellt werden, dass die Daten auf dem Weg zwischen Langzeitarchiv und Benutzer nicht verändert werden, sondern auch, dass ein Benutzer zweifelsfrei feststellen kann, dass die Daten aus genau dem Langzeitarchiv stammen, von dem er die Daten angefordert hat.

Dies wird durch den Einsatz von unterzeichneten Zertifikaten und Verschlüsselung sichergestellt.

Die Konfiguration des Webservers und die Zertifikatsverwaltung entspricht dem im Kapitel 6.3 beschriebenen vorgehen mit Zertifikaten von Let's encrypt.

7.1.1 Prüfung

Zur Überprüfung der getroffenen Massnahmen wurde das SSL Testwerkzeug von Qualys eingesetzt:

- Der Webserver unter `publicdomainproject.org` hat bei der Prüfung ein A⁸⁶ erreicht.

Nur A (und kein A+ wie bei `publicdomain.ch`), weil bei diesem Server keine HTTP Strict Transport Security (HSTS) aktiviert ist. Dies ist nicht möglich, weil der online Radio Server die Streams nicht verschlüsselt ausliefert. Der dazu eingesetzte `icecast2` wäre in der Lage, die Streams des Public Domain Projekts verschlüsselt auszuliefern. Die dadurch verursachte zusätzliche Serverbelastung wird als nicht nutzbringend erachtet, da bei einem Radiostream die Authentizität nicht im gleichen Mass gewährleistet werden muss.

Der detaillierte Report ist im Anhang 14.9 zu finden.

7.2 Korrekturen im MediaWiki

Während der Arbeit wurden immer mal wieder kleinere Probleme an den aktuellen Artikeln oder dem Kategorienbaum entdeckt, die gleich korrigiert wurden. So wurden z. B. mehrere zirkuläre Verbindungen der Kategorien entdeckt und aufgelöst.

⁸⁵ <https://github.com/DE-IBH/apt-dater>

⁸⁶ <https://www.ssllabs.com/ssltest/analyze.html?d=publicdomainproject.org>

8 Schlussfolgerungen/Fazit

In der Evaluationsphase hat sich bestätigt, dass das Dateiformat Flac auch für die Langzeitarchivierung sehr gut geeignet ist. Die Eigenschaften dieses Dateiformats erlauben eine einfache, zuverlässige Prüfung der Datenintegrität. Es ist daher für Archivanwendungen dem verbreiteten WAV Format vorzuziehen.

Das Dokumentenverwaltungssystem NextCloud, das zu Beginn dieser Arbeit eingerichtet wurde, hat sich in den letzten Monaten im täglichen Einsatz bewährt und hat den Umgang mit Dokumenten innerhalb des Public Domain Projekts wesentlich vereinfacht.

Im Bereich der Bestandssicherung der Archivdaten des Public Domain Projekts konnte ein grosser Fortschritt durch die Migration des Archivspeichers von Plan9 auf Gentoo GNU/Linux erzielt werden. Dadurch sind die Festplatten nun besser überwacht und Ausfälle werden den Administratoren per E-Mail gemeldet.

Für den Archivserver liegen nun fast alle benötigten Repräsentationsinformationen vor. Derzeit noch ausschliesslich in digitaler Form, da eine Schublade mit den essentiellsten Repräsentationsinformationen in Papierform nicht mehr während dieser Arbeit zusammengestellt und eingebaut werden konnte. Die fehlenden Repräsentationsinformationen sind zum einen der ATA Kommandostandard und der SATA Verbindungsstandard zur Ansteuerung der Festplatten, da diese nicht frei zugänglich sind und noch gekauft werden muss(t)en und zum anderen die Information, wie die Daten auf der Oberfläche der Festplatte organisiert sind, da diese Information nur der Festplattenhersteller und wenige Datenrettungsfirmen besitzen.

Das favorisierte Monitoringsystem Sensu wurde während dieser Arbeit zwar im Serverraum installiert, aber noch nicht in Betrieb genommen. Das liegt daran, dass im Dezember vor den Festtagen kein Termin für den Umbau im Serverraum der ZHDK mehr zustande kam. Der Umbau hat erst am 4. Januar stattgefunden. Zu Beginn der Masterthesis wurde fälschlicherweise davon ausgegangen, dass die Monitoringlösungen einen Server über bestehende Zugänge wie SSH und Dienste über deren zugängliche Netzwerkports überwachen. Darum besteht mit Sensu jetzt das Problem, dass Gentoo GNU/Linux nicht nativ unterstützt wird und noch unklar ist, wie der Sensu Client manuell installiert werden kann.

Mit dieser Arbeit konnte das Qualitätsniveau der öffentlich zugänglichen Metadaten der archivierten Tonträger von 2-Sterne Open Data auf vier Sterne erhöht werden (siehe dazu Kapitel 5.1), wobei ein Teil der Daten auch schon die Anforderungen von fünf Sternen erfüllt. Der Schritt auf das 5-Sterne Niveau ist weniger eine technische Herausforderung, da die Infrastruktur jetzt zur Verfügung steht, sondern eine Herausforderung der Erschliessung, da alle derzeit erfassten Werke in die neuen Templates überführt werden müssen und dabei zusätzliche Felder mit Referenzlinks zu Authority control files erfasst werden müssen. Die Daten sind mit dem präsentierten Datenmodell maschinenlesbar und mehrheitlich mit weit verbreiteten Ontologien wie Dublin Core und Schema.org repräsentiert. Damit sollten sich auch die vom Public Domain Projekt erhofften Ziele, die einfache Weiternutzung der Metadaten und eine einfachere Zusammenarbeit mit Memoriav und europeana, erfüllen lassen.

9 Literaturverzeichnis

- CAN10:** Chris Cannam, Michael O. Jewell, Christophe Rhodes, Mark Sandler, and Mark d'Inverno, *Linked Data and You: Bringing music research software into the Semantic Web*, , , 2010
- CCSDS652:** , *AUDIT AND CERTIFICATION OF TRUSTWORTHY DIGITAL REPOSITORIES*, CCSDS, CCSDS 652.0-M-1, <http://public.ccsds.org/publications/archive/652x0m1.pdf>, 2011
- CHZ16:** Christoph Zimmermann, *Projektarbeit 2 – Langzeitarchiv für Audiowerke*, 2016, Berne University of Applied Science, http://pool.publicdomainproject.org/index.php/File:Projekt_2_-_Langzeitarchiv_f%C3%BCr_digitale_Audiowerke_Christoph_Zimmermann_mit_Anhang.pdf
- FUR11:** Fabian M. Fürste, *Linked Open Library Data – Bibliographische Daten und ihre Zugänglichkeit im Web der Daten*, Dingel & Frick, 978-3-934997-36-3, , 2011
- GYÖ12:** György Fazekas and Thomas Wilmering, *Semantic Web and Semantic Audio technologies*, 132nd AES Convention, <http://www.isophonics.net/sites/isophonics.net/files/AES132-tutorial.pdf>, 2012
- KEI13:** Christian Keitel, Astrid Schoger, *Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644*, Beuth, 978-3-410-23499-9, , 2013
- NES13:** nestor-Arbeitsgruppe OAIS-Übersetzung / Terminologie, *Referenzmodell für ein Offenes Archiv-Informationssystem – Deutsche Übersetzung, Version 2*, nestor, urn:nbn:de:0008-2013082706, <http://nbn-resolving.de/urn:nbn:de:0008-2013082706>, 2013
- OAIS12:** , *REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS)*, CCSDS, CCSDS 650.0-M-2, <http://public.ccsds.org/publications/archive/650x0m2.pdf>, 2012
- RAI07:** Yves Raimond, Samer Abdallah, Mark Sandler, Frederick Giasson, *The music ontology*, Austrian computer society (OCG), , 2007
-

10 Tabellenverzeichnis

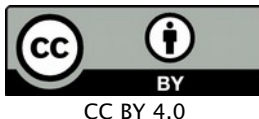
Tabelle 1: Die wichtigsten Begriffe für diese Projektarbeit aus dem OAIS Referenzmodell in deutsch und englisch. Aus [NES13], Seiten 8 bis 16.....	15
Tabelle 2: Vergleich der Begriffe zur Beschreibung von Tonträgern in den Ontologien von MusicBrainz, Schema.org und Music Ontology.....	15
Tabelle 3: Zuweisung der Felder des Storage box Template zu den Ontologien.....	68
Tabelle 4: Zuweisung der Felder des Audio file Template zu den Ontologien.....	72
Tabelle 5: Zuweisung der Felder des Audio storage description Template zu den Ontologien.....	73
Tabelle 6: Zuweisung der Felder des Phonograph cylinder description Template zu den Ontologien.....	73
Tabelle 7: Zuweisung der Felder des Storage media location report Template zu den Ontologien.....	74
Tabelle 8: Zuweisung der Felder der License tag Templates zu den Ontologien.....	75
Tabelle 9: Im Datenmodell definierte Klassen.....	77

11 Abbildungsverzeichnis

Abbildung 1: Beispiel einer Zelluloid Walze: Ansicht auf die Verpackung und die Rillen	20
Abbildung 2: Beispiel einer Zelluloid Walze. Ansicht auf die Stirnseite.....	20
Abbildung 3: Vorgeschlagene Systemarchitektur zur Langzeiterhaltung im Public Domain Projekt [CHZ16], Seite 50.....	22
Abbildung 4: Verbindung des Archivspeichers mit den Produzenten und Benutzern.....	24
Abbildung 5: Bisherige Struktur des Archivspeichers bestehend aus Disc Shelf und Storage controller.....	25
Abbildung 6: Architektur der Monitoringlösung Sensu.....	31
Abbildung 7: Ausschnitt aus der Systemübersicht von Invenio v3. Quelle: https://invenio.readthedocs.io/en/latest/architecture/system-overview.html	33
Abbildung 8: Geplante Struktur des Archivspeichers.....	41
Abbildung 9: Erster Migrationsschritt hin zur neuen Struktur des Archivspeicher.....	42
Abbildung 10: Coriad SR1521 Disc shelf und CLN/21 Storage controller.....	43
Abbildung 11: Ausgaben zu Kernelversion, RAID Status und Festplattenbelegung nach der Migration.....	50
Abbildung 12: Ausgabe der IPMI Management Karte nach der Einrichtung.....	52
Abbildung 13: Die zu Sensu gehörende Weboberfläche Uchiwa nach der Installation im Serverraum.....	56
Abbildung 14: 5-Sterne-Modell für Offene Daten (Open Data), CC0, http://5stardata.info/de/	58
Abbildung 15: UseCase Diagramm der vorgesehenen Zielgruppen, Quelle: [CHZ16], Seite 40.....	59
Abbildung 16: Klassen und deren Beziehungen die von der Music Ontology definiert werden. Quelle: [GYÖ12], Seite 98.....	63
Abbildung 17: Beziehungen zwischen der Music- und Studioontologie Quelle: [GYÖ12], Seite 101.....	64
Abbildung 18: Ausgabe des Templates Storage box.....	68
Abbildung 19: Properties und deren Werte von einem Lagerbehälter (Beispiel).....	69
Abbildung 20: Eingabemaske des Formulars Storage box.....	70
Abbildung 21: Der License tag PD-INT beschreibt, dass ein Werk weltweit gemeinfrei ist	74
Abbildung 22: Der License tag RR-EU beschreibt, dass ein Werk frei kopiert, aber nicht verändert werden darf, da die verwandten Schutzrechte in der EU abgelaufen sind.....	75
Abbildung 23: Datenmodell das für das Public Domain Projekt entwickelt wurde.....	79
Abbildung 24: Ausgabe des Templates Audio record.....	81
Abbildung 25: Ausgabe des Templates Archived item.....	82
Abbildung 26: Ausgabe des Templates Audio track. Wenn die entsprechenden Metadaten vorhanden sind, werden weitere Zeilen eingeblendet.....	83
Abbildung 27: Ausgabe des Templates Digitized audio track.....	84
Abbildung 28: Ausgabe des Templates License investigation. Die grüne Box wird „license tag“ genannt.....	84
Abbildung 29: Ausgabe des Templates Audio release set. Alles, was unterhalb der ersten Tabelle zu sehen ist, wird per semantischer Abfrage von anderen Wikiseiten ausgelesen.....	86
Abbildung 30: Eingabemaske des Formulars Audio record.....	88
Abbildung 31: Eingabemaske des Formulars Audio track, oberer Teil.....	89
Abbildung 32: Ausgabe des Formulars Audio track, unterer Teil.....	90
Abbildung 33: Hinweistext zum Feld Matrix/Stamper ID.....	90
Abbildung 34: Eingabemaske des Formulars Audio release set.....	91
Abbildung 35: Signalflussdiagramm der Audioanalyse mit Sonic Annotator und Vamp Plugins.....	94
Abbildung 36: Ansicht der Ordner in der NextCloud nach der Datenmigration.....	97

12 Lizenz

Dieses Werk ist unter einer Creative Commons Lizenz vom Typ Namensnennung 4.0 International zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie <https://creativecommons.org/licenses/by/4.0/> oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.



Sie dürfen:

Teilen — das Material in jedwedem Format oder Medium vervielfältigen und weiterverbreiten

Bearbeiten — das Material remixen, verändern und darauf aufbauen und zwar für beliebige Zwecke, sogar kommerziell.

Der Lizenzgeber kann diese Freiheiten nicht widerrufen, solange Sie sich an die Lizenzbedingungen halten.

Unter folgenden Bedingungen:

Namensnennung — Sie müssen angemessene Urheber- und Rechteangaben machen, einen Link zur Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Diese Angaben dürfen in jeder angemessenen Art und Weise gemacht werden, allerdings nicht so, dass der Eindruck entsteht, der Lizenzgeber unterstütze gerade Sie oder Ihre Nutzung besonders.

Keine weiteren Einschränkungen — Sie dürfen keine zusätzlichen Klauseln oder technische Verfahren einsetzen, die anderen rechtlich irgendetwas untersagen, was die Lizenz erlaubt.

12.1 Lizenz der verwendeten Bilder

Bei Bildern die von anderen Urhebern stammen ist jeweils die Quelle angegeben. Die Lizenzbedingungen können abweichend sein von der Lizenz dieser Arbeit.

Quelle des Titelbildes: Courtesy of the Recorded Sound Section, MBRS Division, Library of Congress

13 Selbständigkeitserklärung

Ich bestätige, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der im Literaturverzeichnis angegebenen Quellen und Hilfsmittel angefertigt habe. Sämtliche Textstellen, die nicht von mir stammen, sind als Zitate gekennzeichnet und mit dem genauen Hinweis auf ihre Herkunft versehen.

Ort, Datum:

Unterschrift:

14 Anhang

Anhangsverzeichnis

14Anhang.....	106
14.1Einrichtung des MediaWiki Testsystems.....	107
14.1.1Testserver.....	107
14.1.2Entwicklungsrechner.....	107
14.1.2.1Installation dnsmasq.....	108
14.2Verwendete Ontologien von verwandten Projekten.....	109
14.3MediaWiki Code zu den LogiCO Prototypen.....	111
14.3.1.1Definition der externen Ontologie: Smw import LogiCO.....	111
14.3.1.2Template Unit conversion length.....	111
14.3.1.3Property mit Einheitenumrechnung: LogiCO:hasLengthValue.....	111
14.3.1.4Property mit Wertbereichseinschränkung: LogiCO:hasLocation.....	111
14.3.1.5Archive Rütli.....	112
14.3.2Template Storage box.....	112
14.3.3Template Storage box size.....	114
14.3.4Concept Storage box.....	114
14.3.5Formular Storage box.....	115
14.4Beispieldaten.....	116
14.4.1Beispieldaten wie sie bisher menschenlesbar im Browser dargestellt werden..	116
14.4.2Beispieldaten eines Lieds aus der MediaWiki API.....	118
14.4.3RDF/XML Beispieldaten eines Lieds vor der Masterthesis.....	119
14.4.4RDF Beispieldaten eines Lieds nach der Masterthesis.....	122
14.4.5RDF Beispieldaten eines Lagerbehälters.....	131
14.5Listen der erstellten oder geänderten MediaWiki Seiten.....	135
14.5.1.1Archive.....	135
14.5.1.2Demoseiten.....	135
14.5.1.3Liste der importierten Ontologien.....	135
14.5.2Liste der erstellten Properties.....	135
14.5.3Liste der geänderten Templates.....	138
14.5.4Liste der neu erstellten Templates.....	139
14.5.5Liste der erstellten Formulare.....	140
14.5.6Liste der neu erstellten Kategorien.....	140
14.6Reguläre Ausdrücke für Datenmigration in semantische Templates.....	141
14.7Anpassungen an der Konfiguration des MediaWiki Servers.....	142
14.7.1Anpassen der Limiten von Semantic MediaWiki.....	142
14.7.2Installation von MassEditRegex.....	142
14.7.3Fehler zend_mm_heap corrupted beheben:.....	143
14.7.4Setzen der korrekten Serveradresse.....	143
14.8Apache2 Konfiguration.....	144
14.8.1Vhost Konfiguration.....	144
14.8.2SSL.conf.....	146
14.9Security Testberichte von Qualys SSL Labs.....	149
14.10Detailliertere Planung und Aufwandsabschätzung der Systemarchitektur.....	150
14.11Aufgabenstellung Masterthesis.....	151

14.1 Einrichtung des MediaWiki Testsystems

14.1.1 Testserver

Für die Entwicklungs- und Testzwecke wurde ein Testsystem eingerichtet, das dem produktiv laufenden MediaWiki Server entspricht. Im Public Domain Projekt werden alle Dienste von virtualisierten Servern bereitgestellt, dies machte das Einrichten eines Testsystems besonders einfach.

