

Das AVefi-Manual

Grundprinzipien und Richtlinien der Zusammenarbeit von Verbund und Praxis

- Vorabversion 2.0 -

Autoren: Sabrina Silbernagel / Matti Stöhr (TIB); Antje Dittmann /Annette Groschke / Jonas Pitz (SDK); Thomas Ochs / Stefan Matijasevic (FMD); Sven Bingert / Elias Oltmanns (GWDG)

Inhalt

1.	Präambel	1
2.	Grundprinzipien der Zusammenarbeit	1
2.1	Organisationsstruktur im Konsortium	1
2.2	Steuerungsgruppe	2
2.3	Metadatenfachgruppe	2
2.4	Technische Gruppe	2
2.5	Datengeber	2
2.6	Weitere Rollen	3
3.	Aggregation der Metadaten	3
3.1	Allgemeines	3
3.2	AVefi-Datenmodell	3
3.3	Schema - Standardorientierung	5
4.	AVefi-Rechercheplattform	5
4.1	Zielsetzung und Funktionen	5
4.2	Redaktionssystem und Versionierung	6
4.3	Weitere Entwicklungen	6
5.	Konventionen und Empfehlungen zu den Kerndaten	6
5.1	Das Filmwerk	6
5.2	Titel und Archivtitel.....	7
5.3	Produktionsort.....	8
5.4	Jahr.....	8
5.5	Regieangaben.....	8
5.6	Identifikatoren und Normdaten	9
5.7	Kontrollierte Vokabulare in AVefi.....	9
5.8	Werksabgleich beim Import	9
6.	Bestandsauswahl	11
6.1	Allgemeines	11
6.2	Kriterien.....	11
7.	Datenübergabe	11
7.1	Allgemeines	11

7.2	Lizenzierung	11
7.3	Datenlieferung per Export.....	12
7.4	Datenlieferung über eine Schnittstelle	12
7.5	Formate	12
7.6	Arbeitsschritte zur Datenlieferung	12
7.7	How-to und Mapping-Vorlagen.....	13
8.	Technische Infrastruktur	13
8.1	Allgemeines	13
8.2	Das PID-System	14
8.3	Die Type-Registry	14
8.4	Änderungen des technischen Schemas	14
8.5	Schnittstellen und Datenaustausch	15
9.	Kontakt / Projektteam.....	16
10.	Weitere Informationen.....	16

1. Präambel

Im DFG-Projekt AV-EFI - Automatisiertes Verbundsystem für audiovisuelle Bestände über einheitliche Filmidentifikatoren (1. Phase von 11/2023 bis 10/2025 – Eigenschreibweise mit Projektstart *AVefi*) wird ein institutionenübergreifendes, webbasiertes System mit Persistent Identifier (PID) in zentraler Rolle entwickelt. Dieses soll der automatisierten Identifikation und Verknüpfung von Filmen, ihren Fassungen und Exemplaren aus Beständen verschiedener Institutionen dienen. Das PID-System der einheitlichen Filmidentifikatoren (*efi*¹) übernimmt dabei die Funktion einer zentralen Datenbank und führt heterogene filmografische Metadaten aus originalen Datenbanken verschiedener filmhaltender Institutionen strukturiert zusammen. Ziel ist die freie, langfristige Verfügbarkeit und flexible Nachnutzbarkeit dieser Daten für die Öffentlichkeit, insbesondere für die filmgetriebene, wissenschaftliche Forschung und Lehre gleich welcher Disziplin. Weiterhin entsteht ein nationales Findmittel der einzelnen Bestandteile audiovisueller Werke an verschiedenen Institutionen, welches Digitalisierungsprojekte, Langzeitarchivierung und Disastermanagement erleichtern kann. Das entstehende Verbundsystem, einschließlich einer benutzerfreundlichen Recherche- und Redaktionsoberfläche, soll das bereits verfügbare „Ökosystem“ filmischer Identifikatoren und Normdaten (z.B. Filmportal, EIDR, GND, NFDI4Culture) sinnvoll ergänzen und sich in dieses als originäres Angebot mit starkem Bestandsbezug einfügen. Langfristig soll das von Automatisierungen im Metadatenmanagement geprägte System dem Nachweis und der Verknüpfung von AV-Beständen jeglicher Herkunft und Gattung dienen. Im Projekt richtet sich ein besonderer Blick auf heterogenes Filmmaterial wie Unterrichts-, Lehr- und Wissenschaftsfilme oder Amateurfilme. Diese Gattungen sollen erstmals in einem Verbundsystem neben dem „großen“ Film als „kleiner“ Film der Forschung zur Verfügung stehen und bergen besondere Herausforderungen.

Das vorliegende Richtliniendokument in Gestalt eines kontinuierlich iterativ fortgeschriebenen Manuals beschreibt und regelt die wesentlichen, auf bestehenden Standards und Infrastrukturen aufbauenden Rahmenbedingungen, Prinzipien und Workflows in der Zusammenarbeit von Verbundpartnern und Praxispartnern. Es bietet einen Überblick und Orientierung für alle in *AVefi* aktiven Institutionen und Personen und berücksichtigt die Forschungsperspektiven.

Im Projektverlauf werden Vorabversionen des Manuals produziert und öffentlich bereitgestellt. Zum Projektabschluss ist eine persistent zitierfähige Publikation in der Version 1 vorgesehen.

2. Grundprinzipien der Zusammenarbeit

2.1 Organisationsstruktur im Konsortium

Der Begriff „Konsortium“ wird in diesem Zusammenhang verwendet, um die Art der Kollaboration und Zusammenarbeit sowohl zwischen den Verbundpartnern als auch den Praxispartnern zu benennen.

Das Konsortium übernimmt die Verantwortung für den Betrieb des Verbundsystems als Einheit. Es arbeitet gemeinsam daran, die Richtlinien für die konkreten Aktivitäten im Projekt voranzutreiben. Um eine effiziente Organisation sicherzustellen, gibt es eine klare Einteilung in verschiedene Rollen und Gruppen. Diese sind mit den Verbundpartnern und ersten Praxispartnern besetzt. Interessierte Praxispartner sind eingeladen, an den Aktivitäten in den einzelnen Gruppen mitzuwirken. Die Aufgaben und Abläufe innerhalb der Gruppen und Rollen werden, wenn nötig, im Laufe des Projekts weiter angepasst bzw. geschärft.

Weiterhin ist die Organisationsstruktur Teil der Verstetigungsabsicht, das heißt, das Konsortium strebt an, langfristig und nachhaltig zu agieren und mittelfristig weitere Partner als Mitglieder zu gewinnen.

¹ Diese Einheitliche Filmidentifikatoren sind die im Rahmen des Projektes entwickelten „speziell auf Handle basierten, Persistent Identifier (PIDs) und werden im Folgenden als *efi* bezeichnet.

2.2 Steuerungsgruppe

Die Steuerungsgruppe besteht aus den Verbundpartnern und trifft grundsätzliche, interne Entscheidungen rund um das Projekt. Sie hat die Organisationsstruktur mit ihren unterschiedlichen Gruppen und Rollen definiert. Des Weiteren werden in der Steuerungsgruppe Aktivitäten in den Bereichen Metadatenstandard und Infrastruktur besprochen und gestartet, um sicherzustellen, dass alle Mitglieder des Konsortiums auf dem gleichen Stand sind und reibungslos zusammenarbeiten können.