Für diese Arbeit stand ein vorbereiteter Test-PC zur Verfügung. Auf diesem sind das gleiche Betriebssystem (Debian 7, Wheezy) und die gleiche Virtualisierungslösung (KVM mit Proxmox 3.) installiert.

Die MediaWiki Installation musste zuerst auf den aktuellen Stand gebracht werden, da die installierte Semantic MediaWiki Version nicht kompatibel war zum installierten MediaWiki. Die Aktualisierung beinhaltet den Wechsel von der bisherigen MediaWiki LTS (Long term stable) Version 1.23 zur aktuellen LTS Version 1.27, den Wechsel von Semantic MediaWiki 1.9 auf 2.4 und die Aktualisierung aller installierten Extensions. Wo es möglich war, wurden die Extensions neu nicht mehr per Tarball installiert, sondern via Git und dies im Konfigurationsfile dokumentiert. Dies erleichtert wesentlich zukünftige Aktualisierungen der Extensions, die mit einem `git pull && git switch REL_1.<Versionsnummer>` pro Extension erledigt sind.

Nach dem alles auf dem aktuellen Stand war, wurde vom MediaWiki Server ein Backup erstellt und dieses danach auf den Test-PC heruntergeladen. Nach dem die Backupdateien in den lokalen Proxmox-Backupordner verschoben wurden, konnte der MediaWiki Server in zwei Schritten lokal angelegt werden:

- Erstellen einer neuen VM mit der selben VM ID wie auf dem produktiven System (in diesem Fall ID101), die restlichen Einstellungen sind egal, da diese auch im Backup abgelegt werden und entsprechend auch wiederhergestellt werden.
- Im Tab Backup kann nun das Backup des MediaWiki Servers mit der Funktion Restore lokal importiert werden.

Kleine Einstellungen mussten an das lokale Netzwerk angepasst werden, wie dass im Test-PC nur eine Netzwerkkarte vorhanden ist und dem MediaWiki Server musste eine lokale IP Adresse vergeben werden.

14.1.2 Entwicklungsrechner

Der nächste Schritt bestand darin, auf dem Entwicklungsrechner (Debian Sid) eine einfache Möglichkeit zu haben, um zwischen dem Produktiven- und dem Testserver zu wechseln, die andere IP Adressen besitzen. Dazu müssen die selben URLs (z. B. de.publicdomainproject.org) vom DNS andere IP Adressen zurückgeliefert bekommen, sonst funktionieren die Vhost Einträge des Webservers Apache (2.4) nicht mehr.

Gelöst wurde das mit dnsmasq, der auf dem Entwicklungsrechner zusätzlich installiert wurde. Wenn dnsmasq läuft, wird der Testserver angesprochen, sonst wird der produktiv Server angesprochen.

```
Dnsmasq starten:  sudo systemctl start dnsmasq.service
```

```
Dnsmasq stoppen: sudo systemctl stop dnsmasq.service
```

14.1.2.1 Installation dnsmasq

Dnsmasq⁸⁷ ist ein kleiner DNS Server der auf kleinen Routern die Funktion des DHCP Servers übernimmt und gleich als lokaler Cache für DNS abfragen eingesetzt wird. In dieser Installation wird nur die DNS Chaching Funktion genutzt.

```
su
apt-get install resolvconf dnsmasq
systemctl disable dnsmasq.service
```

Danach die in diesem Dokument als Anhang beigefügte Datei publicdomain-testserver.conf in den Ordner /etc/dnsmasq.d/ kopieren und dnsmasq neustarten.

Auf dem Entwicklungsrechner ist Debian mit systemd installiert, entsprechend muss um dnsmasq neuzustarten folgendes Kommando verwendet werden:

```
systemctl restart dnsmasq.service
```

Inhalt der Datei publicdomain-testserver.conf:

```
# Never forward plain names (without a dot or domain part)
domain-needed
# Never forward addresses in the non-routed address spaces.
bogus-priv

# If you don't want dnsmasq to read /etc/resolv.conf or any other
# file, getting its servers from this file instead (see below), then
# uncomment this.
no-resolv

# If you don't want dnsmasq to poll /etc/resolv.conf or other resolv
# files for changes and re-read them then uncomment this.
no-poll

# Add other name servers here, with domain specs if they are for
# non-public domains.
server=192.168.1.1

# Add domains which you want to force to an IP address here.
# The example below send any host in double-click.net to a local
# web-server.
address=/publicdomainproject.org/192.168.1.33
address=/en.publicdomainproject.org/192.168.1.33
address=/de.publicdomainproject.org/192.168.1.33
address=/fr.publicdomainproject.org/192.168.1.33
address=/it.publicdomainproject.org/192.168.1.33
address=/es.publicdomainproject.org/192.168.1.33
address=/pool.publicdomainproject.org/192.168.1.33
```

⁸⁷ <http://www.thekelleys.org.uk/dnsmasq/docs/dnsmasq-man.html>

14.2 Verwendete Ontologien von verwandten Projekten

MusicBrainz

- LinkedBrainz (lb): <https://github.com/LinkedBrainz/r2rml/v0.1#>
- MusicBrainz Schema (ngs): <http://musicbrainz.org/NGS>
- skos: <http://www.w3.org/2004/02/skos/core#>
- Dublin core (dc): <http://purl.org/dc/elements/1.1/>
dct: <http://purl.org/dc/terms/>
- Music Ontology (mo): <http://purl.org/ontology/mo/>
- Friend of a friend (foaf): <http://xmlns.com/foaf/0.1/>
- Open vocabulary (ov): <http://open.vocab.org/terms/>
- Geo: http://www.w3.org/2003/01/geo/wgs84_pos#
- Event: <http://purl.org/NET/c4dm/event.owl#>
- is: <http://purl.org/ontology/is/core#>
- isi: <http://purl.org/ontology/is/inst/>
- Time: <http://www.w3.org/2006/time#>
- Time line (tl): <http://purl.org/NET/c4dm/timeline.owl#>

Dbpedia <http://mappings.dbpedia.org/server/ontology/classes/>

- Provenienz (prov): <http://www.w3.org/ns/prov#>
- dbo: <http://dbpedia.org/ontology/>
- foaf: <http://xmlns.com/foaf/0.1/>
- dct: <http://purl.org/dc/terms/>

Die Dbpedia Ontologie pflegt wenn es geht owl : sameAs Beziehungen, was andere leider nicht machen. Darüber ist dbo stark mit Schema verlinkt.

Wikidata https://www.wikidata.org/wiki/Wikidata:Main_Page

Wikidata nutzt 18 eigene Ontologien

- wikibase: <http://wikiba.se/ontology-beta#>
- wdata: <https://www.wikidata.org/wiki/Special:EntityData/>
- wd: <http://www.wikidata.org/entity/>
- wds: <http://www.wikidata.org/entity/statement/>
- wdref: <http://www.wikidata.org/reference/>
- wdv: <http://www.wikidata.org/value/>
- wdt: <http://www.wikidata.org/prop/direct/>
- p: <http://www.wikidata.org/prop/>

- ps: <http://www.wikidata.org/prop/statement/>
- psv: <http://www.wikidata.org/prop/statement/value/>
- psn: <http://www.wikidata.org/prop/statement/value-normalized/>
- pq: <http://www.wikidata.org/prop/qualifier/>
- pqv: <http://www.wikidata.org/prop/qualifier/value/>
- pqn: <http://www.wikidata.org/prop/qualifier/value-normalized/>
- pr: <http://www.wikidata.org/prop/reference/>
- prv: <http://www.wikidata.org/prop/reference/value/>
- prn: <http://www.wikidata.org/prop/reference/value-normalized/>
- wdno: <http://www.wikidata.org/prop/novalue/>
- skos: <http://www.w3.org/2004/02/skos/core#>
- Schema: <http://schema.org/>
- Creative Commons (cc): <http://creativecommons.org/ns#>
- Geo: <http://www.opengis.net/ont/geosparql#>
- Provenienz (prov): <http://www.w3.org/ns/prov#>

Europeana Sounds <http://www.europeanasounds.eu/>

- Europeana data model (edm): <http://www.europeana.eu/schemas/edm/EDM.xsd>
- Dublin core (dc): <http://purl.org/dc/elements/1.1/>
dct: <http://purl.org/dc/terms/>
- Friend of a friend (foaf): <http://xmlns.com/foaf/0.1/>
- skos: <http://www.w3.org/2004/02/skos/core#>
- rdaGr2: <http://rdvocab.info/ElementsGr2/>
- Geo: http://www.w3.org/2003/01/geo/wgs84_pos#
- Creative Commons (cc): <http://creativecommons.org/ns#>
- Open Digital Rights Language (odrl): <https://www.w3.org/ns/odrl/2/ODRL21>

Memobase (Memoriav) <http://memobase.ch/>

- Memobase Core basierend auf EBUcore, fast alle Felder verweisen auf DCterms

14.3 MediaWiki Code zu den LogiCO Prototypen

14.3.1.1 Definition der externen Ontologie: Smw import LogiCO

```
http://ontology.tno.nl/logico#[http://ontology.tno.nl/logico/ Logistics
Core Ontology]
Box|Category
Carton|Category
contains|Type:Page
hasCapacity|Type:Number
hasLengthValue|Type:Quantity
hasWidthValue|Type:Quantity
hasHeightValue|Type:Quantity
hasWeightValue|Type:Quantity
hasAddress|Type:Text
hasLocation|Type:Page
id|Type:Text
hosts|Type:Page
isLocationOf|Type:Page
isFull|Type:Boolean
isPackagedIn|Type:Page
Location|Category
MoveableResource|Category
Object|Category
Package|Category
PhysicalResource|Category
```

14.3.1.2 Template Unit conversion length

```
<noinclude>
Unit definition for custom units dealing with length

Currently supported are:
</noinclude>
*[[Corresponds to::1 m]]
*[[Corresponds to::0.001 km]]
*[[Corresponds to::10 dm]]
*[[Corresponds to::100 cm]]
*[[Corresponds to::1000 mm]]
```

14.3.1.3 Property mit Einheitenrechnung: LogiCO:hasLengthValue

This relation is used to represent [[imported from::LogiCO:hasLengthValue]].
{Unit conversion length}}

14.3.1.4 Property mit Wertbereichseinschränkung: LogiCO:hasLocation

This relation is used to represent [[imported from::LogiCO:hasLocation]].

This is the inverse of [[owl:inverseOf::Property:LogiCO:isLocationOf]].

The known locations for this property are:

- * [[Allows value::Archive Rütli]]
- * [[Allows value::Archive Küsnacht]]
- * [[Allows value::Office Zürich]]

14.3.1.5 Archive Rüti

The Archive in [[wikipedia:Rüti,_Zürich|Rüti]] is the main archival storage location of the [[:en:Swiss_Foundation_Public_Domain]]. It is located at [[LogiCO:hasAddress::Joweid Zentrum 1 8630 Rüti ZH Schweiz]] in an industrial building. I is currently the location of

```
{{#ask:[[LogiCO:hasLocation::{{FULLPAGENAME}}]]
|format=count
}}
```

boxes containing around 100 78rpm records each.

We maintain there a controlled climate according to the recommendations of IASA and Memoriav in this archive. For this purpose the temperature and humidity is monitored and logged.

== Boxes stored in this archive ==

```
{{#ask:[[LogiCO:hasLocation::{{FULLPAGENAME}}]]
| link=none
| format=ul
| template=Query output isLocationOf
| sort=LogiCO:id
| order=asc
| default=Currently this archive is empty
}}
```

```
[[Category:Location]]
[[Category:PD:Media archives]]
```

14.3.2 Template Storage box

<noinclude>

== HOW TO USE ==

```
<pre>{{Storage box
|ID= {{BASEPAGENAME}} Don't change this!
|Box type= Use one of these values: Euro, Rako, Carton 40cm, Open shelf 1m
|Archive= Name of the Wiki page where this box is located. Ex.: Archive Rüti
|Shelve= Number of the shelf
|Level= Level inside the shelf counting from bottom up (The lowest is 1)
|IsFull= yes or no
}}</pre>
```

== COPY AND PASTE ==

```
<pre>
{{Storage box
|ID= {{BASEPAGENAME}}
|Box type= Carton40cm
|Archive= PD:Archive Rüti
|Shelve=
|Level=
|IsFull= yes
}}</pre>
```

</noinclude>

<includeonly>

== Storage box location ==

```
<div class="hproduct">
<table class="toccolours vevent" style="width: 100%; direction: {{Dir|
{{int:lang}} }};" cellpadding="4">

<!-- ID -->
<tr style="vertical-align: top">
<td id="fileinfotpl_lbl" style="background: #BBB; text-align: right; pad-
ding-right: 0.4em; width: 15%; font-weight:bold">Box ID</td>
<td><span class="ID">{{ #if: {{{id|{{{ID}}}} }} | [[LogiCO:id::{{{id|
{{{ID}}}} }}}]] | }}</span></td>
</tr>

<!-- Box type -->
<tr style="vertical-align: top">
<td id="fileinfotpl_lbl" style="background: #BBB; text-align: right; pad-
ding-right: 0.4em; width: 15%; font-weight:bold">Box type</td>
<td><span class="Box type">{{ #if: {{{box type|{{{Box type}}}} }} | [[IsS-
torageBoxType::{{{box type|{{{Box type}}}} }}}]] | }}</span></td>
</tr>
<tr style="vertical-align: top">
<td id="fileinfotpl_lbl" style="background: #BBB; text-align: right; pad-
ding-right: 0.4em; width: 15%; font-weight:bold">Box size</td>
<td><span class="Box size">{{ #if: {{{box type|{{{Box type}}}} }} | {{Sto-
rage box size|{{{Box type}}}} }} | }}</span></td>
</tr>
<tr style="vertical-align: top">
<td id="fileinfotpl_lbl" style="background: #BBB; text-align: right; pad-
ding-right: 0.4em; width: 15%; font-weight:bold">Box weight</td>
<td><span class="Box size">{{ #if: {{{box type|{{{Box type}}}} }} | {{Sto-
rage box weight|{{{Box type}}}} }} | }}</span></td>
</tr>

<!-- Archive -->
<tr style="vertical-align: top">
<td id="fileinfotpl_add" style="background: #BBB; text-align: right; pad-
ding-right: 0.4em; width: 15%; font-weight:bold">Archive</td>
<td><span class="Archive">{{ #if: {{{archive|{{{Archive}}}} }} |
[[LogiCO:hasLocation::{{{archive|{{{Archive}}}} }}]|{{{archive|
{{{Archive}}}} }}}]] | }}</span></td>
</tr>

<!-- Shelve -->
<tr style="vertical-align: top">
<td id="fileinfotpl_add" style="background: #BBB; text-align: right; pad-
ding-right: 0.4em; width: 15%; font-weight:bold">Shelve</td>
<td><span class="Shelve">{{ #if: {{{shelve|{{{Shelve}}}} }} | [[inShelve::
{{{shelve|{{{Shelve}}}} }}}]] | }}</span></td>
</tr>

<!-- Level -->
<tr style="vertical-align: top">
<td id="fileinfotpl_add" style="background: #BBB; text-align: right; pad-
ding-right: 0.4em; width: 15%; font-weight:bold">Level</td>
<td><span class="Level">{{ #if: {{{level|{{{Level}}}} }} | [[onLevel::
{{{level|{{{Level}}}} }}}]] | }}</span></td>
```

```

</tr>

<!-- IsFull -->
<tr style="vertical-align: top">
<td id="fileinfotpl_add" style="background: #BBB; text-align: right; padding-right: 0.4em; width: 15%; font-weight:bold">Is it full?</td>
<td><span class="IsFull">{{ #if: {{{isfull|{{{IsFull|}}}} }} | [[LogiCO:isFull::{{{isfull|{{{IsFull|}}}} }}]] | }}</span></td>
</tr>
</table>
</div>
<!-- [[Category:{{{PAGENAME}}}] -->
</includeonly>

```

== Records stored in this box ==

```

{{#ask: [[LogiCO:isPackagedIn::{{{FULLPAGENAME}}}]
| link=subject
| format=ul
| order=asc
| default=This box is empty or has unknown content
}}

```

```

<noinclude>
[[Category:Templates]]
</noinclude>

```

14.3.3 Template Storage box size

<noinclude>
This template is used to define the size of an archival storage box. the value depends on the type. Currently RAKO, EURO and Carton40cm are known.

```

[[Category:Templates]]
</noinclude>
<includeonly>
{{#ifeq: {{{1}}} | RAKO | [[LogiCO:hasLengthValue::80 cm]] x [[LogiCO:hasWidthValue::40 cm]] x [[LogiCO:hasHeightValue::40 cm]] |}}
{{#ifeq: {{{1}}} | EURO | [[LogiCO:hasLengthValue::40 cm]] x [[LogiCO:hasWidthValue::40 cm]] x [[LogiCO:hasHeightValue::40 cm]] |}}
{{#ifeq: {{{1}}} | Carton 40cm | [[LogiCO:hasLengthValue::40 cm]] x [[LogiCO:hasWidthValue::40 cm]] x [[LogiCO:hasHeightValue::40 cm]] |}}
{{#ifeq: {{{1}}} | Open shelf 1m | [[LogiCO:hasLengthValue:: 100 cm]] x [[LogiCO:hasWidthValue:: 50 cm]] x [[LogiCO:hasHeightValue::220 cm]] |}}
(Length x Width x Height)

{{#ifeq: {{{1}}} | RAKO | [[Category:Box]]|}}
{{#ifeq: {{{1}}} | Euro | [[Category:Box]]|}}
{{#ifeq: {{{1}}} | Carton 40cm | [[Category:Carton]]|}}
{{#ifeq: {{{1}}} | Open shelf 1m | [[Category:Package]]|}}
</includeonly>

```

14.3.4 Concept Storage box

```

{{#concept: [[Category:Package]]
|All storage boxes that are used in our archives
}}

```

14.3.5 Formular Storage box

```
<noinclude>
```

This is the "Storage box" form.

To create a page with this form, enter the page name below;

if a page with that name already exists, you will be sent to a form to edit that page.

```
{{#forminput:form=Storage box}}
```

```
</noinclude><includeonly>
```

```
<div id="wikiPreview" style="display: none; padding-bottom: 25px; margin-bottom: 25px; border-bottom: 1px solid #AAAAAA;"></div>
```

```
{{{for template|Storage box}}}
```

```
{| class="formtable"
```

```
!
```

```
| {{{field|ID|hidden|default={{BASEPAGENAME}} }}}
```

```
|-
```

```
! Box type:
```

```
| {{{field|Box type|input type=radio|mandatory|property=IsStorageBox-Type|default=Carton40cm}}}
```

```
|-
```

```
! Archive:
```

```
| {{{field|Archive|input type=radio|mandatory|property=LogiC0:hasLocation|default=Archive Rüti}}}
```

```
|-
```

```
! Shelve:
```

```
| {{{field|Shelve}}}
```

```
|-
```

```
! Level:
```

```
| {{{field|Level|input type=regexp|regexp=/^[1-9]{0,1}$/|message=A number between 1 and 9}}}
```

```
|-
```

```
! IsFull:
```

```
| {{{field|IsFull|input type=radio|mandatory|property=LogiC0:isFull|default=true}}}
```

```
|}
```

```
{{{end template}}}
```

```
'''Free text:'''
```

```
{{{standard input|free text|rows=10}}}
```

```
{{{standard input|summary}}}
```

```
{{{standard input|minor edit}}} {{{standard input|watch}}}
```

```
{{{standard input|save}}} {{{standard input|preview}}} {{{standard input|changes}}} {{{standard input|cancel}}}
```

```
</includeonly>
```

14.4 Beispieldaten

14.4.1 Beispieldaten wie sie bisher menschenlesbar im Browser dargestellt werden **Audio file**

Image(s) see below (if available)



Label Busy-Bee Records
Cat. no. D47
Order number 3259
Matrix/StamperID 3259-7-22
1st release date 1907
1st recording date 1907
Coupling date 1907
Cutout date unknown
Place of recording Chicago, Illinois (United States)
Description Busy-Bee D47
Author(s)/Composer(s) David Wallis Reeves (1838-1900)
Lyricist(s) none
Music arranger(s) none
Conductor(s) none
Performer(s) Unknown artist (band)
Vocal range Instrumental
Title/Work Second Regiment Connecticut National Guard March (1888)
Content none
Genre(s) March music

FLAC FLAC (http://pool.publicdomainproject.org/audio/flac/genres/march_music/unknown_artist/busy-bee-d47-3259.flac)

Ogg (Vorbis/Theora) Ogg (http://pool.publicdomainproject.org/audio/ogg/genres/march_music/unknown_artist/busy-bee-d47-3259.ogg) · Ogg (Commons)

PD CH 1 January 1971

PD EU 1 January 1971

PD USA 1 January 1971

PD INT 1 January 2001

References

Licensing

*This work is in the **public domain** because its copyright has **expired**.*

*This applies **worldwide**.*



Retrieved from "<http://pool.publicdomainproject.org/index.php?title=Busy-bee-d47-3259&oldid=30096>"

Categories: Busy-Bee Records | 1900s music | March music | David Wallis Reeves
| Unknown artist | Second Regiment Connecticut National Guard March
| PD CH 1971 | PD EU 1971 | PD USA 1971 | PD INT 2001 | PD:RAKO-SD25-00005
| FLAC sound files | Ogg sound files | Audio file licenses | Carl Flisch collection


```

&lt;/div&gt;
&lt;/div&gt;

=={{int:license-header}}==
{{PD-INT}}

[[Category:Busy-Bee Records]] &lt;!- Label information
--&gt;
[[Category:1900s music]] &lt;!- Music decade (1st release
date is important) --&gt;
[[Category:March music]] &lt;!- Genre --&gt;
[[Category:David Wallis Reeves]] &lt;!- Composer --&gt;
[[Category:Unknown artist]] &lt;!- Performer --&gt;
[[Category:Second Regiment Connecticut National Guard March]]
&lt;!- Musical composition --&gt;
[[Category:PD CH 1971]] &lt;!- Public Domain --&gt;
[[Category:PD EU 1971]] &lt;!- Public Domain --&gt;
[[Category:PD USA 1971]] &lt;!- Public Domain --&gt;
[[Category:PD INT 2001]] &lt;!- Public Domain --&gt;
[[Category:PD:RAK0-SD25-00005]] &lt;!- Archive container
--&gt;
[[Category:FLAC sound files]] &lt;!- FLAC --&gt;
[[Category:Ogg sound files]] &lt;!- Ogg --&gt;
[[Category:Audio file licenses]]
[[Category:Carl Flisch collection]] &lt;!- License and des-
cription file --&gt;
</rev>
</revisions>
</page>
</pages>
</query>
</api>

```

14.4.3 RDF/XML Beispieldaten eines Lieds vor der Masterthesis

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
  <!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-schema#'>
  <!ENTITY owl 'http://www.w3.org/2002/07/owl#'>
  <!ENTITY swivt 'http://semantic-mediawiki.org/swivt/1.0#'>
  <!ENTITY wiki 'http://publicdomainproject.org/index.php/Special:URIRe-
solver/'>
  <!ENTITY category 'http://publicdomainproject.org/index.php/Special:URI-
Resolver/Category-3A'>
  <!ENTITY property 'http://publicdomainproject.org/index.php/Special:URI-
Resolver/Property-3A'>
  <!ENTITY wikiurl 'http://pool.publicdomainproject.org/index.php/'>
]>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:swivt="http://semantic-mediawiki.org/swivt/1.0#"
  xmlns:wiki="http://publicdomainproject.org/index.php/Special:URIResol-
ver/"

```

```

    xmlns:category="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3A"
    xmlns:property="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3A">

    <owl:Ontology
rdf:about="http://pool.publicdomainproject.org/index.php/Special:ExportRDF/Busy-bee-d47-3259">
        <swivt:creationDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2016-11-16T21:38:28+01:00</swivt:creationDate>
        <owl:imports rdf:resource="http://semantic-mediawiki.org/swivt/1.0"/>
    </owl:Ontology>
    <swivt:Subject rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Busy-2Dbee-2Dd47-2D3259">
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3ABusy-2DBee_Records"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3A1900s_music"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3AMarch_music"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3ADavid_Wallis_Reeves"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3AUnknown_artist"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3ASecond_Regiment_Connecticut_National_Guard_March"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3APD_CH_1971"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3APD_EU_1971"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3APD_USA_1971"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3APD_INT_2001"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3APD-3ARAKO-2DSD25-2D00005"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3AFLAC_sound_files"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3Aogg_sound_files"/>
            <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3AAudio_file_licenses"/>

```



```

    <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3ACarl_Flisch_collection"/>
    <rdfs:label>Busy-bee-d47-3259</rdfs:label>
    <swivt:page rdf:resource="http://pool.publicdomainproject.org/in-
dex.php/Busy-bee-d47-3259"/>
    <rdfs:isDefinedBy rdf:resource="http://pool.publicdomainprojec-
t.org/index.php/Special:ExportRDF/Busy-bee-d47-3259"/>
    <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSche-
ma#integer">0</swivt:wikiNamespace>
    <swivt:wikiPageModificationDate
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2016-04-
30T18:48:35Z</swivt:wikiPageModificationDate>
    <property:Modification_date-23aux
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2457509.2837384</pro-
perty:Modification_date-23aux>
    <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XML-
Schema#string">Busy-bee-d47-3259</swivt:wikiPageSortKey>
    </swivt:Subject>
    <owl:DatatypeProperty rdf:about="http://semantic-
mediawiki.org/swivt/1.0#creationDate"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3ABusy-2DBee_Records"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3A1900s_music"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3AMarch_music"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3ADavid_Wallis_Reeves"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3AUnknown_artist"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3ASecond_Regiment_Connecticut_National_Guard_March"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3APD_CH_1971"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3APD_EU_1971"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3APD_USA_1971"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3APD_INT_2001"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3APD-3ARAK0-2DSD25-2D00005"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegoriy-3AFLAC_sound_files"/>

```

```

    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3A0gg_sound_files"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3AAudio_file_licenses"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3ACarl_Flisch_collection"/>
    <owl:ObjectProperty rdf:about="http://semantic-mediawiki.org/swivt/1.0#page"/>
    <owl:DatatypeProperty rdf:about="http://semantic-mediawiki.org/swivt/1.0#wikiNamespace"/>
    <owl:DatatypeProperty rdf:about="http://semantic-mediawiki.org/swivt/1.0#wikiPageModificationDate"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AModification_date-23aux"/>
    <owl:DatatypeProperty rdf:about="http://semantic-mediawiki.org/swivt/1.0#wikiPageSortKey"/>
    <!-- Created by Semantic MediaWiki, https://www.semantic-mediawiki.org/ -->
</rdf:RDF>

```

14.4.4 RDF Beispieldaten eines Lieds nach der Masterthesis

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rdf:RDF [
    <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
    <!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-schema#'>
    <!ENTITY owl 'http://www.w3.org/2002/07/owl#'>
    <!ENTITY swivt 'http://semantic-mediawiki.org/swivt/1.0#'>
    <!ENTITY wiki 'http://publicdomainproject.org/index.php/Special:URIResolver/'>
    <!ENTITY category 'http://publicdomainproject.org/index.php/Special:URIResolver/Category-3A'>
    <!ENTITY property 'http://publicdomainproject.org/index.php/Special:URIResolver/Property-3A'>
    <!ENTITY wikiurl '//pool.publicdomainproject.org/index.php/'>
]>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:swivt="http://semantic-mediawiki.org/swivt/1.0#" xmlns:wiki="http://publicdomainproject.org/index.php/Special:URIResolver/" xmlns:category="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3A" xmlns:property="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3A"
xmlns:cc="http://creativecommons.org/ns#"
xmlns:dbpedia="http://dbpedia.org/ontology"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:mo="http://purl.org/ontology/mo/" xmlns:Schema="http://schema.org">

    <owl:Ontology rdf:about="//pool.publicdomainproject.org/index.php/Special:ExportRDF/Polydor-67575a-859ge9_demo">
        <swivt:creationDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2017-01-12T18:51:40+01:00</swivt:creationDate>

```

```

        <owl:imports rdf:resource="http://semantic-
mediawiki.org/swikt/1.0"/>
    </owl:Ontology>
    <swikt:Subject rdf:about="http://publicdomainproject.org/index.php/Spe-
cial:URIResolver/Polydor-2D67575a-2D859ge9_demo">
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3AFranz_Schubert"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3ASymphony"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3APaul_van_Kempen"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3APolydor_Records"/>
        <rdf:type rdf:resource="http://schema.org/MusicRecording"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3ADigitized_audio_track"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3AFLAC_sound_files"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3ASiemens-2DPolydor_album_67575-2D67577"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3A1940s_music"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3ASymphonies"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3ADresden_Philharmonic"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3ASymphony_No._8_-28Schubert-29"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3APD_CH_1992"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3APD_EU_2012"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3APD_USA_1992"/>
        <rdf:type
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/C
ategory-3APD_INT_2012"/>
        <rdfs:label>Polydor-67575a-859ge9_demo</rdfs:label>
        <swikt:page rdf:resource="//pool.publicdomainproject.org/index.-
php/Polydor-67575a-859ge9_demo"/>
        <rdfs:isDefinedBy rdf:resource="//pool.publicdomainproject.org/in-
dex.php/Special:ExportRDF/Polydor-67575a-859ge9_demo"/>

```

```

        <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">0</swivt:wikiNamespace>
        <cc:permits
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Reproduction</cc:per-
mits>
        <cc:permits
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Distribution</cc:per-
mits>
        <cc:permits
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Derivative
Works</cc:permits>
        <property:ComposerHasMoreInformation rdf:resource="http://en.wiki-
pedia.org/wiki/Franz_Schubert"/>
        <property:ComposerHasMusicBrainzID
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">f91e3a88-24ee-4563-
8963-fab73d2765ed</property:ComposerHasMusicBrainzID>
        <property:ComposerYearBorn
rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1797</property:Compo-
serYearBorn>
        <property:ComposerYearBorn-23aux
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2377401.5</property:C
omposerYearBorn-23aux>
        <property:ComposerYearDied
rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1828</property:Compo-
serYearDied>
        <property:ComposerYearDied-23aux
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2388722.5</property:C
omposerYearDied-23aux>
        <property:ConductedBy rdf:datatype="http://www.w3.org/2001/XML-
Schema#string">Paul van Kempen</property:ConductedBy>
        <property:ConductorHasMoreInformation rdf:resource="http://en.wi-
kipedia.org/wiki/Paul_van_Kempen"/>
        <property:ConductorHasMusicBrainzID
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">4b393889-c76d-47d-
f-add1-c4ab007f5e6e</property:ConductorHasMusicBrainzID>
        <property:ConductorYearBorn
rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1893</property:Conduc-
torYearBorn>
        <property:ConductorYearBorn-23aux
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2412464.5</property:C
onductorYearBorn-23aux>
        <property:ConductorYearDied
rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1955</property:Conduc-
torYearDied>
        <property:ConductorYearDied-23aux
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2435108.5</property:C
onductorYearDied-23aux>
        <dbpedia:composer
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Franz Schubert</dbpe-
dia:composer>
        <dc:contributor
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Paul van
Kempen</dc:contributor>
        <dc:contributor
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Dresden
Philharmonic</dc:contributor>