Die Gruppe organisiert regelmäßig stattfindende Veranstaltungen und Meetings, um den Austausch zwischen den Partnern zu fördern. Darüber hinaus repräsentiert die Steuerungsgruppe das Konsortium auf verschiedenen Veranstaltungen nach außen und betreibt Öffentlichkeitsarbeit, um das Projekt bekannt zu machen und potenzielle Partner anzusprechen.

Im Falle einer Verstärkung des Projekts, kann die Steuerungsgruppe weitere Mitglieder hinzufügen oder Mitglieder ggf. entfernen.

Ansprechpartner: Sabrina Silbernagel, Matti Stöhr (TIB)

2.3 Metadatenfachgruppe

Die Metadatenfachgruppe ist dafür verantwortlich, die Bedürfnisse und Anforderungen der verschiedenen Mitglieder des Konsortiums zu sammeln, um das Metadatenschema unter Beachtung nationaler und internationaler Standards erweitern zu können. Sie nimmt Änderungsvorschläge entgegen und entscheidet über deren Umsetzung.

Außerdem entwickelt sie Prinzipien zur Disambiguierung und Duplikatsprüfung beim Datenimport und legt Prinzipien der Metadatenredaktion und Versionierung fest. Die Fachgruppe überprüft die inhaltlich-fachliche Umsetzbarkeit und stimmt die technische Umsetzung mit den relevanten Partnern im Konsortium ab.

Dabei bezieht die Fachgruppe Praxispartner gleichberechtigt in den Prozess ein, indem sie diese aktiv in die Entwicklung des Projekts einbindet.

Die Mitglieder der Metadatenfachgruppe gehören in der Regel auch der Gruppe der Datengeber sowie der nutzenden Fachcommunity an.

Ansprechpartner: Annette Groschke (SDK), Thomas Ochs (FMDU)

2.4 Technische Gruppe

Die technische Gruppe besteht aus den Teilbereichen Backend- und Frontend-Entwicklung. Die Gruppe implementiert die aus der Metadatenfachgruppe abgestimmten und spezifizierten Anforderungen. Darüber hinaus betreibt sie sowohl die technische PID-Infrastruktur als auch ein nutzerzentriertes Frontend (Rechercheoberfläche) und entwickelt diese jeweils weiter.

Dies geschieht in enger Abstimmung zwischen den beiden Entwicklungsbereichen und mit der Metadatenfachgruppe.

Ansprechpartner: Backend: Elias Oltmanns, Sven Bingert (GWDG) – Frontend: Stefan Stretz (TIB)

2.5 Datengeber

Im Unterschied zu den anderen Gruppen sind hier keine Einzelpersonen, sondern Institutionen vertreten. Mitglieder der Gruppe sind berechtigt, *efi* zu erzeugen und dem Verbundsystem Metadaten zu liefern. Dies kann zum einen durch eine automatisierte Schnittstelle und der Implementierung der erzeugten *efi* im eigenen System geschehen, zum anderen können Metadaten auch durch Exporte aus dem jeweiligen Quellsystem zur Verfügung gestellt werden.

Bei einer Verstärkung des Projekts können weitere Institutionen in die Datengeber-Gruppe aufgenommen werden. Interessenten werden hierbei in ihrer Entscheidungsfindung inhaltlich und technisch unterstützt. Die Aufnahme neuer Datengeber erfolgt nach Entscheidung durch die Steuerungsgruppe.

Ansprechpartner: Technisch: Sven Bingert (GWDG) – Inhaltlich: Antje Dittmann (SDK)

2.6 Weitere Rollen

Neben den oben genannten Gruppen gibt es zwei weitere Rollen, die nicht Teil des Konsortiums sind.

Beratende Fachcommunity

Die beratende Fachcommunity liefert mit ihrer Fachexpertise wichtige Impulse für die Weiterentwicklung der Plattform.

Experten ausgewählter Fachcommunity-Institutionen bringen sich mit ihrer Expertise in den Themenbereichen Werksdisambiguierung, Metadatenmanagement Amateurfilm sowie Anbindung an übergreifende Normdatensysteme ein.

Ansprechpartner: Sabrina Silbernagel, Matti Stöhr (TIB)

Nutzende Fachcommunity

Die Gruppe setzt sich aus Vertretern der interdisziplinären wissenschaftlichen Community bzw. den potenziellen Nutzern der Plattform zusammen.

Sie gibt Hinweise auf Bedürfnisse und Anforderungen der Community in Bezug auf Usability und Recherchemöglichkeiten. Des Weiteren testet die Gruppe in einem iterativen Prozess die Funktionen der Rechercheoberfläche, und gibt Rückmeldungen, die in den Entwicklungsprozess einfließen.

Darüber hinaus agieren Fachcommunity-Institutionen als Multiplikatoren für die nutzende Community.

Ansprechpartner: Organisatorisch: Sabrina Silbernagel, Matti Stöhr (TIB) – Technisch: Stefan Stretz (TIB)

3. Aggregation der Metadaten

3.1 Allgemeines

Es gelten die FAIR-Prinzipien für Forschungsdaten: „Ziel ist es, dass Forschungsdaten für Menschen und Maschinen optimal aufbereitet und zugänglich sind. Dies impliziert aber nicht, dass jeder Datensatz uneingeschränkt nachnutzbar ist. Vielmehr zielen die FAIR-Prinzipien darauf, im Rahmen des rechtlich und technisch möglichen, Datenbestände für neue Nutzungsszenarien zu öffnen. Mit der Anwendung der FAIR-Prinzipien soll die Wiederverwendbarkeit von Datenbeständen verbessert werden.“²

Zentrales Ziel des Projekts ist maximale Effizienz, die durch ein hohes Automatisierungslevel zu erreichen ist: Der Up- und Download soll standardmäßig per Schnittstelle erfolgen. Die Kompatibilität der Metadaten wird durch Verwendung standardisierter Formate und Normdaten angestrebt.

Zusätzliche Vernetzung wird durch die Verwendung weiterer Identifikatoren-Systeme unterstützt und angestrebt. Die Metadaten werden durch Schnittstellen und Exporte offen lizenziert als Open Data nachnutzbar – offene Formate stehen zur freien Verfügung.