```

```

    <dc:creator
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Franz
Schubert</dc:creator>
    <dc:description
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Part 1: 1st movement:
Allegro moderato (part 1), Instrumental</dc:description>
    <dc:description
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">http://en.wikipe-
dia.org/wiki/Symphony_No._8_(Schubert)</dc:description>
    <dc:publisher
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Polydor
Records</dc:publisher>
    <dc:title
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Symphony No. 8 in B
minor, 1st movement: Allegro moderato</dc:title>
    <property:Dcterms-3Acreated
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1940-05-
25Z</property:Dcterms-3Acreated>
    <property:Dcterms-3Acreated-23aux
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2429774.5</property:D
cterms-3Acreated-23aux>
    <dcterms:hasFormat rdf:resource="http://pool.publicdomainprojec-
t.org/audio/flac/genres/classical/symphonic/dresden_philharmonic/poly-
dor-67575a-859ge9.flac"/>
    <dcterms:hasFormat rdf:resource="http://pool.publicdomainprojec-
t.org/audio/ogg/genres/classical/symphonic/dresden_philharmonic/poly-
dor-67575a-859ge9.ogg"/>
    <dcterms:isPartOf rdf:resource="http://publicdomainproject.org/in-
dex.php/Special:URIResolver/Polydor_67575_demo"/>
    <dcterms:issued
rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1941</dcterms:issued>
    <dcterms:issued
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2429995.5</dcterms:is
sued>
    <dcterms:license rdf:resource="http://creativecommons.org/public-
domain/mark/1.0/">
    <dcterms:references rdf:datatype="http://www.w3.org/2001/XMLSche-
ma#string">Coupling date is printed on record,
[http://www.charm.rhul.ac.uk/discography/disco.html CHARM]</dcterms:refe-
rences>
    <property:FlacFileLink rdf:resource="http://pool.publicdomainpro-
ject.org/audio/flac/genres/classical/symphonic/dresden_philharmonic/polydor-
67575a-859ge9.flac"/>
    <foaf:depiction
rdf:resource="http://pool.publicdomainproject.org/images/f/f4/Poly-
dor-67575a-859ge9.jpg"/>
    <property:IsOnRecord
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/P
olydor_67575_demo"/>
    <mo:available_as rdf:resource="http://pool.publicdomainprojec-
t.org/audio/flac/genres/classical/symphonic/dresden_philharmonic/poly-
dor-67575a-859ge9.flac"/>
    <mo:available_as rdf:resource="http://pool.publicdomainprojec-
t.org/audio/ogg/genres/classical/symphonic/dresden_philharmonic/poly-
dor-67575a-859ge9.ogg"/>
    <mo:record_side
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">A</mo:record_side>

```

```

    <property:OggFileLink rdf:datatype="http://www.w3.org/2001/XMLSchema#string">http://pool.publicdomainproject.org/audio/ogg/genres/classical/symphonic/dresden_philharmonic/polydor-67575a-859ge9.ogg</property:OggFileLink>
    <property:Pdch
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1992-01-01Z</property:Pdch>
    <property:Pdch-23aux rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2448622.5</property:Pdch-23aux>
    <property:Pdeu
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-01-01Z</property:Pdeu>
    <property:Pdeu-23aux rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2455927.5</property:Pdeu-23aux>
    <property:Pdint
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-01-01Z</property:Pdint>
    <property:Pdint-23aux rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2455927.5</property:Pdint-23aux>
    <property:Pdusa
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1992-01-01Z</property:Pdusa>
    <property:Pdusa-23aux rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2448622.5</property:Pdusa-23aux>
    <property:PerformerHadFunction
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Orchestra</property:PerformerHadFunction>
    <property:PerformerHasMoreInformation rdf:resource="http://en.wikipedia.org/wiki/Dresden_Philharmonic"/>
    <property:PerformerHasMusicBrainzID
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">6fc009d4-6f05-4ffa-8151-a560508f6f4c</property:PerformerHasMusicBrainzID>
    <Schema:audio
rdf:resource="http://pool.publicdomainproject.org/audio/flac/genres/classical/symphonic/dresden_philharmonic/polydor-67575a-859ge9.flac"/>
    <Schema:audio
rdf:resource="http://pool.publicdomainproject.org/audio/ogg/genres/classical/symphonic/dresden_philharmonic/polydor-67575a-859ge9.ogg"/>
    <Schema:dateCreated rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1940-05-25Z</Schema:dateCreated>
    <Schema:dateCreated rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2429774.5</Schema:dateCreated>
    <Schema:funder rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3ACarl_Flisch_collection"/>
    <Schema:genre
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Symphony</Schema:genre>
    <Schema:image
rdf:resource="http://pool.publicdomainproject.org/images/f/f4/Polydor-67575a-859ge9.jpg"/>
    <Schema:isAccessibleForFree
rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">>true</Schema:isAccessibleForFree>
    <Schema:license rdf:resource="http://creativecommons.org/publicdomain/mark/1.0/">
    <Schema:locationCreated rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Dresden (German Reich)</Schema:locationCreated>

```

```

    <Schema:thumbnailUrl rdf:resource="http://pool.publicdomainproject.org/images/thumb/f/f4/Polydor-67575a-859ge9.jpg/200px-Polydor-67575a-859ge9.jpg"/>
    <property:StamperID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">67575 A-M, 859<sup>&lt;&lt;/sup>&lt;/sup>&lt;/property:StamperID>
    <property:TitleLink
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/wikipedia-3ASymphony_No._8_(Schubert)"/>
    <swivt:wikiPageModificationDate
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2017-01-07T21:18:48Z</swivt:wikiPageModificationDate>
    <property:Modification_date-23aux
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2457761.3880556</property:Modification_date-23aux>
    <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Polydor-67575a-859ge9 demo</swivt:wikiPageSortKey>
    </swivt:Subject>
    <swivt:Subject rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Polydor_67575_demo">
    <rdfs:label>Polydor 67575 demo</rdfs:label>
    <swivt:page rdf:resource="//pool.publicdomainproject.org/index.php/Polydor_67575_demo"/>
    <rdfs:isDefinedBy rdf:resource="//pool.publicdomainproject.org/index.php/Special:ExportRDF/Polydor_67575_demo"/>
    <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">0</swivt:wikiNamespace>
    <mo:track
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/Polydor-2D67575a-2D859ge9_demo"/>
    <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Polydor 67575 demo</swivt:wikiPageSortKey>
    </swivt:Subject>
    <owl:ObjectProperty rdf:about="http://purl.org/ontology/mo/track">
    <rdfs:label>Mo:track</rdfs:label>
    <swivt:page rdf:resource="//pool.publicdomainproject.org/index.php/Property:Mo:track"/>
    <rdfs:isDefinedBy rdf:resource="//pool.publicdomainproject.org/index.php/Special:ExportRDF/Property:Mo:track"/>
    <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">110</swivt:wikiNamespace>
    <swivt:specialImportedFrom
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">mo track http://purl.org/ontology/mo/</swivt:specialImportedFrom>
    <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Mo:track</swivt:wikiPageSortKey>
    <swivt:type rdf:resource="http://semantic-mediawiki.org/swivt/1.0#_wpg"/>
    </owl:ObjectProperty>
    <owl:DatatypeProperty rdf:about="http://semantic-mediawiki.org/swivt/1.0#creationDate"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3AFranz_Schubert"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Category-3ASymphony"/>

```

```

    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3APaul_van_Kempen"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3APolydor_Records"/>
        <owl:Class rdf:about="http://schema.org/MusicRecording"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3ADigitized_audio_track"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3AFLAC_sound_files"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3ASiemens-2DPolydor_album_67575-2D67577"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3A1940s_music"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3ASymphonies"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3ADresden_Philharmonic"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3ASymphony_No._8_-28Schubert-29"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3APD_CH_1992"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3APD_EU_2012"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3APD_USA_1992"/>
    <owl:Class
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Ca-
tegor-y-3APD_INT_2012"/>
        <owl:ObjectProperty rdf:about="http://semantic-
mediawiki.org/swivt/1.0#page"/>
        <owl:DatatypeProperty rdf:about="http://semantic-
mediawiki.org/swivt/1.0#wikiNamespace"/>
        <owl:DatatypeProperty
rdf:about="http://creativecommons.org/ns#permits"/>
        <owl:ObjectProperty
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Pro-
perty-3AComposerHasMoreInformation"/>
            <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AComposerHasMusicBrainzID"/>
            <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AComposerYearBorn"/>
            <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AComposerYearBorn-23aux"/>
            <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AComposerYearDied"/>

```



```

    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AComposerYearDied-23aux"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AConductedBy"/>
    <owl:ObjectProperty
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AConductorHasMoreInformation"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AConductorHasMusicBrainzID"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AConductorYearBorn"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AConductorYearBorn-23aux"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AConductorYearDied"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AConductorYearDied-23aux"/>
    <owl:DatatypeProperty rdf:about="http://dbpedia.org/ontology/composer"/>
    <owl:DatatypeProperty rdf:about="http://purl.org/dc/elements/1.1/contributor"/>
    <owl:DatatypeProperty rdf:about="http://purl.org/dc/elements/1.1/creator"/>
    <owl:DatatypeProperty rdf:about="http://purl.org/dc/elements/1.1/description"/>
    <owl:DatatypeProperty rdf:about="http://purl.org/dc/elements/1.1/publisher"/>
    <owl:DatatypeProperty
rdf:about="http://purl.org/dc/elements/1.1/title"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3ADcTerms-3Acreated"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3ADcTerms-3Acreated-23aux"/>
    <owl:ObjectProperty rdf:about="http://purl.org/dc/terms/hasFormat"/>
    <owl:ObjectProperty rdf:about="http://purl.org/dc/terms/isPartOf"/>
    <owl:DatatypeProperty rdf:about="http://purl.org/dc/terms/issued"/>
    <owl:ObjectProperty rdf:about="http://purl.org/dc/terms/license"/>
    <owl:DatatypeProperty rdf:about="http://purl.org/dc/terms/references"/>
    <owl:ObjectProperty
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AFlacFileLink"/>
    <owl:ObjectProperty rdf:about="http://xmlns.com/foaf/0.1/depiction"/>
    <owl:ObjectProperty
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AIsOnRecord"/>
    <owl:ObjectProperty
rdf:about="http://purl.org/ontology/mo/available_as"/>
    <owl:DatatypeProperty
rdf:about="http://purl.org/ontology/mo/record_side"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AoggFileLink"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3APdch"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3APdch-23aux"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3APdeu"/>

```

```

    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3APdeu-23aux"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3APdint"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3APdint-23aux"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3APdusa"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3APdusa-23aux"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3APerformerHadFunction"/>
    <owl:ObjectProperty
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Pro-
perty-3APerformerHasMoreInformation"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3APerformerHasMusicBrainzID"/>
    <owl:ObjectProperty rdf:about="http://schema.orgaudio"/>
    <owl:DatatypeProperty rdf:about="http://schema.orgdateCreated"/>
    <owl:ObjectProperty rdf:about="http://schema.orgfunder"/>
    <owl:DatatypeProperty rdf:about="http://schema.orggenre"/>
    <owl:ObjectProperty rdf:about="http://schema.orgimage"/>
    <owl:DatatypeProperty rdf:about="http://schema.orgisAccessibleForFree"/>
    <owl:ObjectProperty rdf:about="http://schema.orglicense"/>
    <owl:DatatypeProperty rdf:about="http://schema.orglocationCreated"/>
    <owl:ObjectProperty rdf:about="http://schema.orgthumbnailUrl"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AStamperID"/>
    <owl:ObjectProperty
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Pro-
perty-3ATitleLink"/>
    <owl:DatatypeProperty rdf:about="http://semantic-
mediawiki.org/swikt/1.0#wikiPageModificationDate"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AModification_date-23aux"/>
    <owl:DatatypeProperty rdf:about="http://semantic-
mediawiki.org/swikt/1.0#wikiPageSortKey"/>
    <owl:DatatypeProperty rdf:about="http://semantic-
mediawiki.org/swikt/1.0#specialImportedFrom"/>
    <owl:ObjectProperty rdf:about="http://semantic-
mediawiki.org/swikt/1.0#type"/>
    <!-- Created by Semantic MediaWiki, https://www.semantic-mediawiki.org/
-->
</rdf:RDF>

```

14.4.5 RDF Beispieldaten eines Lagerbehälters

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
  <!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-schema#'>
  <!ENTITY owl 'http://www.w3.org/2002/07/owl#'>
  <!ENTITY swivt 'http://semantic-mediawiki.org/swivt/1.0#'>
  <!ENTITY wiki 'http://publicdomainproject.org/index.php/Special:URIRe-
solver/'>
  <!ENTITY category 'http://publicdomainproject.org/index.php/Special:URI-
Resolver/Category-3A'>
  <!ENTITY property 'http://publicdomainproject.org/index.php/Special:URI-
Resolver/Property-3A'>
  <!ENTITY wikiurl '//pool.publicdomainproject.org/index.php/'>
]>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://ww-
w.w3.org/2002/07/owl#" xmlns:swivt="http://semantic-
mediawiki.org/swivt/1.0#" xmlns:wiki="http://publicdomainproject.org/index.-
php/Special:URIResolver/" xmlns:category="http://publicdomainproject.org/in-
dex.php/Special:URIResolver/Category-3A" xmlns:property="http://publicdo-
mainproject.org/index.php/Special:URIResolver/Property-3A"
xmlns:LogiC0="http://ontology.tno.nl/logico#">

  <owl:Ontology rdf:about="//pool.publicdomainproject.org/index.php/Speci-
al:ExportRDF/EURO-SD25-00008_demo">
    <swivt:creationDate rdf:datatype="http://www.w3.org/2001/XMLSche-
ma#dateTime">2017-01-12T18:31:58+01:00</swivt:creationDate>
    <owl:imports rdf:resource="http://semantic-
mediawiki.org/swivt/1.0"/>
  </owl:Ontology>
  <swivt:Subject rdf:about="http://publicdomainproject.org/index.php/Spe-
cial:URIResolver/EURO-2DSD25-2D00008_demo">
    <rdf:type rdf:resource="http://ontology.tno.nl/logico#Carton"/>
    <rdfs:label>EURO-SD25-00008 demo</rdfs:label>
    <swivt:page rdf:resource="//pool.publicdomainproject.org/index.-
php/EURO-SD25-00008_demo"/>
    <rdfs:isDefinedBy rdf:resource="//pool.publicdomainproject.org/in-
dex.php/Special:ExportRDF/EURO-SD25-00008_demo"/>
    <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSche-
ma#integer">0</swivt:wikiNamespace>
    <property:InShelve
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">5</property:InShelve>
    <property:IsStorageBoxType
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Carton 40cm</proper-
ty:IsStorageBoxType>
    <LogiC0:hasHeightValue rdf:datatype="http://www.w3.org/2001/XML-
Schema#double">0.4</LogiC0:hasHeightValue>
    <LogiC0:hasLengthValue rdf:datatype="http://www.w3.org/2001/XML-
Schema#double">0.4</LogiC0:hasLengthValue>
    <LogiC0:hasLocation
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/A
rchive_R-C3-BCti"/>
    <LogiC0:hasWeightValue rdf:datatype="http://www.w3.org/2001/XML-
Schema#double">35</LogiC0:hasWeightValue>
```

```

    <LogiCO:hasWidthValue rdf:datatype="http://www.w3.org/2001/XMLSchema#double">0.4</LogiCO:hasWidthValue>
    <LogiCO:id
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">EURO-SD25-00008
demo</LogiCO:id>
    <LogiCO:isFull rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">true</LogiCO:isFull>
    <property:OnLevel
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2</property:OnLevel>
    <property:Has_query
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/EURO-2DSD25-2D00008_demo-23_QUERY6095b5d83790f5f60d44b64fd8a129eb"/>
    <swivt:wikiPageModificationDate
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2017-01-06T17:28:45Z</swivt:wikiPageModificationDate>
    <property:Modification_date-23aux
rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2457760.2282986</property:Modification_date-23aux>
    <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">EURO-SD25-00008 demo</swivt:wikiPageSortKey>
    </swivt:Subject>
    <swivt:Subject rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/EURO-2DSD25-2D00008_demo-23_QUERY6095b5d83790f5f60d44b64fd8a129eb">
    <swivt:masterPage rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/EURO-2DSD25-2D00008_demo"/>
    <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">0</swivt:wikiNamespace>
    <property:Query_depth rdf:datatype="http://www.w3.org/2001/XMLSchema#double">1</property:Query_depth>
    <property:Query_format rdf:datatype="http://www.w3.org/2001/XMLSchema#string">ul</property:Query_format>
    <property:Query_size rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2</property:Query_size>
    <property:Query_string rdf:datatype="http://www.w3.org/2001/XMLSchema#string">[[LogiCO:isPackagedIn::EURO-SD25-00008 demo]]</property:Query_string>
    <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">EURO-SD25-00008 demo</swivt:wikiPageSortKey>
    </swivt:Subject>
    <swivt:Subject rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Polydor_67575_demo">
    <rdfs:label>Polydor 67575 demo</rdfs:label>
    <swivt:page rdf:resource="//pool.publicdomainproject.org/index.php/Polydor_67575_demo"/>
    <rdfs:isDefinedBy rdf:resource="//pool.publicdomainproject.org/index.php/Special:ExportRDF/Polydor_67575_demo"/>
    <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">0</swivt:wikiNamespace>
    <LogiCO:isPackagedIn
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/EURO-2DSD25-2D00008_demo"/>
    <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Polydor 67575 demo</swivt:wikiPageSortKey>
    </swivt:Subject>
    <swivt:Subject rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Polydor_67575-2D67577_demo">

```

```

        <rdfs:label>Polydor 67575-67577 demo</rdfs:label>
        <swivt:page rdf:resource="//pool.publicdomainproject.org/index.php/Polydor_67575-67577_demo"/>
        <rdfs:isDefinedBy rdf:resource="//pool.publicdomainproject.org/index.php/Special:ExportRDF/Polydor_67575-67577_demo"/>
        <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">0</swivt:wikiNamespace>
        <LogiC0:isPackagedIn
rdf:resource="http://publicdomainproject.org/index.php/Special:URIResolver/EURO-2DSD25-2D00008_demo"/>
        <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Polydor 67575-67577 demo</swivt:wikiPageSortKey>
        </swivt:Subject>
        <owl:ObjectProperty rdf:about="http://ontology.tno.nl/logico#isPackagedIn">
        <rdfs:label>LogiC0:isPackagedIn</rdfs:label>
        <swivt:page rdf:resource="//pool.publicdomainproject.org/index.php/Property:LogiC0:isPackagedIn"/>
        <rdfs:isDefinedBy rdf:resource="//pool.publicdomainproject.org/index.php/Special:ExportRDF/Property:LogiC0:isPackagedIn"/>
        <swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">110</swivt:wikiNamespace>
        <swivt:specialImportedFrom
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">LogiC0 isPackagedIn
http://ontology.tno.nl/logico#</swivt:specialImportedFrom>
        <swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">LogiC0:isPackagedIn</swivt:wikiPageSortKey>
        <swivt:type rdf:resource="http://semantic-mediawiki.org/swivt/1.0#_wpg"/>
        </owl:ObjectProperty>
        <owl:DatatypeProperty rdf:about="http://semantic-mediawiki.org/swivt/1.0#creationDate"/>
        <owl:Class rdf:about="http://ontology.tno.nl/logico#Carton"/>
        <owl:ObjectProperty rdf:about="http://semantic-mediawiki.org/swivt/1.0#page"/>
        <owl:DatatypeProperty rdf:about="http://semantic-mediawiki.org/swivt/1.0#wikiNamespace"/>
        <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AInShelve"/>
        <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AIsStorageBoxType"/>
        <owl:DatatypeProperty rdf:about="http://ontology.tno.nl/logico#hasHeightValue"/>
        <owl:DatatypeProperty rdf:about="http://ontology.tno.nl/logico#hasLengthValue"/>
        <owl:ObjectProperty rdf:about="http://ontology.tno.nl/logico#hasLocation"/>
        <owl:DatatypeProperty rdf:about="http://ontology.tno.nl/logico#hasWeightValue"/>
        <owl:DatatypeProperty rdf:about="http://ontology.tno.nl/logico#hasWidthValue"/>
        <owl:DatatypeProperty rdf:about="http://ontology.tno.nl/logico#id"/>
        <owl:DatatypeProperty rdf:about="http://ontology.tno.nl/logico#isFull"/>
        <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AOnLevel"/>

```

```

    <owl:ObjectProperty
rdf:about="http://publicdomainproject.org/index.php/Special:URIResolver/Property-3AHas_query"/>
    <owl:DatatypeProperty rdf:about="http://semantic-
mediawiki.org/swikt/1.0#wikiPageModificationDate"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AModification_date-23aux"/>
    <owl:DatatypeProperty rdf:about="http://semantic-
mediawiki.org/swikt/1.0#wikiPageSortKey"/>
    <owl:ObjectProperty rdf:about="http://semantic-
mediawiki.org/swikt/1.0#masterPage"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AQuery_depth"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AQuery_format"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AQuery_size"/>
    <owl:DatatypeProperty rdf:about="http://publicdomainproject.org/index.-
php/Special:URIResolver/Property-3AQuery_string"/>
    <owl:DatatypeProperty rdf:about="http://semantic-
mediawiki.org/swikt/1.0#specialImportedFrom"/>
    <owl:ObjectProperty rdf:about="http://semantic-
mediawiki.org/swikt/1.0#type"/>
    <!-- Created by Semantic MediaWiki, https://www.semantic-mediawiki.org/
-->
</rdf:RDF>

```

14.5 Listen der erstellten oder geänderten MediaWiki Seiten

Die Listen in diesem Anhang dienen vor allem der Migration vom Testsystem auf das Produktionssystem im Anschluss an diese Masterthesis.

Neu erstellte Seiten können mit der Import/Export Funktion von MediaWiki migriert werden. Geänderte Seiten können mit dem Werkzeug WikilImporter⁸⁸ migriert werden.

14.5.1.1 Archive

Office Zürich

Archive Küsnacht

Archive Rüti

14.5.1.2 Demoseiten

EURO-SD25-00008 demo

Polydor-67575a-859ge9 demo

Polydor 67575-67577 demo

Polydor 67575 demo

14.5.1.3 Liste der importierten Ontologien

Smw import Schema

Smw import mo

Smw import dc

Smw import dcterms

Smw import cc

Smw import LogiCO

Smw import owl

14.5.2 Liste der erstellten Properties

Cc:attributionName

Cc:attributionURL

Cc:deprecatedOn

Cc:jurisdiction

Cc:legalcode

Cc:morePermissions

Cc:permits

Cc:prohibits

Cc:requires

Cc:useGuidelines

ComposerHasMoreInformation

ComposerHasMusicBrainzID

ComposerYearBorn

ComposerYearDied

ConditionReport

ConductedBy

⁸⁸ <https://www.mediawiki.org/wiki/WikilImporter>

Conductor
ConductorHasMoreInformation
ConductorHasMusicBrainzID
ConductorYearBorn
ConductorYearDied
CouplingDate
CoverArtistHasMoreInformation
CoverArtistYearBorn
CoverArtistYearDied
CutoutDate
Dbo:composer
Dbo:coverArtist
Dbo:lyrics
Dc:contributor
Dc:coverage
Dc:creator
Dc:date
Dc:description
Dc:format
Dc:language
Dc:publisher
Dc:relation
Dc:rights
Dc:source
Dc:subject
Dc:title
Dc:type
Dcterms:bibliographicCitation
Dcterms:created
Dcterms:dateCopyrighted
Dcterms:format
Dcterms:hasFormat
Dcterms:hasPart
Dcterms:hasVersion
Dcterms:isFormatOf
Dcterms:isPartOf
Dcterms:issued
Dcterms:license
Dcterms:medium
Dcterms:references
DigitizingNotes
FlacFileLink
Foaf:depiction
InShelve
IsInCollection
IsOnRecord
IsStorageBoxType
ItemNotes
LicenseNotes
LogiCO:contains

LogiCO:hasAddress
LogiCO:hasCapacity
LogiCO:hasHeightValue
LogiCO:hasLengthValue
LogiCO:hasLocation
LogiCO:hasWeightValue
LogiCO:hasWidthValue
LogiCO:id
LogiCO:isFull
LogiCO:isLocationOf
LogiCO:isPackagedIn
LyricistHasMusicBrainzID
LyricistYearBorn
LyricistYearDied
MediumConditionReport
Mo:available as
Mo:musicbrainz
Mo:musicbrainz guid
Mo:record
Mo:record side
Mo:track
Mo:track number
MusicArrangerHasMusicBrainzID
MusicArrangerYearBorn
MusicArrangerYearDied
Notes
OggFileLink
OnLevel
OnRecordSide
OrderNumber
Owl:inverseOf
Pdch
Pdeu
Pdint
Pdusa
PerformerHadFunction
PerformerHasMoreInformation
PerformerHasMusicBrainzID
PerformerYearBorn
PerformerYearDied
Schema:alternateName
Schema:associatedMedia
Schema:audio
Schema:byArtist
Schema:catalogNumber
Schema:citation
Schema:contentUrl
Schema:contributor
Schema:copyrightHolder
Schema:creator

Schema:dateCreated
Schema:datePublished
Schema:description
Schema:discussionUrl
Schema:editor
Schema:encodesCreativeWork
Schema:encodingFormat
Schema:exampleOfWork
Schema:fileFormat
Schema:funder
Schema:genre
Schema:hasPart
Schema:image
Schema:inAlbum
Schema:inLanguage
Schema:isAccessibleForFree
Schema:isPartOf
Schema:license
Schema:locationCreated
Schema:name
Schema:numTracks
Schema:publisher
Schema:sameAs
Schema:schemaVersion
Schema:thumbnailUrl
Schema:track
Schema:uploadDate
Schema:url
StamperID
TitleLink
UsedCleaningProcess
UsedDigitizingProcess

14.5.3 Liste der geänderten Templates

Audio file
Audio storage description
CI
CI/layout
FLAC missing
Image missing
Label missing
PD-1923
PD-1923/layout
PD-Author
PD-CH
PD-CH/layout
PD-EU
PD-EU/layout
Pd-ineligible
PD-ineligible/en

PD-ineligible/layout
PD-INT
PD-INT/layout
PD-Italy-audio
PD-Italy-audio/layout
PD-Layout
PD-old-70-1923
PD-old/layout
PD-Open Clipart Library
PD-OpenClipart
PD-self
PD-textlogo
PD-textlogo/en
PD-textlogo/layout
PD-traditional
PD-US
PD-US-95
PD-US-95/layout
PD-US-record
PD-US/layout
PD-user
Phonograph cylinder description
RR-CH
RR-EU
Tango project/layout

14.5.4 Liste der neu erstellten Templates

Archived item
Audio medium
Audio medium missing
Audio record
Audio record missing
Audio track
Cleaning process missing
Container missing
Digitized audio track
Digitizing process missing
ExtractAudioFileUrl
ExtractFlacFileUrl
License investigation
Link to audio medium missing
Query output isLocationOf
Release set
Release set not SMW
Render contributor information
Smw composer
Smw conductor
Smw coverArtist
Smw depiction
Smw license

- Smw List record content
- Smw lyricist
- Smw music arranger
- Smw performer
- Smw Query release set record content
- Smw record content table intro
- Smw record content table outro
- Smw record content table row
- Storage box
- Storage box size
- Storage box weight
- Storage media location report
- Storage medium information
- Unit conversion length

14.5.5 Liste der erstellten Formulare

- Audio track
- Audio medium
- Audio release set
- Storage box

14.5.6 Liste der neu erstellten Kategorien

- Archived item
- Audio record
- Audio release set
- Audio track
- Audio tracks
- AudioObject
- Box
- Carton
- Class
- CreativeWork
- Digitized audio track
- Location
- Logistics object
- MediaObject
- MoveableResource
- MusicAlbum
- MusicPlaylist
- MusicRecording
- MusicRelease
- Package
- Pages with too many expensive parser function calls
- PD files lacking an image
- PD INT
- PhysicalResource
- Semantic MediaWiki helpers
- Thing

14.6 Reguläre Ausdrücke für Datenmigration in semantische Templates

Reguläre Ausdrücke um Bilderlinks bei Templates zu korrigieren. Sie sind vorgesehen zur Verwendung mit MassEditRegex.

`_Image variable Audio file template_`

Search for: `/(\{\{Audio file(.|\n)*\}|Image=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1 $3.jpg`

Search for: `/(\{\{Audio file(.|\n)*\})(\|Image1=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1|Image= $4.jpg`

Search for: `/(\{\{Audio file(.|\n)*\})(\|Image2=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1|Image1= $4.jpg`

Search for: `/(\{\{Audio file(.|\n)*\})(\|Image3=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1|Image2= $4.jpg`

Search for: `/(\{\{Audio file(.|\n)*\})(\|Image4=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1|Image3= $4.jpg`

`_Image variable Audio storage description template_`

Search for: `/(\{\{Audio storage description(.|\n)*\}|Image=).*\[\[.*Image:(.*)\.jpg.*\]\].*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1 $3.jpg \n|Image2= $4.jpg`

Search for: `/(\{\{Audio storage description(.|\n)*\}|Image=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1 $3.jpg`

Search for: `/(\{\{Audio storage description(.|\n)*\})(\|Image1=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1|Image= $4.jpg`

Search for: `/(\{\{Audio storage description(.|\n)*\})(\|Image2=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1|Image1= $4.jpg`

Search for: `/(\{\{Audio storage description(.|\n)*\})(\|Image3=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1|Image2= $4.jpg`

Search for: `/(\{\{Audio storage description(.|\n)*\})(\|Image4=).*\[\[.*Image:(.*)\.jpg.*\]\]/`
Replace with: `$1|Image3= $4.jpg`

14.7 Anpassungen an der Konfiguration des MediaWiki Servers

In diesem Kapitel des Anhangs sind die während dieser Arbeit vorgenommenen Änderungen an der Konfiguration des MediaWiki Servers dokumentiert.

14.7.1 Anpassen der Limiten von Semantic MediaWiki

Um die Concepts zu erstellen, die für das effiziente ausfüllen der Formulare nötig sind, mussten die Limiten von Semantic MediaWiki in der Konfiguration angepasst werden. Semantic MediaWiki unterscheidet zwischen schwierigen (hard) und normalen Abfragen. Schwierige Abfragen werden offline über einen Cronjob berechnet und zwischengespeichert, sie werden ein mal pro Tag aktualisiert.

Damit neu hinzugefügte Labels, Lagerkisten etc. aber sofort im Formular ausgewählt werden können, mussten weitere Suchfeatures als einfach eingestuft werden. Dazu gehören alle Features, die in der Konfigurationsvariable `smwgQFeatures` aktiviert sind. Die Datei `/var/www/mediawiki-conf/ExtensionSettings.php` musste wie folgt angepasst werden:

```
# SMW creates and alters temporary tables for certain semantic queries.
# If you don't wish to allow such privileges, you can disable these specific query features by adding the following:
$smwgQMaxSize = 12;
$smwgQMaxDepth = 4;
$smwgQFeatures = SMW_CATEGORY_QUERY | SMW_CONCEPT_QUERY |
SMW_CONJUNCTION_QUERY | SMW_DISJUNCTION_QUERY;
$smwgQConceptMaxSize = 20;
$smwgQConceptMaxDepth = 8;
$smwgQConceptFeatures = SMW_PROPERTY_QUERY | SMW_CATEGORY_QUERY | SMW_NAME-
SPACE_QUERY | SMW_CONJUNCTION_QUERY | SMW_DISJUNCTION_QUERY;
```

Durch diese Änderungen muss Semantic MediaWiki temporäre Tabellen in der Datenbank anlegen und wieder entfernen können. Dazu benötigt der Datenbankbenutzer, unter dem das Semantic MediaWiki die Datenbank anspricht, zusätzlich die Datenbankberechtigungen `CREATE TEMPORARY TABLES` und `DROP`.

MySQL Konsole starten:

```
mysql pool_publicdomain
```

Rechte hinzufügen:

```
mysql> GRANT CREATE TEMPORARY TABLES, DROP ON publicdomainwiki_pool.* TO
'public_pool'@'localhost' ;
```

14.7.2 Installation von MassEditRegex

Für diese Arbeit mussten im MediaWiki diverse Änderungen gemacht werden an bestehenden Wikiseiten und den Werten, die den Templates übergeben werden. Dies betrifft jeweils mehrere Seiten und es sollen nicht alle Seiten von Hand geändert werden müssen. Dazu bieten sich reguläre Ausdrücke an, die auf die zu ändernden Seiten angewendet werden.

Eine Erweiterung, die diese Funktionalität bietet ist „Replace Text“⁸⁹. Während der Arbeit wurde aber festgestellt, dass Replace Text Probleme hat, die sich so ausgewirkt haben, dass die Vorschau zwar die korrekten Wikiseiten auflistet, aber danach trotzdem keine Änderungen vorgenommen wurden. In der Folge wurde darum eine weitere Extension „MassEditRegex“⁹⁰ installiert. Die beiden Extensions benötigen eine leicht andere Syntax.

Die Installation erforderte folgende Arbeitsschritte:

```
cd /var/www/mediawiki-extensions/  
git clone https://phabricator.wikimedia.org/diffusion/EMER/extension-masseditregex.git  
mv extension-masseditregex/ MassEditRegex/  
cd MassEditRegex  
git switch REL1_27  
nano /var/www/mediawiki-conf/ExtensionSettings.php
```

Folgende Zeilen müssen in die ExtensionSettings.php eingefügt werden:

```
## MassEditRegex (git)  
## The MassEditRegex extension allows administrators perform a single edit  
## across multiple pages in one step, by running a regular expression over  
## the  
## content of each page.  
require_once "$wgOtherExtensions/MassEditRegex/MassEditRegex.php";  
$wgGroupPermissions['sysop']['masseditregex'] = true;
```

14.7.3 Fehler zend_mm_heap corrupted beheben :

Während der Arbeit funktionierte das Anzeigen der Seite „Recent Changes“ plötzlich nicht mehr. Das Apache Fehlerlog zeigte nur den Fehler zend_mm_heap corrupted an. Dieser Fehler kann gemäss der Diskussion auf Stackoverflow⁹¹ diverse Ursachen haben und je nach dem sind andere Fehlerbehebungen nötig. In unserem Fall war eine Änderung der Opcache Einstellungen nötig:

```
/etc/php5/mods-available/opcache.ini  
opcache.fast_shutdown=0
```

14.7.4 Setzen der korrekten Serveradresse

Als die ersten Properties eingefügt wurden, die URLs von Bildern bereitstellen, wurde festgestellt, dass die exportierten URLs in den Properties falsch waren. Überall stand als Domäne localhost anstatt pool.publicdomainproject.org.

Eigentlich wird diese Information aus der Apache „ServerName“ Konfigurationsvariable ausgelesen, die korrekt gesetzt war. In MediaWiki gibt es eine eigene Konfigurationsvariable um die URL des Servers zu setzen. Dies musste für jede Sprachversion (Subdomain) konfiguriert werden. Folgender Eintrag ist dazu in der LocalSettings.php nötig:

```
$wgServer = "//pool.publicdomainproject.org";  
$wgCanonicalServer = "https://pool.publicdomainproject.org";
```

⁸⁹ https://www.mediawiki.org/wiki/Extension:Replace_Text

⁹⁰ <https://www.mediawiki.org/wiki/Extension:MassEditRegex>

⁹¹ <https://stackoverflow.com/questions/2247977/what-does-zend-mm-heap-corrupted-mean>

14.8 Apache2 Konfiguration

14.8.1 Vhost Konfiguration

Inhalt der Datei: /etc/apache2/sites-enabled/000-default.conf

```
<VirtualHost *:80>
    # The ServerName directive sets the request scheme, hostname and port
that
    # the server uses to identify itself. This is used when creating
    # redirection URLs. In the context of virtual hosts, the ServerName
    # specifies what hostname must appear in the request's Host: header
to
    # match this virtual host. For the default virtual host (this file)
this
    # value is not decisive as it is used as a last resort host regard-
less.
    # However, you must set it for any further virtual host explicitly.
    ServerName publicdomain.ch

    ServerAdmin webmaster@localhost
    DocumentRoot /html

    <Directory /html/>
        Options FollowSymLinks
        Require all granted
        AllowOverride All
    </Directory>

    # Available loglevels: trace8, ..., tracel, debug, info, notice,
warn,
    # error, crit, alert, emerg.
    # It is also possible to configure the loglevel for particular
    # modules, e.g.
    #LogLevel info ssl:warn
    LogLevel warn

    ErrorLog ${APACHE_LOG_DIR}/error.log
    CustomLog ${APACHE_LOG_DIR}/access.log combined

    # For most configuration files from conf-available/, which are
    # enabled or disabled at a global level, it is possible to
    # include a line for only one particular virtual host. For example
the
    # following line enables the CGI configuration for this host only
    # after it has been globally disabled with "a2disconf".
    #Include conf-available/serve-cgi-bin.conf

    # SSL Engine Switch:
    # Enable/Disable SSL for this virtual host.
    SSLEngine off
</VirtualHost>

<VirtualHost *:443>
    # The ServerName directive sets the request scheme, hostname and port
that
    # the server uses to identify itself. This is used when creating
```



```

# redirection URLs. In the context of virtual hosts, the ServerName
# specifies what hostname must appear in the request's Host: header
to
# match this virtual host. For the default virtual host (this file)
this
# value is not decisive as it is used as a last resort host regard-
less.
# However, you must set it for any further virtual host explicitly.
ServerName publicdomain.ch

ServerAdmin webmaster@localhost
DocumentRoot /html

<Directory /html/>
    Options FollowSymLinks
    Require all granted
    AllowOverride All
</Directory>

# Available loglevels: trace8, ..., tracel, debug, info, notice,
warn,
# error, crit, alert, emerg.
# It is also possible to configure the loglevel for particular
# modules, e.g.
#LogLevel info ssl:warn
LogLevel warn

ErrorLog ${APACHE_LOG_DIR}/error.log
CustomLog ${APACHE_LOG_DIR}/access.log combined

# For most configuration files from conf-available/, which are
# enabled or disabled at a global level, it is possible to
# include a line for only one particular virtual host. For example
the
# following line enables the CGI configuration for this host only
# after it has been globally disabled with "a2disconf".
#Include conf-available/serve-cgi-bin.conf

# SSL Engine Switch:
# Enable/Disable SSL for this virtual host.
SSLEngine on

# Enable HTTP Strict Transport Security
# While redirecting all traffic to HTTPS is good, it may not comple-
tely
# prevent man-in-the-middle attacks. Thus administrators are encoura-
ged
# to set the HTTP Strict Transport Security header, which instructs
# browsers to not allow any connection to the instance using HTTP,
and
# it attempts to prevent site visitors from bypassing invalid
# certificate warnings.
<IfModule mod_headers.c>
    Header always set Strict-Transport-Security "max-
age=15552000; includeSubDomains; preload"
</IfModule>
</VirtualHost>

```

14.8.2 SSL.conf

Inhalt der Datei: /etc/apache2/mods-enabled/ssl.conf

```
<IfModule mod_ssl.c>

    # Pseudo Random Number Generator (PRNG):
    # Configure one or more sources to seed the PRNG of the SSL library.
    # The seed data should be of good random quality.
    # WARNING! On some platforms /dev/random blocks if not enough entropy
    # is available. This means you then cannot use the /dev/random device
    # because it would lead to very long connection times (as long as
    # it requires to make more entropy available). But usually those
    # platforms additionally provide a /dev/urandom device which doesn't
    # block. So, if available, use this one instead. Read the mod_ssl
    # Manual for more details.
    #
    SSLRandomSeed startup builtin
    SSLRandomSeed startup file:/dev/urandom 512
    SSLRandomSeed connect builtin
    SSLRandomSeed connect file:/dev/urandom 512

    ##
    ## SSL Global Context
    ##
    ## All SSL configuration in this context applies both to
    ## the main server and all SSL-enabled virtual hosts.
    ##

    #
    # Some MIME-types for downloading Certificates and CRLs
    #
    AddType application/x-x509-ca-cert .crt
    AddType application/x-pkcs7-crl .crl

    # Pass Phrase Dialog:
    # Configure the pass phrase gathering process.
    # The filtering dialog program ('builtin' is a internal
    # terminal dialog) has to provide the pass phrase on stdout.
    SSLPassPhraseDialog exec:/usr/share/apache2/ask-for-passphrase

    # Inter-Process Session Cache:
    # Configure the SSL Session Cache: First the mechanism
    # to use and second the expiring timeout (in seconds).
    # (The mechanism dbm has known memory leaks and should not be
    used).
    #SSLSessionCache dbm:${APACHE_RUN_DIR}/ssl_scache
    SSLSessionCache shmcb:${APACHE_RUN_DIR}/ssl_scache(512000)
    SSLSessionCacheTimeout 300

    # Semaphore:
    # Configure the path to the mutual exclusion semaphore the
    # SSL engine uses internally for inter-process synchronization.
    # (Disabled by default, the global Mutex directive consolidates by
    default
```

```

# this)
#Mutex file:${APACHE_LOCK_DIR}/ssl_mutex ssl-cache

# SSL Cipher Suite:
# List the ciphers that the client is permitted to negotiate. See
the
# ciphers(1) man page from the openssl package for list of all
available
# options.
# Enable only secure ciphers:
#SSLCipherSuite HIGH:!aNULL
SSLCipherSuite ECDHE-RSA-AES128-GCM-SHA256:ECDHE-ECDSA-AES128-GCM-S-
HA256:ECDHE-RSA-AES256-GCM-SHA384:ECDHE-ECDSA-A
ES256-GCM-SHA384:DHE-RSA-AES128-GCM-SHA256:DHE-DSS-AES128-GCM-
SHA256:KEDH+AESGCM:ECDHE-RSA-AES128-SHA256:ECDHE-ECDSA-AES12
8-SHA256:ECDHE-RSA-AES128-SHA:ECDHE-ECDSA-AES128-SHA:ECDHE-RSA-AES256-S-
HA384:ECDHE-ECDSA-AES256-SHA384:ECDHE-RSA-AES256-SH
A:ECDHE-ECDSA-AES256-SHA:DHE-RSA-AES128-SHA256:DHE-RSA-AES128-SHA:DHE-DSS-
AES128-SHA256:DHE-RSA-AES256-SHA256:DHE-DSS-AES2
56-SHA:DHE-RSA-AES256-SHA:AES128-GCM-SHA256:AES256-GCM-SHA384:AES128-
SHA256:AES256-SHA256:AES128-SHA:AES256-SHA:AES:CAMELL
IA:DES-CBC3-SHA:!aNULL:!eNULL:!EXPORT:!DES:!RC4:!MD5:!PSK:!aECDH:!EDH-DSS-
DES-CBC3-SHA:!EDH-RSA-DES-CBC3-SHA:!KRB5-DES-CBC
3-SHA

# SSL server cipher order preference:
# Use server priorities for cipher algorithm choice.
# Clients may prefer lower grade encryption. You should enable this
# option if you want to enforce stronger encryption, and can afford
# the CPU cost, and did not override SSLCipherSuite in a way that
puts
# insecure ciphers first.
# Default: Off
SSLHonorCipherOrder on

# The protocols to enable.
# Available values: all, SSLv3, TLSv1, TLSv1.1, TLSv1.2
# SSL v2 is no longer supported
SSLProtocol all -SSLv3

# Allow insecure renegotiation with clients which do not yet sup-
port the
# secure renegotiation protocol. Default: Off
#SSLInsecureRenegotiation on

# Whether to forbid non-SNI clients to access name based virtual
hosts.
# Default: Off
#SSLStrictSNIVHostCheck On

# The only CRIME workaround is to work on the third condition, and
to disable
# TLS compression. This is the default in recent Apache releases,
as in recent
# browsers. On earlier Apache releases, the following configuration
directive
# lets you disable it

```