3.2 AVefi-Datenmodell

Das erarbeitete Datenmodell ist darauf ausgelegt, Sammlungsbestandsdaten eindeutig einer haltenden Institution zuzuordnen und gleichzeitig Beziehungen zwischen den Beständen unterschiedlicher Institutionen darzustellen. Dazu werden *efi* für verschiedene Klassen von Objekten mit jeweils eigenen Anforderungen

² <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>. Diese und alle folgenden Netzquellen wurden am 25.02.2025 auf Zugreifbarkeit geprüft.

erzeugt. So wird unterschieden zwischen (Film)werken, Manifestationen und Exemplaren (Items) im Sinne des Regelwerks der FIAF:

- Das **Werk** ist eine Entität, die den geistigen oder künstlerischen Inhalt und den Prozess der Realisierung in einem kinematografischen Medium umfasst, z.B. wie der Film heißt, wann er gedreht wurde, wer ihn gedreht hat, wer darin zu sehen war, worum es geht usw.
- Unter einer **Variante** wird eine Fassung eines Werkes verstanden, die durch entsprechende Ausarbeitung oder bearbeitende Gestaltung (u.a. Kürzung, Neuschnitt, Synchronisierung) entstanden ist. Die Grenzen zwischen einer neuen Variante und einem neuen Werk sind fließend. In einer dreistufigen Hierarchie kann bei geringfügigen Änderungen am Werk statt einer Variante eine Manifestation gebildet werden, daher ist die Beschreibung von Varianten im Kontext AVefi optional.
- **Manifestationen** sind die Verkörperung eines Filmwerks bzw. einer Variante. Manifestationen umfassen alle analogen, digitalen und Online-Medien. Informationen der Ebene Manifestation enthalten üblicherweise eine Beschreibung dessen, was die zugehörigen Exemplare idealerweise ausmachen sollte, unabhängig von den im Archiv vorhandenen Exemplaren. Ein Beispiel: Die ursprüngliche Originallänge eines Films beträgt 90 Minuten, aber in dem im Archiv vorhandenen Exemplar fehlt Filmmaterial, sodass es kürzer ist.
- **Exemplare** sind das physische Produkt, also konkrete Kopien, einer Manifestation eines Werks bzw. einer Variante. Ein Exemplar kann aus einer oder mehreren Komponenten bestehen, d.h. das gesamte Exemplar kann aus einer Rolle oder fünf Spulen, zwei VHS-Kassetten oder einer DVD bestehen. Ein Exemplar-Datensatz kann bei Bedarf Felder oder Bereiche für separate Barcodes und Zustandsinformationen für jede Komponente des Exemplars (z.B. jede Spule) enthalten. Das Exemplar kann vollständig, unvollständig oder ein Fragment sein. Bei rein digitalen Medien ist ein Exemplar definiert als die Verfügbarkeit der Datei, unabhängig von der Anzahl der eventuell vorhandenen Sicherungskopien.
- **Wichtig:** Das AVefi-Datenmodell orientiert sich grundsätzlich am FIAF-Manual, weicht jedoch in der Beschreibung von Manifestationen und Exemplaren bewusst leicht von dessen Empfehlungen ab. Aus Gründen der Datenreinheit wurde entschieden, technische Eigenschaften wie Farbe, Ton oder Dauer ausschließlich auf einer Ebene zu erfassen. Da diese Merkmale nach Erfahrung der Praxispartner in der Regel am Exemplar dokumentiert werden und aufgrund des starken Bestandsbezuges der AVefi-Plattform, sind sie im AVefi-Modell ebenfalls dort verortet. Das bedeutet konkret: Nicht die idealtypischen technischen Merkmale der Manifestation, sondern die tatsächlichen Eigenschaften der konkreten Kopie werden erfasst.

Grundsätzlich kann ein Exemplar im AVefi-Kontext nicht für sich stehen und benötigt ein Werk und eine Manifestation, um vollständig dreistufig erfasst zu sein. Umgekehrt sollten keine Werke und Manifestationen erfasst werden, denen keine Exemplare zugeordnet sind.

Bei datenliefernden Institutionen ohne dreistufige Unterscheidung erfolgt durch die Überführung ins AVefi-Schema die Umwandlung in die drei Ebenen.

Der *efi* für eine Manifestation oder ein Exemplar bezieht sich bei AVefi immer auf den Bestand einer einzelnen teilnehmenden Institution. *Efi* für Exemplare werden zwingend mit einem *efi* für die Manifestation verknüpft, zu der sie gehören. Hält jedoch eine weitere Institution exakte Kopien einiger Exemplare, so ist für den Bestandsnachweis eine Registrierung jeweils eigener *efi* für Exemplare und zugehörige Manifestation durch die zweite Institution vorgesehen. In den Metadaten kann auf die Kopie der jeweils anderen Institution verwiesen werden, aber die Exemplare der beiden Institutionen sind immer durch einen eigenen PID identifizierbar.

Im Gegensatz dazu strebt AVefi auf der Ebene des Filmwerks einheitliche Identifikatoren an. Das heißt, *efi* für Filmwerke sollen ausdrücklich eindeutig sein und von allen gemeinsam genutzt werden. Insbesondere können Manifestationen aus den Beständen unterschiedlicher Institutionen mit demselben Werk verknüpft sein. Entsprechend können und sollen die Metadaten auf Werksebene von allen Beteiligten ergänzt und korrigiert werden.

3.3 Schema - Standardorientierung

Das in diesem Projekt entwickelte AVefi-Schema³ orientiert sich an dem FIAF Moving Image Cataloguing Manual⁴ und an der Gemeinsamen Normdatei der Deutschen Nationalbibliothek⁵. Das Metadatenschema ist zudem an FIAFcore⁶ und EBUCore⁷ angepasst. Insbesondere die Nähe zum FIAF-Manual bietet den Zugang zu einer umfassenden und frei verfügbaren Dokumentation. Diese bezieht auch andere Standards wie EN 15744⁸ und EN 15907⁹ mit ein. Auf diese Weise wird sichergestellt, dass das AVefi-Schema sowohl möglichst nachvollziehbar als auch kompatibel mit anderen etablierten Standards ist.

FIAFcore ist eine RDF-Ontologie, die auf dem FIAF-Manual basiert und für einen Datenaustausch zwischen filmhaltenden Institutionen entwickelt wird. Das AVefi-Schema ist so konzipiert, dass es sich direkt auf FIAFcore mappen lässt. Filmobjekte werden mit Hilfe der drei Ebenen Werk, Manifestation und Exemplar abgebildet. Varianten beschreiben zusätzlich Fassungen von Werken, die sich nur geringfügig vom Originalwerk unterscheiden. Nach FIAFcore werden Werke und Varianten als WerkVariante auf einer Stufe erfasst. Verschiedene Ereignisse, wie das Produktionsereignis, beschreiben Prozesse im Lebenszyklus des Filmobjekts, zum Beispiel die Entstehung des Werks. An den Ereignissen können Personen und andere Agenten durch verschiedene Aktivitäten wie Regie oder Produktion beteiligt sein.

Dazu kommen kontrollierte Vokabulare, zum Beispiel für Filmgattungen, Filmcredits oder Datenträger, die sich an FIAFcore und anderen etablierten Vokabularen orientieren und in der Metadatenfachgruppe diskutiert und erweitert werden. Wenn möglich, wird auch auf Normdateien zurückgegriffen, zum Beispiel die GND für Personen und Schlagworte.

Ein Überblick, Erläuterungen und Beispiele zum Schema finden sich in der Dokumentation unter folgendem Link: <https://av-efi.github.io/av-efi-schema/>.

4. AVefi-Rechercheplattform

4.1 Zielsetzung und Funktionen

Das Projekt AVefi verfolgt das Ziel, Filmbestände institutionsübergreifend **sichtbar, durchsuchbar und vergleichbar** zu machen. Die Rechercheplattform dient als zentrale Anlaufstelle, auf der Nutzer*innen Daten zu Werken, Manifestationen und Exemplaren **finden, entdecken und miteinander in Beziehung setzen** können.

Bereits in der Alpha-Phase bietet die Plattform folgende Kernfunktionen:

- **Facettennavigation und Suchfilter** ermöglichen eine gezielte Eingrenzung und kombinierbare Suchstrategien.
- **Unterschiedliche Ergebnisansichten** (z. B. Listenansicht) unterstützen flexible Nutzungsszenarien.
- Eine **Detailansicht** zeigt die Metadaten zu Werken, Manifestationen und Exemplaren in strukturierter Form.
- **Exportmöglichkeiten:** Einzelne Datensätze sowie angelegte Favoriten-Listen können exportiert und außerhalb der Plattform weiterverwendet werden.
- Über eine **Merkliste** können Werke gesammelt und jederzeit wieder aufgerufen werden.

³ <https://github.com/AV-EFI/av-efi-schema>

⁴ <https://www.fiafnet.org/pages/E-Resources/Cataloguing-Manual.html>

⁵ https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html

⁶ <https://fiafcore.org>

⁷ <https://tech.ebu.ch/metadata/ebucore>

⁸ http://filmstandards.org/fsc/index.php/EN_15744

⁹ http://filmstandards.org/fsc/index.php/EN_15907

- Die **Vergleichsansicht** ermöglicht es, Metadaten zu einem Werk nebeneinander darzustellen und Unterschiede hervorzuheben.
- Wo möglich, bestehen **Verlinkungen zu den Quelldatenbanken** der Daten liefernden Institutionen, um weiterführende Informationen zugänglich zu machen.
- Aufgrund des **diversen Zielpublikums** werden derzeit verschiedene **Nutzungsmodi** erprobt, die sich dynamisch an das jeweilige **Nutzungsverhalten und die Vorerfahrung der Nutzer*innen** anpassen.

Die **Alpha-Phase** startete am 28. Mai 2025. Seitdem ist die Plattform unter <https://www.av-efi.net/> frei zugänglich. Da die Weiterentwicklung laufend erfolgt, sind noch nicht alle vorgesehenen Funktionen vollständig verfügbar.

4.2 Redaktionssystem und Versionierung

Parallel zur Rechercheplattform wird ein **Redaktionssystem** aufgebaut, mit dem die **Bearbeitung von Metadaten** und die **Zusammenführung identischer Werke** ermöglicht werden soll.

Geplante Funktionen sind:

- Eine **Datensatzhistorie**, die nachvollziehbar macht, wer wann welche Änderungen vorgenommen hat.
- Die Möglichkeit, **inaktive Datensätze** (z. B. Manifestationen oder Exemplare nach Kassation) zu kennzeichnen, ohne diese aus der Recherche zu entfernen.
- **Tombstone-Seiten** für zusammengeführte oder gelöschte Datensätze. Diese sichern die Nachvollziehbarkeit und garantieren, dass auch veraltete Einträge weiterhin auffindbar und zitierfähig bleiben.

Die Umsetzung einer umfassenden **Versionierung** ist für das Folgeprojekt vorgesehen und wird in späteren Dokumentationen detailliert beschrieben.

4.3 Weitere Entwicklungen

Einige Funktionen befinden sich derzeit in der Erprobung oder sind geplant:

- **Merge-Logik:** Erste Ansätze für das automatische Zusammenführen von Datensätzen werden aktuell getestet.
- **Manuelle Korrekturen:** Die Möglichkeit, falsch zusammengeführte Werke wieder zu trennen, ist vorgesehen, aber noch nicht genauer definiert.

5. Konventionen und Empfehlungen zu den Kerndaten

5.1 Das Filmwerk

Bestehende Werks-Identifikatoren wie EIDR, ISAN oder Filmportal UID erleichtern den automatisierten Abgleich und sollten daher mitgeliefert werden.

Werk-Aggregate (z.B. Sammel-DVDs) werden entweder als eine Manifestation, verknüpft mit mehreren Werken, dargestellt oder als ein Sammelwerk mit mehreren Werken in Teil-von-Beziehung und einer Manifestation am Sammelwerk.

Da der Fokus auf der Registrierung von Beständen liegt, wird wenig Konsolidierung vor dem Upload vorgegeben. Der Prozess soll niedrigschwellig gestaltet werden, wobei der Minimalprozess lediglich das Erzeugen eines *efi* umfasst. Es besteht kein Anspruch auf Vollständigkeit der Metadaten, für die Erzeugung eines *efi* sind nur wenige Kernfelder erforderlich.

AVefi orientiert sich an dem Metadatenstandard EN 15744¹⁰. Dessen Pflichtfelder sind:

- der Titel des Werks und
- die datenliefernde Institution

Alle Datensätze, die den Mindestanforderungen genügen, erhalten einen *efi*. Eine eindeutige Unterscheidbarkeit zwischen den Datensätzen ist gegeben, wenn Informationen zu den *Kerndaten* **Titel, Produktionsort, Jahr** und **Regie** vorhanden sind.

5.2 Titel und Archivtitel

Idealerweise sollte für Werke der **Originaltitel** angegeben werden. Alternativ sind auch andere Titel, wie beispielsweise der Verleihtitel, zulässig. Untertitel werden unverändert übernommen. Die Sprache des Titels muss nicht angegeben werden; sie wird nicht auf der AVefi-Rechercheplattform angezeigt.

Der Originaltitel entspricht dem Preferred title. Der Verleihtitel wird als alternative Bezeichnung angezeigt.

- Originaltitel -> Preferred title
- Verleihtitel -> Alternative title

Folgennummern, beispielsweise bei Folgen einer Serie, werden im Titel vermerkt. Dies wird in der Praxis getestet, vorläufig werden keine zusätzlichen Felder für Folgennummern eingeführt. Bei Serienwerken ist auf die Disambiguierung der einzelnen Folge durch Folgennummern am Episodentitel zu achten.

Artikel werden vorangestellt. Zusätzlich kann ein Sortiertitel mit angegeben werden, bei dem der Artikel mit Komma abgetrennt an den Titel hintenangestellt wird.

Sollte **kein Titel** vorhanden sein, wird ein Archivtitel gebildet. Dazu wird folgende Konvention empfohlen, die sich an den FIAF-Empfehlungen orientiert:

Archivtitel setzen sich zusammen aus:

- **Name der Quelle oder Sammlung:** geben Sie hier den Namen der Filmsammlung an, falls vorhanden
- **Kategorie:** Falls gewünscht, kann an zweiter Stelle des Archivtitels eine Kategorie genannt werden, um die Auffindbarkeit z.B. aller Betriebsausflüge zu erleichtern. Infrage kommende Kategorien wären z.B.: Betriebsausflug, Familienfilm [= Home Movie], Filmtagebuch [Institutionsspezifische Bezeichnung (Haus des Dokumentarfilms) für Aufnahmen aus dem Zweiten Weltkrieg, die von Wehrmachtsoldaten hergestellt wurden.], Fragment, Kompilation, Privataufnahme [Jeder Familienfilm ist eine Privataufnahme, aber nicht jede Privataufnahme ist ein Familienfilm. Beispiel: Katzenvideo], Reise-film [selbstgedrehte Aufnahmen einer privaten Reise]
- **Wer oder was:** im Film gezeigte Personen, Ereignisse (Skifahren, Bügeln, Zugfahrt, Einschulung etc.) oder Objekte (Auto, Fahrrad, elektrische Eisenbahn)
- **Wo:** Ort
- **Wann:** Zeitspanne - zusätzlich zur Verzeichnung des Produktionszeitraums in den dafür vorgesehenen Datenfeldern kann im Archivtitel z.B. ein genaues Datum (TT-MM-JJJJ) verzeichnet werden oder das Jahr der Aufnahme kann hier wiederholt genannt werden, wenn gewünscht (Weihnachten 1974).