```
SSLCompression off

    SSLCertificateFile /etc/apache2/publicdomain.ch/fullchain.cer
    SSLCertificateKeyFile /etc/apache2/publicdomain.ch/publicdo-
main.ch.key
#     SSLOpenSSLConfCmd DHParameters "/etc/apache2/dhparams.pem"
</IfModule>
```

14.9 Security Testberichte von Qualys SSL Labs

You are here: [Home](#) > [Projects](#) > [SSL Server Test](#) > publicdomain.ch

SSL Report: publicdomain.ch (195.176.247.106)

Assessed on: Fri, 02 Dec 2016 11:14:57 UTC | [Hide](#) | [Clear cache](#)

[Scan Another »](#)

Summary

Overall Rating

A+

Certificate

Protocol Support

Key Exchange

Cipher Strength

0 20 40 60 80 100

Visit our [documentation page](#) for more information, configuration guides, and books. Known issues are documented [here](#).

HTTP Strict Transport Security (HSTS) with long duration deployed on this server. [MORE INFO »](#)

Certificate #1: RSA 2048 bits (SHA256withRSA)



Server Key and Certificate #1

Subject	publicdomain.ch Fingerprint SHA1: 9a438fcc8c07d2e315ab9fae2ffb65923611bf46 Pin SHA256: P69sqp71417nbr1JPIfEZeamV0vY14Gwnw7SFYPaC8g=
Common names	publicdomain.ch
Alternative names	publicdomain.ch www.publicdomain.ch
Valid from	Wed, 02 Nov 2016 12:09:00 UTC
Valid until	Tue, 31 Jan 2017 12:09:00 UTC (expires in 1 month and 29 days)
Key	RSA 2048 bits (e 65537)
Weak key (Debian)	No
Issuer	Let's Encrypt Authority X3 AIA: http://cert.int-x3.letsencrypt.org/
Signature algorithm	SHA256withRSA
Extended Validation	No
Certificate Transparency	No
OCSP Must Staple	No
Revocation information	OCSP OCSP: http://ocsp.int-x3.letsencrypt.org/
Revocation status	Good (not revoked)
Trusted	Yes



Additional Certificates (if supplied)

Certificates provided	2 (2481 bytes)
Chain issues	None

#2

Subject	Let's Encrypt Authority X3 Fingerprint SHA1: e6a3b45b062d509b3382282d196efe97d5956ccb Pin SHA256: YLh1dUR9y6Kja30RrAn7JKnbQG/uEILMkBgFF2Fuihg=
Valid until	Wed, 17 Mar 2021 16:40:46 UTC (expires in 4 years and 3 months)
Key	RSA 2048 bits (e 65537)
Issuer	DST Root CA X3
Signature algorithm	SHA256withRSA



Certification Paths



[Click here to expand](#)

Configuration



Protocols

TLS 1.2	Yes
TLS 1.1	Yes
TLS 1.0	Yes
SSL 3	No
SSL 2	No



Cipher Suites (SSL 3+ suites in server-preferred order; deprecated and SSL 2 suites at the end)

TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 (0xc02f)	ECDH secp256r1 (eq. 3072 bits RSA) FS	128
TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384 (0xc030)	ECDH secp256r1 (eq. 3072 bits RSA) FS	256
TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 (0xc02e)	DH 2048 bits FS	128
TLS_DHE_RSA_WITH_AES_256_GCM_SHA384 (0xc02f)	DH 2048 bits FS	256
TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 (0xc027)	ECDH secp256r1 (eq. 3072 bits RSA) FS	128
TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA (0xc013)	ECDH secp256r1 (eq. 3072 bits RSA) FS	128
TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384 (0xc028)	ECDH secp256r1 (eq. 3072 bits RSA) FS	256
TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA (0xc014)	ECDH secp256r1 (eq. 3072 bits RSA) FS	256
TLS_DHE_RSA_WITH_AES_128_CBC_SHA256 (0xc026)	DH 2048 bits FS	128
TLS_DHE_RSA_WITH_AES_128_CBC_SHA (0xc011)	DH 2048 bits FS	128
TLS_DHE_RSA_WITH_AES_256_CBC_SHA256 (0xc02b)	DH 2048 bits FS	256
TLS_DHE_RSA_WITH_AES_256_CBC_SHA (0xc012)	DH 2048 bits FS	256
TLS_RSA_WITH_AES_128_GCM_SHA256 (0xc02c)		128
TLS_RSA_WITH_AES_256_GCM_SHA384 (0xc02d)		256
TLS_RSA_WITH_AES_128_CBC_SHA256 (0xc020)		128
TLS_RSA_WITH_AES_256_CBC_SHA256 (0xc021)		256
TLS_RSA_WITH_AES_128_CBC_SHA (0xc010)		128
TLS_RSA_WITH_AES_256_CBC_SHA (0xc014)		256
TLS_DHE_RSA_WITH_CAMELLIA_256_CBC_SHA (0xc088)	DH 2048 bits FS	256
TLS_RSA_WITH_CAMELLIA_256_CBC_SHA (0xc084)		256
TLS_DHE_RSA_WITH_CAMELLIA_128_CBC_SHA (0xc045)	DH 2048 bits FS	128
TLS_RSA_WITH_CAMELLIA_128_CBC_SHA (0xc041)		128
TLS_RSA_WITH_3DES_EDE_CBC_SHA (0xc02a)		112



Handshake Simulation

Android 2.3.7	No SNI ²	RSA 2048 (SHA256)	TLS 1.0	TLS_DHE_RSA_WITH_AES_128_CBC_SHA	DH 2048	FS
Android 4.0.4		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
Android 4.1.1		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
Android 4.2.2		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
Android 4.3		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
Android 4.4.2		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Android 5.0.0		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Android 6.0		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Android 7.0		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Baidu Jan 2015		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
BingPreview Jan 2015		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Chrome 49 / XP SP3		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Chrome 51 / Win 7	R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Firefox 31.3.0 ESR / Win 7		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Firefox 47 / Win 7	R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Firefox 49 / XP SP3		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Firefox 49 / Win 7	R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Googlebot Feb 2015		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
IE 6 / XP	No FS ¹ No SNI ²		Server closed connection			
IE 7 / Vista		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS

Handshake Simulation

IE 8 / XP No FS ¹ No SNI ²	RSA 2048 (SHA256)	TLS 1.0	TLS_RSA_WITH_3DES_EDE_CBC_SHA
IE 8-10 / Win 7 R	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
IE 11 / Win 7 R	RSA 2048 (SHA256)	TLS 1.2	TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 DH 2048 FS
IE 11 / Win 8.1 R	RSA 2048 (SHA256)	TLS 1.2	TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 DH 2048 FS
IE 10 / Win Phone 8.0	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
IE 11 / Win Phone 8.1 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
IE 11 / Win Phone 8.1 Update R	RSA 2048 (SHA256)	TLS 1.2	TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 DH 2048 FS
IE 11 / Win 10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Edge 13 / Win 10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Edge 13 / Win Phone 10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Java 6u45 No SNI ²	Client does not support DH parameters > 1024 bits		
	RSA 2048 (SHA256) TLS 1.0 TLS_DHE_RSA_WITH_AES_128_CBC_SHA DH 2048		
Java 7u25	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
Java 8u31	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
OpenSSL 0.9.8y	RSA 2048 (SHA256)	TLS 1.0	TLS_DHE_RSA_WITH_AES_128_CBC_SHA DH 2048 FS
OpenSSL 1.0.1 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
OpenSSL 1.0.2e R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Safari 5.1.9 / OS X 10.6.8	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
Safari 6 / iOS 6.0.1	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 6.0.4 / OS X 10.8.4 R	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
Safari 7 / iOS 7.1 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 7 / OS X 10.9 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 8 / iOS 8.4 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 8 / OS X 10.10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 9 / iOS 9 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Safari 9 / OS X 10.11 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Safari 10 / iOS 10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Safari 10 / OS X 10.12 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Apple ATS 9 / iOS 9 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Yahoo Slurp Jan 2015	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
YandexBot Jan 2015	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS

(1) Clients that do not support Forward Secrecy (FS) are excluded when determining support for it.

(2) No support for virtual SSL hosting (SNI). Connects to the default site if the server uses SNI.

(3) Only first connection attempt simulated. Browsers sometimes retry with a lower protocol version.

(R) Denotes a reference browser or client, with which we expect better effective security.

(All) We use defaults, but some platforms do not use their best protocols and features (e.g., Java 6 & 7, older IE).



Protocol Details

	No, server keys and hostname not seen elsewhere with SSLV2
DROWN (experimental)	<p>(1) For a better understanding of this test, please read this longer explanation</p> <p>(2) Key usage data kindly provided by the Censys network search engine; original DROWN test here</p> <p>(3) Censys data is only indicative of possible key and certificate reuse; possibly out-of-date and not complete</p>
Secure Renegotiation	Supported
Secure Client-Initiated Renegotiation	No
Insecure Client-Initiated Renegotiation	No
BEAST attack	Not mitigated server-side (more info) TLS 1.0: 0xc013
POODLE (SSLv3)	No, SSL 3 not supported (more info)
POODLE (TLS)	No (more info)
Downgrade attack prevention	Yes, TLS_FALLBACK_SCSV supported (more info)
SSL/TLS compression	No
RC4	No
Heartbeat (extension)	Yes
Heartbleed (vulnerability)	No (more info)
OpenSSL CCS vuln. (CVE-2014-0224)	No (more info)
OpenSSL Padding Oracle vuln. (CVE-2016-2107)	No (more info)
Forward Secrecy	Yes (with most browsers) ROBUST (more info)
ALPN	No
NPN	No
Session resumption (caching)	Yes
Session resumption (tickets)	Yes
OCSP stapling	No

Protocol Details

Strict Transport Security (HSTS)	Yes max-age=15552000; includeSubDomains; preload
HSTS Preloading	Not in: Chrome Edge Firefox IE
Public Key Pinning (HPKP)	No
Public Key Pinning Report-Only	No
Long handshake intolerance	No
TLS extension intolerance	No
TLS version intolerance	No
Incorrect SNI alerts	No
Uses common DH primes	No
DH public server param (Ys) reuse	No
SSL 2 handshake compatibility	Yes

**HTTP Requests**

1 <https://publicdomain.ch/> (HTTP/1.1 200 OK)

**Miscellaneous**

Test date	Fri, 02 Dec 2016 11:12:45 UTC
Test duration	131.974 seconds
HTTP status code	200
HTTP server signature	Apache/2.4.10 (Debian)
Server hostname	-

SSL Report v1.25.2

You are here: [Home](#) > [Projects](#) > [SSL Server Test](#) > publicdomainproject.org

SSL Report: publicdomainproject.org (195.176.247.101)

Assessed on: Fri, 02 Dec 2016 11:24:47 UTC | [Hide](#) | [Clear cache](#)

[Scan Another »](#)

Summary

Overall Rating

A

Certificate

Protocol Support

Key Exchange

Cipher Strength

0 20 40 60 80 100

Visit our [documentation page](#) for more information, configuration guides, and books. Known issues are documented [here](#).

Certificate #1: RSA 2048 bits (SHA256withRSA)



Server Key and Certificate #1

Subject	publicdomainproject.org Fingerprint SHA1: 2431cdd61ae38622c4b61ec046a1161ce3be5a0b Pin SHA256: K9C0arfoRdhMBW4PyW8LKQk4CdKCdJSxxQyAwpkb/wk=
Common names	publicdomainproject.org
Alternative names	de.publicdomainproject.org en.publicdomainproject.org es.publicdomainproject.org fr.publicdomainproject.org it.publicdomainproject.org pool.publicdomainproject.org publicdomainproject.org
Valid from	Tue, 01 Nov 2016 11:15:00 UTC
Valid until	Mon, 30 Jan 2017 11:15:00 UTC (expires in 1 month and 27 days)
Key	RSA 2048 bits (e 65537)
Weak key (Debian)	No
Issuer	Let's Encrypt Authority X3 AIA: http://cert.int-x3.letsencrypt.org/
Signature algorithm	SHA256withRSA
Extended Validation	No
Certificate Transparency	No
OCSP Must Staple	No
Revocation information	OCSP OCSP: http://ocsp.int-x3.letsencrypt.org/
Revocation status	Good (not revoked)
Trusted	Yes



Additional Certificates (if supplied)

Certificates provided	2 (2649 bytes)
Chain issues	None
#2	
Subject	Let's Encrypt Authority X3 Fingerprint SHA1: e6a3b45b062d509b3382282d196efe97d5956ccb Pin SHA256: YLh1dUR9y6Kja30RrAn7JKnbQG/uEtLMkBgFF2Fuihg=
Valid until	Wed, 17 Mar 2021 16:40:46 UTC (expires in 4 years and 3 months)
Key	RSA 2048 bits (e 65537)
Issuer	DST Root CA X3
Signature algorithm	SHA256withRSA



Certification Paths



[Click here to expand](#)

Configuration



Protocols

TLS 1.2	Yes
TLS 1.1	Yes
TLS 1.0	Yes
SSL 3	No
SSL 2	No



Cipher Suites (SSL 3+ suites in server-preferred order; deprecated and SSL 2 suites at the end)

TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 (0xc02f)	ECDH secp256r1 (eq. 3072 bits RSA) FS	128
TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384 (0xc030)	ECDH secp256r1 (eq. 3072 bits RSA) FS	256
TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 (0x9e)	DH 2048 bits FS	128
TLS_DHE_RSA_WITH_AES_256_GCM_SHA384 (0x9f)	DH 2048 bits FS	256
TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 (0xc027)	ECDH secp256r1 (eq. 3072 bits RSA) FS	128
TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA (0xc013)	ECDH secp256r1 (eq. 3072 bits RSA) FS	128
TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384 (0xc028)	ECDH secp256r1 (eq. 3072 bits RSA) FS	256
TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA (0xc014)	ECDH secp256r1 (eq. 3072 bits RSA) FS	256
TLS_DHE_RSA_WITH_AES_128_CBC_SHA256 (0x67)	DH 2048 bits FS	128
TLS_DHE_RSA_WITH_AES_128_CBC_SHA (0x33)	DH 2048 bits FS	128
TLS_DHE_RSA_WITH_AES_256_CBC_SHA256 (0x6b)	DH 2048 bits FS	256
TLS_DHE_RSA_WITH_AES_256_CBC_SHA (0x39)	DH 2048 bits FS	256
TLS_RSA_WITH_AES_128_GCM_SHA256 (0x9c)		128
TLS_RSA_WITH_AES_256_GCM_SHA384 (0x9d)		256
TLS_RSA_WITH_AES_128_CBC_SHA256 (0x3c)		128
TLS_RSA_WITH_AES_256_CBC_SHA256 (0x3d)		256
TLS_RSA_WITH_AES_128_CBC_SHA (0x2f)		128
TLS_RSA_WITH_AES_256_CBC_SHA (0x35)		256
TLS_DHE_RSA_WITH_CAMELLIA_256_CBC_SHA (0x88)	DH 2048 bits FS	256
TLS_RSA_WITH_CAMELLIA_256_CBC_SHA (0x84)		256
TLS_DHE_RSA_WITH_CAMELLIA_128_CBC_SHA (0x45)	DH 2048 bits FS	128
TLS_RSA_WITH_CAMELLIA_128_CBC_SHA (0x41)		128
TLS_RSA_WITH_3DES_EDE_CBC_SHA (0xa)		112



Handshake Simulation

Android 2.3.7	No SNI ²	RSA 2048 (SHA256)	TLS 1.0	TLS_DHE_RSA_WITH_AES_128_CBC_SHA	DH 2048	FS
Android 4.0.4		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
Android 4.1.1		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
Android 4.2.2		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
Android 4.3		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
Android 4.4.2		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Android 5.0.0		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Android 6.0		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Android 7.0		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Baidu Jan 2015		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS
BingPreview Jan 2015		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Chrome 49 / XP SP3		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Chrome 51 / Win 7	R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Firefox 31.3.0 ESR / Win 7		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Firefox 47 / Win 7	R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Firefox 49 / XP SP3		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Firefox 49 / Win 7	R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
Googlebot Feb 2015		RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ECDH secp256r1	FS
IE 6 / XP	No FS ¹ No SNI ²		Server closed connection			
IE 7 / Vista		RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	ECDH secp256r1	FS

Handshake Simulation

IE 8 / XP No FS ¹ No SNI ²	RSA 2048 (SHA256)	TLS 1.0	TLS_RSA_WITH_3DES_EDE_CBC_SHA
IE 8-10 / Win 7 R	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
IE 11 / Win 7 R	RSA 2048 (SHA256)	TLS 1.2	TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 DH 2048 FS
IE 11 / Win 8.1 R	RSA 2048 (SHA256)	TLS 1.2	TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 DH 2048 FS
IE 10 / Win Phone 8.0	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
IE 11 / Win Phone 8.1 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
IE 11 / Win Phone 8.1 Update R	RSA 2048 (SHA256)	TLS 1.2	TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 DH 2048 FS
IE 11 / Win 10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Edge 13 / Win 10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Edge 13 / Win Phone 10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Java 6u45 No SNI ²	Client does not support DH parameters > 1024 bits		
	RSA 2048 (SHA256) TLS 1.0 TLS_DHE_RSA_WITH_AES_128_CBC_SHA DH 2048		
Java 7u25	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
Java 8u31	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
OpenSSL 0.9.8y	RSA 2048 (SHA256)	TLS 1.0	TLS_DHE_RSA_WITH_AES_128_CBC_SHA DH 2048 FS
OpenSSL 1.0.1l R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
OpenSSL 1.0.2e R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Safari 5.1.9 / OS X 10.6.8	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
Safari 6 / iOS 6.0.1	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 6.0.4 / OS X 10.8.4 R	RSA 2048 (SHA256)	TLS 1.0	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA ECDH secp256r1 FS
Safari 7 / iOS 7.1 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 7 / OS X 10.9 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 8 / iOS 8.4 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 8 / OS X 10.10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 ECDH secp256r1 FS
Safari 9 / iOS 9 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Safari 9 / OS X 10.11 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Safari 10 / iOS 10 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Safari 10 / OS X 10.12 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Apple ATS 9 / iOS 9 R	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
Yahoo Slurp Jan 2015	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS
YandexBot Jan 2015	RSA 2048 (SHA256)	TLS 1.2	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 ECDH secp256r1 FS

(1) Clients that do not support Forward Secrecy (FS) are excluded when determining support for it.

(2) No support for virtual SSL hosting (SNI). Connects to the default site if the server uses SNI.

(3) Only first connection attempt simulated. Browsers sometimes retry with a lower protocol version.

(R) Denotes a reference browser or client, with which we expect better effective security.

(All) We use defaults, but some platforms do not use their best protocols and features (e.g., Java 6 & 7, older IE).



Protocol Details

	No, server keys and hostname not seen elsewhere with SSLV2