Aus technischen Gründen ist die **Zeichenlänge** für Titel auf 250 Zeichen begrenzt. Bei Überschreitung erfolgt ein Hinweis an die datenliefernde Institution.

Beispiele für Titelangaben:

- *Originaltitel:* Metropolis
- *Originaltitel:* Das Boot
- *Sortiertitel:* Boot, Das.
- *Episodentitel:* Die Wespe. 2
- *Archivtitel:* Privataufnahmen Vergnügungspark USA, 1940er Jahre

¹⁰ http://www.filmstandards.org/fsc/index.php/EN_15744

- *Archivtitel:* Sammlung Max Mustermann, Reisetagebuch, Dresden, 1970-1971

5.3 Produktionsort

Üblicherweise wird das Land angegeben, in dem die Produktionsfirma des Werks ihren Geschäftssitz hat. Bei Amateurfilmen besteht eine höhere Wahrscheinlichkeit, dass verschiedene Werke beispielsweise gleiche Titel und Jahresangaben aufweisen („Weihnachten, 1974“). Daher kann hier für den Produktionsort eine genauere Angabe, z.B. der Stadt, sinnvoll sein. Zudem wird empfohlen, bei der Angabe von Städten auch das Land zu erfassen, um eine spätere Facettensuche und Filterfunktionen zu erleichtern. Die Anreicherung von geographischen Daten mit Identifikatoren des Getty Thesaurus of Geographic Names (TGN) wird empfohlen, so dass ein Match auch mit unterschiedlicher Schreibweise stattfinden kann. Ein mit dem Wert „unbekannt“ belegter Produktionsort gilt als unbelegt.

Beispiele für Ortsangaben:

- *Produktionsort:* Bundesrepublik Deutschland
- *Produktionsort:* Frankreich (TGN ID: <http://vocab.getty.edu/page/tgn/7008038>)
- *Produktionsort:* Dresden
Produktionsort: Deutschland (TGN ID: <http://vocab.getty.edu/page/tgn/7000084>)
- *Produktionsort:* Unbekannt

5.4 Jahr

Ein Jahr oder ein vollständigeres Datum ist für die Identifizierung eines Filmwerks unerlässlich. Im Rahmen von AVefi wird das Produktionsjahr und/oder das Veröffentlichungsdatum zum Abgleich der Werke verwendet. Datumsangaben richten sich nach dem ISO 8601-2019 Standard, speziell nach dem dort integrierten Extended Date/Time Format (EDTF) der Library of Congress. Zugelassen sind alle Schreibweisen der Stufe o sowie die Qualifikatoren „?“ für ein ungewisses Datum (plus/minus 5 Jahre) und „~“ für einen ungefähren Zeitpunkt (plus/minus 1 Jahr). Ein mit „unbekannt“ belegtes Datumsfeld gilt als unbelegt. Eine Abweichung von plus/minus 1 Jahr wird im Abgleich toleriert.

Beispiele für Datumsangaben:

- 2015-04-24: Vollständiges Datum, also Jahr, Monat und Tag
- 2015-04: Jahr und Monat
- 2015: Jahr
- 2015~: Ungefähr 2015, an einem Zeitpunkt zwischen 2014 und 2016
- 2015?: Ungewisses Datum 2015, an einem Zeitpunkt zwischen 2010 und 2020
- 2015/2016: Ein Zeitraum von 2015 bis 2016
- 2015-04-24/2016-06: Ein Zeitraum vom 24.04.2015 bis Juni 2016

5.5 Regieangaben

Die Regie wird nach dem Schema „Nachname, Vorname“ eingetragen. Die Verwendung von Identifikatoren der GND wird empfohlen, da diese bei der Werksdisambiguierung von Vorteil sind, z.B. im Fall von unterschiedlichen Schreibweisen eines Namens im Freitextfeld. Ein mit „unbekannt“ ausgewiesener Wert gilt als unbelegt. Mehrfachnennungen sind möglich.

Beispiele für Regieangaben:

- *Regie:* Troller, Georg Stefan
- *Regie:* Petzold, Christian (GND ID: <http://d-nb.info/gnd/134218272>)
- *Regie:* Wachowski, Lana (GND ID: <http://d-nb.info/gnd/122552571>)
Regie: Wachowski, Lilly (GND ID: <http://d-nb.info/gnd/122552601>)
- *Regie:* Unbekannt

5.6 Identifikatoren und Normdaten

Die Metadaten aller Verbund- und Praxispartner referenzieren ganz im Sinne der Linked Open Data etablierte Identifikatoren- und Normdatensysteme wie etwa die GND für Personen oder der TGN für Orte. Um die Sichtbarkeit im GND-Explorer zu erhöhen, sind Identifikatoren zu Personen, Orten und Schlagworten sinnvoll. Die Verwendung von Normdaten mit eindeutigen Identifikatoren wird nicht vorausgesetzt, ist aber nachdrücklich erwünscht, um stabilere Abgleichroutinen zu ermöglichen.

Für **Personendaten** und **Schlagworte** wird die Verwendung von GND-Daten mit entsprechender Referenzierung empfohlen: <https://lobid.org/gnd>. Pro Werk können technisch bedingt 99 Schlagworte geliefert werden.

Weitere für Personendaten unterstützte Normdatenbanken in AVefi sind: Filmportal, Virtual International Authority File (VIAF) sowie Wikidata. Diese Liste kann bei Bedarf erweitert werden.

Für **Datumsangaben** wird der EDTF Standard der Library of Congress verwendet: <https://www.loc.gov/standards/datetime/>.

Für **geografische Daten** wird der Getty Thesaurus of Geographic Names (TGN) genutzt: <https://www.getty.edu/research/tools/vocabularies/tgn/>.

5.7 Kontrollierte Vokabulare in AVefi

Die kontrollierten Vokabulare finden sich in der AVefi-GitHub-Dokumentation unter Enumerations (https://av-efi.github.io/av-efi-schema/schema_overview/#enumerations). Relevante Enumerations sind zum Beispiel Werkstypen, Titeltypen, Credits, beteiligte Parteien (Personen & Körperschaften), Sprache, Format. Das Vokabular basiert in erster Linie auf den Empfehlungen des FIAF Manuals. In einigen Fällen, zum Beispiel beim Format und bei Gattungen, wurde das Vokabular um Begriffe ergänzt, die sich während der Testphase als besonders gebräuchlich erwiesen haben.

5.8 Werksabgleich beim Import

Für einen automatisierten Abgleich von Werksdatensätzen hat sich das Konsortium auf die vier Kernfelder **Titel**, **Produktionsort**, **Jahr**, **Regie** festgelegt. Dabei ist nur das Titelfeld ein Pflichtfeld für die Datenübernahme, die weiteren Kernfelder sind optional. Der Abgleich nutzt neben den vier Feldern auch einen gegebenenfalls vorhandenen Identifikator am Werksdatensatz. Für jedes Kernfeld werden gewisse vordefinierte Toleranzen (Schwellenwerte) zugelassen.

Ein **Match** im System findet dann statt, wenn die Inhalte der vier Datenfelder Titel, Produktionsort, Jahr, Regie übereinstimmen. Haben zwei Werksdatensätze denselben **Identifikator** (externe Identifikatoren wie z.B. die Filmportal UID oder *efi*), gilt dies als automatische Übereinstimmung.

Beim automatischen Abgleich von Werken werden folgende geringfügige Abweichungen zugelassen:

- beim **Titel**: Groß- und Kleinschreibung, Sonderzeichen, doppelte Buchstaben, verbundene Untertitel, Episodentitel
- beim **Produktionsjahr** des Werkes plus/minus ein Jahr
- bei den **Produktionsorten** muss **mindestens eine** Übereinstimmung zwischen beiden Werksdatensätzen vorhanden sein, entweder über eine Zeichenkette oder einen hinterlegten Identifier.
- Für einen Match muss in zwei Werksdatensätzen mindestens ein genannter **Regisseur** übereinstimmen, ebenfalls entweder über eine Zeichenkette oder einen hinterlegten Identifier.

Beispiele:

- Wenn der Name im Feld 'Regie' abweichend geschrieben und gleichzeitig keine GND-ID vorhanden ist, findet kein Match statt. Beide Datensätze bleiben erhalten und können manuell im Redaktionssystem gemergt werden.

Beispiel:

Titel: Undine
Produktionsort: Deutschland
Produktionsjahr: 2019-2019
Regie: Petzoldt, Christian

Titel: Undine
Produktionsort: Deutschland
Produktionsjahr: 2019-2019
Regie: Petzold, Christian

- Trotz abweichender Schreibweise des Namens im Feld „Regie“ bei zwei Werksdatensätzen kommt es zu einem Match, wenn beide Einträge die gleiche GND-ID haben. Beide Schreibweisen werden im resultierenden Werksdatensatz verzeichnet und können händisch im Redaktionssystem nachgearbeitet werden.

Beispiel:

Titel: Smultronstället
Produktionsort: Schweden
Produktionsjahr: 1957-1957
Regie: Bergmann, Ingmar (<https://d-nb.info/gnd/118509519>)

Titel: Smultronstället
Produktionsort: Schweden
Produktionsjahr: 1957-1957
Regie: Bergman, Ingmar (<https://d-nb.info/gnd/118509519>)

Haben zwei Werksdatensätze für **Orte** oder **Regie** dieselben **Identifikatoren**, gilt dies als automatische Übereinstimmung der jeweiligen Felder.

Folgende Fälle können beim Import von Werken auftreten:

- 1. Fall: **Match + No Update**: zwischen dem neu gelieferten Werksdatensatz und den bereits vorhandenen Werken im System konnte eine 100-prozentige Übereinstimmung festgestellt werden. Ein Update der Werksmetadaten ist daher nicht nötig.
- 2. Fall **No Match + Create**: zwischen dem neu gelieferten Werksdatensatz und den bereits vorhandenen Werken im System konnte keine Übereinstimmung festgestellt werden. Es findet eine Rückmeldung statt, dass ein neues Werk angelegt wurde.
- 3. Fall **Match + Update 1**: Das Werk wurde bereits im System registriert; dasselbe Werk wird noch einmal vom selben Datengeber geliefert und gematcht. Die neu gelieferten Metadaten ersetzen und vervollständigen auf Wunsch des Datengebers die alten Metadaten. Dieser Fall tritt nur ein, wenn zw. Erstlieferung und weiteren Datenlieferungen kein Datenupdate von anderen Institutionen stattgefunden hat.
- 4. Fall **Match + Update 2**: ein neuer Werksdatensatz matcht mit einem bereits vorhandenen Werk. Ergänzungen und Anreicherungen werden auf Wunsch des zweiten Datengebers aus der Institution in das PID-System übertragen. Keine Metadaten werden gelöscht oder korrigiert.

Das Zusammenführen von zwei Datensätzen, die unter den oben definierten Bedingungen gematcht wurden, wird als **Merge** bezeichnet. Einer der beiden Datensätze wird dann übernommen und gegebenenfalls mit Informationen aus dem anderen Datensatz angereichert. Eine Anreicherung kann zum Beispiel stattfinden, wenn bei einem Match in einem Datensatz mehr Angaben für ein Feld gemacht wurden, als in dem gematchten Datensatz. Unterschiedliche Angaben, die einmal mit einer Normdaten-ID (z.B. GND) und einmal ohne (bei Schlagworten, Personen und Orten) vorliegen, werden zusammengeführt. Generell wird jeder Merge protokolliert, sodass er auch im Nachhinein nachvollziehbar ist.

Wenn ein Werk gemergt wurde, werden Schlagworte zusammengeführt, auch wenn eines mit und eines ohne Identifikator vergeben wurde: Werk X hat das Schlagwort „Natur“ ohne GND-ID; Werk Y hat das Schlagwort „Natur“ mit GND-ID. Es handelt sich bei X und Y um dasselbe Werk, die Datensätze werden zusammengeführt und das Schlagwort „Natur“ ist nur einmal im Werksdatensatz verzeichnet, da die Zeichenkette identisch war.

Beim Abgleich werden Amateurfilme (Gattung = Amateurfilm im Schema) anders als bei anderen Gattungen nicht gemergt, wenn das Jahr nur um 1 abweicht. Dies wurde so entschieden, da es im Amateurfilmbereich oft

Werke gibt, die sich nur durch das Produktionsjahr unterscheiden lassen (Beispiel: Titel: Weihnachten, Produktionsort: Deutschland, Jahr: 1977, Regie: Heinz Müller versus Titel: Weihnachten, Produktionsort: Deutschland, Jahr: 1978, Regie: Heinz Müller).

Für den Fall, dass ein Nutzer ein fälschlicherweise gemergtes Werk erkennt, kann er diesen Sachverhalt dem datenhaltenden Archiv mitteilen. Die datengegebende Institution hat die Möglichkeit, sich im AVefi-Redaktionssystem anzumelden und ein gemergtes Werk in zwei neue Werke zu trennen. Der dabei entstehende neue *efi* wird manuell ins Lokalsystem eingetragen. Bei diesem Vorgang entsteht ein Tombstone-Datensatz der das fälschlicherweise gemergte Werk repräsentiert. Dieser behält den initial vergebenen *efi*, da diese evtl. bereits zitiert wurde. Die zwei Werke bekommen zwei neue Datensätze mit zwei neuen *efi*.

6. Bestandsauswahl

6.1 Allgemeines

Die Bestandsauswahl soll breitgefächerte Abgleichroutinen und Ähnlichkeitsanalysen ermöglichen. Sie sieht keine Ein- bzw. Beschränkung vor. Das Ziel der aufeinander abgestimmten Bestandsauswahl im Rahmen des Projektes ist, dass sich Parameter für die Abgleichs- und Ähnlichkeitsanalysen definieren lassen. Daraus resultieren Schnittmengen verschiedener Institutionen, die der Analyse als Basis zur Verfügung stehen. Für die technischen Abgleichanalysen können so die definierten Fälle der Metadatenfachgruppe getestet und auf die technische Praxistauglichkeit geprüft werden.