DROWN (experimental)	<p>(1) For a better understanding of this test, please read this longer explanation</p> <p>(2) Key usage data kindly provided by the Censys network search engine; original DROWN test here</p> <p>(3) Censys data is only indicative of possible key and certificate reuse; possibly out-of-date and not complete</p>
Secure Renegotiation	Supported
Secure Client-Initiated Renegotiation	No
Insecure Client-Initiated Renegotiation	No
BEAST attack	Not mitigated server-side (more info) TLS 1.0: 0xc013
POODLE (SSLv3)	No, SSL 3 not supported (more info)
POODLE (TLS)	No (more info)
Downgrade attack prevention	Yes, TLS_FALLBACK_SCSV supported (more info)
SSL/TLS compression	No
RC4	No
Heartbeat (extension)	Yes
Heartbleed (vulnerability)	No (more info)
OpenSSL CCS vuln. (CVE-2014-0224)	No (more info)
OpenSSL Padding Oracle vuln. (CVE-2016-2107)	No (more info)
Forward Secrecy	Yes (with most browsers) ROBUST (more info)
ALPN	No
NPN	No
Session resumption (caching)	Yes
Session resumption (tickets)	Yes
OCSP stapling	No

Protocol Details

Strict Transport Security (HSTS)	No
HSTS Preloading	Not in: Chrome Edge Firefox IE
Public Key Pinning (HPKP)	No
Public Key Pinning Report-Only	No
Long handshake intolerance	No
TLS extension intolerance	No
TLS version intolerance	No
Incorrect SNI alerts	No
Uses common DH primes	No
DH public server param (Ys) reuse	No
SSL 2 handshake compatibility	Yes

**HTTP Requests**

1 <https://publicdomainproject.org/> (HTTP/1.1 200 OK)

**Miscellaneous**

Test date	Fri, 02 Dec 2016 11:22:34 UTC
Test duration	133.254 seconds
HTTP status code	200
HTTP server signature	Apache/2.4.10 (Debian)
Server hostname	-

SSL Report v1.25.2

14.10 Detailliertere Planung und Aufwandsabschätzung der Systemarchitektur

Metadaten in strukturierter Form nach aussen zur Verfügung stellen

Erweiterte Suchfunktionen und Linked Data API anbieten. Dies ist auch nötig um die derzeit erfassten Metadaten intern in strukturierter Form weiterzuverwenden, da diese bisher nicht in strukturierter Form sondern nur in Form von Wikisyntax vorliegen.

Projekt vorläufig: 279 h

Arbeitspaket	Teilpaket	Beschreibung	Zeitaufwand [h]
Server vorbereiten			
<i>Total Arbeitspaket: 223 h</i>	Einarbeiten in Semantic MediaWiki	Es muss viel ausprobiert und getestet werden bevor die produktive Plattform geändert wird. Dazu wird ein Backup von der virtuellen Maschine des MediaWiki Servers auf einem Testrechner eingerichtet.	10
	Testsystem einrichten		8
	Semantic MediaWiki und Extensions auf aktuellen Stand bringen	Jetzige Version ist nicht kompatibel mit jetzigem MediaWiki	5
	Für die vorhandenen Metadaten die zu benutzenden Ontologien und Vokabulare definieren und Auswahl begründen		40
	Auswahl an Ontologien mit Partnern absprechen, ob diese damit umgehen können.		16
	Anpassungen an der Ontologieauswahl vornehmen		16
	Die ausgewählten Ontologien und deren Properties in Semantic MediaWiki einpflegen	Externe Ontologien können genutzt werden, es muss trotzdem für jedes zu verwendende Property eine eigene Definition erstellt werden mit Datentyp, Beschreibung, URL zur externen Ontologie.	24
	Migration des bestehenden Templates		16
	Neu benötigte Kategorien erstellen		8
	Erstellen von Seitenklassen	Noch zu wenig eingearbeitet um komplett abzuschätzen ob dies für unseren Anwendungsfall hilfreich ist. Aufwand daher einfach ein Platzhalter.	16
	Erstellen von semantischen Formularen	Am Ende können die Daten über Formulare eingegeben und editiert werden. Es muss kein kompliziertes Template mehr ausgefüllt werden.	24
	Anleitung zur Verwendung des Formulars erstellen. In Arbeitsablaufdokumentation integrieren		16
	Dokumentieren		24

Aufräumen und Harmonisieren der bestehenden Metadaten nach neuem Schema	Anleitungen und Checklisten erstellen Schulung Arbeit betreuen	Kann separat von den anderen Arbeiten gemacht werden, Semantic MediaWiki benötigt dies nicht zwingend. Externe Abfragen werden so aber stark vereinfacht, da ein bekanntes standardisiertes Protokoll zum Einsatz kommt.	8
SPARQL Endpoint anbieten <i>Total Arbeitspaket: 56 h</i>	Tripple Store evaluieren		
	Bevorzugte zwei Lösungen austesten (installieren oder Demonstrationsserver nutzen)	Ausprobieren ob die angepriesenen Funktionen auch vorhanden und Funktionsfähig sind. Testen ob die Dokumentation ausreicht eine Installation erfolgreich durch zu führen.	16
	Tripple Store installieren und konfigurieren	Die meisten Tripple Stores sind Java und erfordern einen Apache Tomcat Server, der auch neu aufgesetzt werden muss.	16
	In Semantic Media Wiki SPARQL Enpoint nutzen	Dazu muss Semantic MediaWiki die Daten komplett neu in den Triple Store speichern.	4
	Testen		4
	Dokumentieren		8
Technische Metadaten für Endnutzer verfügbar, sichtbar machen <i>Arbeitspaket vorläufig: 68 h</i>	Evaluieren ob SemanticMediaWiki technische Metadaten auslesen und anbieten kann	Technische Metadaten sind z. B. Bitrate, Dateiformat, Spielzeit, Zeitpunkt der letzten Überprüfung etc. Alles was direkt mit den Dateien und deren Inhalt zu tun hat und ohne Menschliches zu tun bekannt ist.	8
	Limiten dokumentieren		8
	Mögliche Lösungswege evaluieren		40
	Auwandsabschätzung der zwei realistischsten Lösungswege		formation 12
	Entwicklung und Implementation sichtbar machen von technischen Metadaten für Endnutzer		zum Schätzen

Vertrauenswürdiger Archivspeicher

Total Projekt: 516 h

Arbeitspaket	Teilpaket	Beschreibung	Zeitaufwand [h]
Monitoring Total Arbeitspaket: 144 h		Der Zustand der Server und der Speichermedien muss überwacht werden und bei Unregelmässigkeiten müssen die zuständigen Personen informiert werden Das Monitoring soll z. B. mit der regelmässigen Datenprüfung verbunden sein.	16
	Anforderungsanalyse des Monitorings		8
	Anforderungen an das Monitoringsystem formulieren		4
	Open Source Monitoringsysteme suchen		16
	Gefundene Monitoringsysteme gemäss Anforderungen bewerten		32
	Bevorzugte zwei Lösungen austesten (installieren oder Demonstrationsserver nutzen)	Ausprobieren ob die angepriesenen Funktionen auch vorhanden und Funktionsfähig sind. Testen ob die Dokumentation ausreicht eine Installation erfolgreich durch zu führen.	8
	Entscheidung für System fällen und technische Umsetzung planen		4
	Hardware beschaffen bzw. vorhandene Hardware aufrüsten		40
	Monitoringsystem aufsetzen, Überwachungen, Meldungen und Zugänge einrichten und gemäss Anforderungen testen		16
	Dokumentation		
Regelmässige Datenprüfung			

Total Arbeitspaket: 82 h

Anforderungen an die regelmässige Datenprüfung formulieren	8
Open Source Lösungen suchen	4
Gefundene Lösungen gemäss Anforderungen bewerten	8
Entscheidung für System fällen und technische Umsetzung planen	6
Lösung zur regelmässigen Datenprüfung aufsetzen, einrichten und gemäss Anforderungen testen	24
Integration in Monitoringsystem	16
Dokumentation	16

Versioniertes Speichersystem

Total Arbeitspaket: 156 h

Anforderungen an das Versionierungssystem formulieren	8
Open Source Versionierungssysteme suchen	4
Gefundene Versionierungssysteme gemäss Anforderungen bewerten	16
Bevorzugte zwei Lösungen austesten (installieren oder Demonstrationsserver nutzen)	Ausprobieren ob die angepriesenen Funktionen auch vorhanden und Funktionsfähig sind. Testen ob die Dokumentation ausreicht eine Installation erfolgreich durch zu führen. 32
Entscheidung für System fällen und technische Umsetzung planen	8
Hardware beschaffen bzw. vorhandene Hardware aufrüsten	4

	Versionierungssystem aufsetzen und gemäss Anforderungen testen		40
	Bestehende Flac Dateien ins Versionierungssystem migrieren		4
	Migrierte Dateien auf Korrektheit überprüfen	Dazu muss schon ein System zur Datenprüfung vorhanden sein und ein bekannter Zustand der Dateien vor der Migration vorliegen bzw. ein Skript entwickelt werden, dass diese Prüfung vornimmt.	8
	Technische Dokumentation		16
	Anwender Dokumentation		16
Server des Archivspeichers frisch mit Gentoo GNU/Linux aufsetzen <i>Total Arbeitspaket: 32 h</i>			
	Testsystem einrichten	Das neue System muss auf einem Testsystem aufgesetzt werden, da der Archivspeicher nur kurze Zeit offline sein kann	4
	Basissystem aufsetzen und konfigurieren		8
	Sicherstellen, dass Gentoo die aktuell benutzten Quellcodepakete abspeichert		4
	Aktuell eingesetzte Serverdienste installieren und einrichten		4
	Migration im Rechenzentrum durchführen		8
	Dokumentation		4
Räumlich getrenntes Backup <i>Total Arbeitspaket: 16 h</i>		Derzeit ist noch kein Partner/Sponsor hierfür vorhanden	
	Prüfen ob das angebotene System den Anforderungen entspricht und ob Zugangsdaten stimmen etc.		2
	Initiales Backup ausführen		2
	Backup Plan erstellen		3
	Regelmässiges Backup einrichten gemäss Plan		5
	Dokumentation		4

Repräsentationsinformation auf Papier und digital

Total Arbeitspaket: 86 h

Die Repräsentationsinformation enthält alle Informationen die nötig sind die Speichermedien und die darauf enthaltenen Daten zu interpretieren. Z. B. Flac Dateiformat- und Dateisystemspezifikation

Überblick der benötigten Repräsentationsinformation erstellen		16
Nötige Repräsentationsinformation in digitaler Form beschaffen		32
Analysieren welche Repräsentationsinformation nötig ist um das Speichermedium auslesen zu können	Diese Information muss in Papierform vorgehalten werden	8
Repräsentationsinformation in digitaler langzeittauglicher Form auf den Speichermedien ablegen	Dies werden einfache Text- oder PDF/A-Dateien sein. Entsprechend PDF/A Spezifikation auch ablegen.	8
In Papierform nötige Repräsentationsinformation ausdrucken		8
Schublade für Papierinformation beschaffen und im Serverraum einbauen		6
Dokumentation		8

Authentizität des Benutzerzugriffs sicherstellen

Projekt vorläufig: 0 h

Arbeitspaket	Teilpaket	Beschreibung	Zeitaufwand [h]
Verschlüsselung auf öffentlichem Server einrichten <i>Total Arbeitspaket: 10 h</i>	Serverzertifikat beantragen	Eine verschlüsselte Verbindung stellt sicher, dass zwischen Server und Endnutzer die Daten nicht verändert wurden	1
	Aktuell zu verwendende Verschlüsselungsalgorithmen bestimmen		1
	Serverkonfiguration anpassen		3
	Automatische Zertifikatsaktualisierung einrichten Dokumentation		2 3
Hashwerte der zur Verfügung gestellten Musikdateien zugänglich machen <i>Arbeitspaket vorläufig: 24 h</i>		Die Hashwerte ermöglichen die Überprüfung ob die Dateien korrekt übertragen wurden und dass sie zwischenzeitlich nicht verändert wurden.	
	Mögliche Lösungswege evaluieren	Diese Aufgabe hängt zusammen mit dem Übernahmeprozess bzw. der Darstellung von technischen Metadaten.	16
	Auwandsabschätzung der zwei realistischsten Lösungswege		8
	Detailplanung des Arbeitspaketes erstellen und Aufwandsabschätzung bis zur Realisation		zu viel zum Schätzen

Basis Infrastruktur für Verwaltung und Erhaltungsplanung

Basierend auf der vorgeschlagenen Systemarchitektur aus der Semesterarbeit FS 2016 *Langzeitarchiv für Audiowerke*.

Projekt vorläufig: 164 h

<i>Arbeitspaket</i>	<i>Teilpaket</i>	<i>Beschreibung</i>	<i>Zeitaufwand [h]</i>
Dokumentenverwaltungssystem <i>Total Arbeitspaket: 90 h</i>			
	Anforderungen ausformulieren		6
	Open Source DMS suchen		4
	Gefundene DMS gemäss Anforderungen bewerten		8
	Entscheid mit Stiftungsrat		2
	DMS aufsetzen, Zugänge einrichten und gemäss Anforderungen testen		24

Bestehende Daten migrieren	Die Daten sind auf verschiedenen Computern verteilt. Je Computer 2 h eingerechnet	8
Arbeitsplatzrechner für Synchronisation mit DMS einrichten	4 Arbeitsplatzrechner plus 3 Mobiltelefone	6
Anleitung zum Einrichten der Synchronisation erstellen		8
Dokumentation		24
<p>Issue Tracker / Aufgabenplanungswerkzeug <i>Total Arbeitspaket: 74 h</i></p>		
Anforderungen ausformulieren		8
Open Source Issue Tracker Projektplanungswerkzeuge suchen		4
Gefundene Systeme gemäss Anforderungen bewerten		12
Entscheid mit Stiftungsrat		2

System aufsetzen, Zugangssynchronisation mit Media-Wiki einrichten, gemäss Anforderungen testen

48

Software entwickeln die regelmässig zu erledigende Aufgaben des Managements und der Erhaltungsplanung regelmässig in dieses neue System einträgt

zu viel zum Schätzen

Neues AIP auf Matroska Basis entwickeln

Total Projekt: 520 h

Arbeitspaket	Teilpaket	Beschreibung	Zeitaufwand [h]
AIP Anforderungen verfeinern Arbeiten in Zusammenhang mit dem Matroska Container Format <i>Total Arbeitspaket: 80 h</i>	In Matroska Standard einarbeiten, Formatdefinition lesen		32
	Stand des IETF cellar Projekt untersuchen und weitere Entwicklung abschätzen		8
	Umfang und Ausdrucksmöglichkeiten der Metatags des Matroska Containers, Flac Formats und des PNG For- mats (für Cover Bilder) vergleichen und tabellarisch fest- halten		40
Arbeiten mit die einzubettenden se- mantischen Metadaten <i>Total Arbeitspaket: 80 h</i>		XML Dateien die neben den Metadaten tags die eigentlichen Erhaltungsmetadaten beinhalten. Vorgeschlagen sind derzeit RDF Daten serialisiert als RDF/XML	
	Anforderungen an die einzubettenden Erhaltungsmeta- daten definieren		16
	Libraries und Tools evaluieren, die das einbetten und ex- trahieren von Dateien in Matroska Container erlauben		24
	Die zwei am besten geeigneten Tools/Libraries mit ei- nem Prototypen ausprobieren		40
Unterstützung und Fähigkeiten von typischen Mediaplayern untersuchen und tabellarisch festhalten			

Total Arbeitspaket: 80 h

Die weit verbreiteten Mediaplayer in den Zielgruppen zusammensuchen		8
Mediaplayer auf Testsystemen installieren bzw. Testsysteme suchen (Windows, OS X, Android, iOS)		16
Testdateien zusammenstellen und selber Testdateien für wichtige Features (Dateianhänge, Metadatenunterstützung) erstellen		24
Mediaplayer testen und vergleichen		16
Dokumentieren, tabellarisch Resultate festhalten		16

AIP Definition erstellen
Total Arbeitspaket: 176 h

Mappings zwischen Matroska, Flac und PNG Metatags definieren	Sowohl der Matroska Container wie die enthaltenen Flac (Audio) und PNG (Cover Bild) Daten haben Metadatenfelder. Diese sollen soweit wie möglich und von üblichen Playern unterstützt auch genutzt werden. Diese Felder sind die, die von der allgemeinen Zielgruppe genutzt werden.	40
Struktur und Inhalt der einzubettenden Erhaltungsmetadaten definieren	Dazu gehört auch, ob es eine oder mehrere Dateien sind, welcher MIME Type verwendet wird und der/die Dateinamen.	40
Encoder Einstellungen für die PNG Bilder festlegen	Kompatibilität zu Libraries und Mediaplayer sicherstellen. Nur zum Basis Standard gehörende Features, Merkmale nutzen. Üblichste Einstellungen nutzen.	24
Encoder Einstellungen für die Flac Audiodaten festlegen	Kompatibilität zu Libraries und Mediaplayer sicherstellen. Nur zum Basis Standard gehörende Features, Merkmale nutzen. Üblichste Einstellungen nutzen.	16
Dateinamenkonvention erstellen		16
Vollständige AIP Definition schreiben		40

AIP Konformitätsprüfung entwickeln
Total Arbeitspaket: 104 h

Anforderungen ausformulieren	Ein SIP oder ein erstelltes AIP soll automatisch darauf geprüft werden, ob es den formalen Anforderungen entspricht.	16
------------------------------	--	----

Open Source Werkzeuge suchen	8
Gefundene Werkzeuge gemäss Anforderungen bewerten	40
Dokumentation	16
Aufwand zur Entwicklung und Implementation einer automatisierten Konformitätsprüfung abschätzen	24
Entwicklung und Implementation einer automatisierten Konformitätsprüfung	zu wenig Information zum Schätzen

Übernahmeprozess mit Matroska AIP

Vorläufig Projekt: 0 h

Arbeitspaket	Teilpaket	Beschreibung	Zeitaufwand [h]
Anforderungsanalyse des Übernahmeprozesses durchführen		Vergleichdaten von anderen Institutionen einholen Anforderungen aus OAIS Modell zusammentragen Anforderungen aus Audit Prozess zusammentragen	
Umsetzung planen		Maktroska Werkzeuge evaluieren	
Technische Umsetzung			
Umsetzung im Arbeitsablauf		Schulung Arbeit betreuen	

Migration bestehendes in neues AIP

Die bestehenden AIPs werden so aufbereitet und migriert, dass sie in den Übernahmeprozess eingespeist werden können. Als möglichst vollständiges SIP, das dann nach einer Prüfung freigegeben wird

Projekt vorläufig: 120 h

Arbeitspaket	Teilpaket	Beschreibung	Zeitaufwand [h]
Migration planen und Ablaufdokumentation erstellen			40
Anforderungen an Software zur Migration ausarbeiten Vorbereitungsarbeiten			24
	Software zur Migration entwickeln		zu wenig Information zum Schätzen
	Migration auf Testsystem durchführen und Ergebnis detailliert prüfen		24
Migration durchführen	Alte AIPs und Repräsentationsinformation auf separatem Backup ablegen, Backup prüfen	Angenommen wird, dass die Repräsentationsinformation vorhanden ist.	16
	Protokolldaten der Migration genau prüfen		16
	Fehlende Metadaten in den automatisch generierten SIPs ergänzen	Vollständig migrierbare AIPs werden direkt wieder als AIP abgelegt. Problematische AIPs werden zur weiteren manuellen Bearbeitung zu SIPs migriert.	zu wenig Information zum Schätzen

14.11 Aufgabenstellung Masterthesis

Human Interface Technology

Thesis

Student: Christoph Zimmermann
Advisor: Daniel Debrunner
Semester: full time, 4th semester

Subject Title: **Long time archive for audio works**

Abstract:

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, completing metadata and investigating the copyright status of old audio records, mainly 78 rpms (Shellac records).

The goals of this master thesis are:

- Evaluate and compare the LinkedData ontologies relevant for the scope of the Public Domain Project
- Develop a unified meta data template based on this ontologies using Semantic MediaWiki, migrate existing data and run them on a test system
- Evaluate and implement a semi-automated document management system (DMS) used internally for administration, maintenance, issue tracking
- Enable external secure login to this DMS using state of the art security mechanisms
- Evaluate a trustworthy storage system for the digitized audio works, develop a migration strategy to implement it

Literature:

- [Schrim] Das OAI-Modell für die Langzeitarchivierung, 1st edition 2014, DIN, *Sabine Schrimpf*
- [Keyser] Indexing – From thesauri to the Semantic Web, 1st edition 2012, Chandos Publishing, *Pierre de Keyser*
- [Fürste] Linked Open Library Data, Bibliographische Daten und ihre Zugänglichkeit im Web der Daten, 2011, Dinges & Frick, *Fabian M. Fürste*

Project start: Mo 19.09.2016
Report delivery: Fr. 20.01.2017
Presentation date: Fr 27.01.2017

Assessment: 25 min presentation plus questions/answers
thesis report

Date: 9th September 2016