6.2 Kriterien

Das Konsortium strebt eine Anzahl von circa 1.000 Werken pro Partner an. Verbund- und Praxispartner stellen diese Werksdatensätze für die Entwicklung des Verbundsystems zur Verfügung. Es findet keine Beschränkung hinsichtlich der Gattung statt, das heißt, es können Spiel-, Dokumentar-, Experimental-, Werbe-, Amateurfilme wie auch jede andere Gattung gewählt werden. Die Besonderheit des Projektes AVefi ist es, Bestände jeder Filmgattung im Verbundsystem zu registrieren und darzustellen.

Die breite und offene Auswahl an Beständen führt in letzter Konsequenz zu mehr Anknüpfungspunkten und einem hohen Grad an Repräsentation in der Forschungscommunity.

7. Datenübergabe

7.1 Allgemeines

Die Kriterien zur Bestandsauswahl dienen als Schwerpunktsetzung, um die verschiedenen Abgleichszenarien sowie das Clustern von Amateurfilmen technisch zu erproben. In Zukunft werden die Archive selbst über die Auswahl der Daten entscheiden.

In der ersten Projektphase werden als Schlagworte nur deutsche Begriffe akzeptiert.

Datengeber nennen eine Ansprechperson, der die erforderliche Mapping-Arbeit auf das AVefi-Schema in Zusammenarbeit mit der Metadatenfachgruppe umsetzen kann.

7.2 Lizenzierung

Es können nur Daten eingebracht werden, deren Erfassung, Weitergabe und gemeinfreie Veröffentlichung rechtlich zulässig sind. Die eingebrachten Metadaten stehen allen Interessierten in verschiedenen Formaten und über verschiedene Bezugswege zur Verfügung. Sie sind sowohl als einzelne Datensätze als auch als gesamtes Datenset kostenfrei unter „Creative Commons Zero“-Bedingungen (CCO 1.0) zur freien Nutzung abrufbar. Das Verbundsystem AVefi sichert sich hierfür über die Steuerungsgruppe vertraglich bei allen Datengebern ab, sollte eine dezidierte CCO-Lizenzierung nicht bereits öffentlich dokumentiert sein.

7.3 Datenlieferung per Export

Datenlieferungen mittels eines manuellen Exports sind nach Absprache möglich und bieten einen niedrigschwelligen Zugang zu den Vorteilen des Projekts. Der Nachteil dieser Vorgehensweise ist das nur manuell durchführbare Rückspielen des aktuellen *efi* zurück an die Institutionsdatenbank. Inwieweit diese für AVefi arbeitsintensive Variante in einen Regelbetrieb übernommen werden kann, wird derzeit geprüft.

Bei der Entwicklung von Tools zur Erzeugung von AVefi konformen JSON-Datensätzen empfiehlt sich ein Blick auf unser *efi-conv* Repository: <https://github.com/AV-EFI/efi-conv/>. Dieses stellt ein Check-Modul zur Verfügung, mit dessen Hilfe die erzeugten Dateien validiert werden können (siehe README). Außerdem sind Submodule für verschiedene Mappings enthalten, die als Ausgangspunkt für eigene Entwicklungen genutzt werden können, sofern das sinnvoll erscheint.

Wichtig: Um die eindeutige Beziehung zwischen den Ebenen Werk, Manifestation und Exemplar zu gewährleisten, müssen dem Export für jeden Datensatz entweder lokale oder für den Export erzeugte Identifier beiliegen. Details dazu hier: https://av-efi.github.io/av-efi-schema/has_identifier/.

Voraussetzung für die Teilnahme am Projekt über Exporte ist es, die erzeugten *efi* zurück ins Lokalsystem zu übertragen.

7.4 Datenlieferung über eine Schnittstelle

Die Vorteile einer Datenlieferung über eine API sind: Effizienzsteigerung durch hohes Automatisierungslevel, die *efi* können automatisch übernommen werden inklusive aller Modifikationen und bei Abgabe der Daten erscheint immer der Verweis auf die aktuelle Version der Metadaten. Zu betonen ist die Möglichkeit des Up- und Downloads per Schnittstelle, die eine Aktualisierung der Metadaten in der Quelldatenbank ermöglicht.

7.5 Formate

In der aktuellen Phase werden folgende Formate für Datenlieferungen akzeptiert:

1. JSON
2. XML
3. RDF
4. CSV

Eine endgültige Festlegung der Formate steht noch aus, da derzeit Erfahrungen mit verschiedenen Formaten gesammelt werden und basierend auf dem damit verbundenen Aufwand entschieden wird, wie flexibel sich die Datenlieferung künftig gestalten lässt. Als Faustregel gilt, dass JSON, idealerweise nach einem Schema, gut zu verarbeiten ist. XML-Dateien sind ebenfalls vielversprechend, vorausgesetzt, es existiert ein Schema und die Dateien werden nachweislich gegen dieses validiert. Auch RDF wird als potenziell vorteilhaft angesehen. CSV-Dateien, also auch Excel-Exporte, können ebenfalls verarbeitet werden. Allerdings besteht bei allen Formaten die Gefahr, dass unstrukturierte Daten innerhalb eines Feldes abgelegt werden, wobei das Risiko bei CSV höher erscheint als beispielsweise bei RDF. Allgemein gilt: Je strukturierter die Daten, desto besser.

7.6 Arbeitsschritte zur Datenlieferung

1. Mapping von Feldinhalten aus Lokaldatenbank (LDB) auf das [AVefi-Schema](#)
2. Export der Metadaten aus LDB
3. Überarbeitung der Metadaten, um Schemakonformität zu erreichen, vor oder nach Export
4. Übergabe des Exports mit lokalem oder für den Export erzeugtem Identifier an AVefi-Team
5. Einspielung der Daten in AVefi-System
6. Export der *efi* aus AVefi und Import der *efi* in LDB
7. Bei Nutzung einer Schnittstelle erfolgen die Schritte 2-6 weitestgehend automatisch

7.7 How-to und Mapping-Vorlagen

Die Anforderungen für Datengeber an das Mapping und die Datenübergabe werden durch die gesammelten Erfahrungen während der Projektlaufzeit kontinuierlich weiterentwickelt.

Im „How-to“-Dokument, welches als konkretisierende Anlage zu diesem Manual dient, werden verschiedene Wege der Datenübergabe sowie die wichtigsten Voraussetzungen an die Datenübergabe beschrieben.¹¹

Ein Beispieldatensatz für AVefi konformes JSON findet sich unter:

https://github.com/AV-EFI/av-efi-schema/blob/main/examples/json/menschen_am_sonntag_all_in_one.json

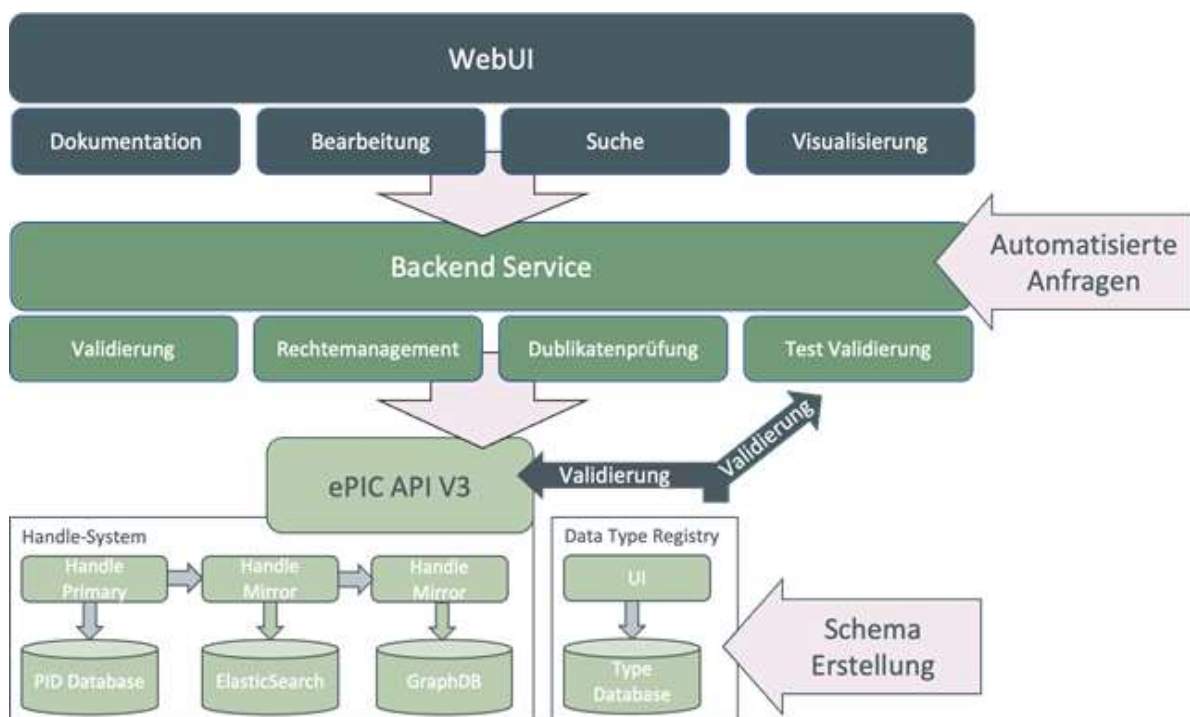
Weitere Details zu Mapping-Vorlagen folgen in einer späteren Version dieses Dokuments.

Die inhaltliche Hilfestellung beim Mapping erfolgt durch die Metadatenfachgruppe. Bei technischen Fragen erfolgt eine Weiterleitung an die technische Gruppe.

8. Technische Infrastruktur

8.1 Allgemeines

Das PID-System bildet die Grundlage der technischen Infrastruktur und umfasst sämtliche Informationen zu den registrierten Werken, Manifestationen und Exemplaren. Zusätzlich wird eine separate Datenbank bereitgestellt, die für die Benutzeroberfläche (UI) erforderlich ist. Ein Back-End-Server stellt sämtliche logischen Funktionen zur Verfügung, um mit den gesicherten Daten zu arbeiten. Die Algorithmen zur Ähnlichkeitsanalyse werden im Backend implementiert. Bei der Identifikation kritischer Fälle ist vorgesehen, diese über die WebUI im Rahmen eines redaktionellen Konfliktmanagements zu bearbeiten.



¹¹ Die erste Vorabversion des How-to-Dokuments befindet sich im Anhang.

8.2 Das PID-System

Das PID-System basiert auf dem Handle-System¹². Das [ePIC Konsortium](#) stellt durch geeignete Policies und technische Schnittstellen bzw. Application Programming Interfaces (API) die Persistenz der Identifikatoren sicher. Dazu zählt auch eine Spiegelung der Daten zwischen den am ePIC Konsortium beteiligten Partnerorganisationen, zu denen die GWDG gehört. Die Validierung der in das PID-System übermittelten Daten wird von einer erweiterten API realisiert. Das Handle-System ist seit vielen Jahren etabliert und bildet die technische Basis von einer enormen Anzahl an PID die in verschiedenen Kontexten und Communities in den letzten Jahren erstellt wurden und in Zukunft erstellt werden. Durch die Verwendung des Handle-Systems besteht die Möglichkeit, Information im Handle-Record als Attribute abzulegen. Dadurch sind alle notwendigen Informationen direkt beim Auflösungsprozess der PID erreichbar, ohne dass weitere Schnittstellen oder Dienste abgefragt werden müssen.

8.3 Die Type-Registry

In der Type-Registry werden Datentypen definiert, die für strukturierte Daten (im JSON-Format) bestimmte formale, validierbare Anforderungen festlegen. So kann beispielsweise für Datumsfelder ein bestimmtes Format der Darstellung oder für ein Feld wie Titeltyp (titleType) eine feste Werteliste vorgegeben werden. Datentypen für komplexere JSON-Objekte können andere Datentypen referenzieren, um Teilstrukturen festzulegen. Mit diesen Datentypen können die APIs des PID-Systems eingehende Daten validieren und somit sicherstellen, dass PID immer einem vorgegebenen Schema entsprechen. Ein JSON-Schema der akzeptierten PID kann ebenfalls per API abgerufen und lokal zur Validierung eingesetzt werden. Die Type-Registry sowie der Dienst zur Erstellung der Validierungsschemata werden von der GWDG im Rahmen eines internationalen Projekts betrieben und weiterentwickelt.

8.4 Änderungen des technischen Schemas

Technisch werden, wie im vorhergehenden Abschnitt beschrieben, Datentypen in der Type-Registry definiert, sodass das resultierende JSON-Schema den Anforderungen entspricht, die in der Metadatengruppe erarbeitet wurden. Die Entwicklung des Schemas wird unter <https://github.com/AV-EFI/av-efi-schema> öffentlich dokumentiert.

Anforderungen und Änderungsvorschläge werden in der Metadatenfachgruppe diskutiert und für ein Release zusammengestellt. Sobald ein release-fähiger Stand erreicht ist, wird dieser durch die Steuerungsgruppe genehmigt und von der Technikgruppe im nachfolgend beschriebenen Prozess umgesetzt.

1. Alle Änderungsanträge werden als Issue auf GitHub angelegt (<https://github.com/AV-EFI/av-efi-schema>).
2. Für das Issue wird eine Deadline nach Vorgabe aus dem Steuerungsgremium eingetragen.
3. Für das Issue wird eine verantwortliche Person aus der Technikgruppe eingetragen.
4. Änderungen werden in der Modellierungssprache LinkML ins Repositorium eingepflegt und durch eine im README dokumentierte Pipeline in verschiedene Derivate (JSON-Schema, Python-Bindings, etc.) übernommen.
5. Durch ein weiteres Kommando (ebenfalls im README dokumentiert) werden die Änderungen in die Data-Type-Registry übernommen. Sofern dabei neue Datentypen angelegt werden, erfolgt eine Erfassung der zugehörigen PID im Ausgangsrepositorium als Teil der LinkML Definition der jeweiligen Elemente.
6. Die automatisch erzeugte Schema-Dokumentation wird publiziert (https://av-efi.github.io/av-efi-schema/schema_overview).
7. Die Versionsnummer im Git-Repositorium wird aktualisiert.

¹² <https://www.handle.net>

8.5 Schnittstellen und Datenaustausch

Die Metadaten der registrierten *efi* im AVefi-System werden per REST-Schnittstelle im JSON-Format öffentlich zugänglich gemacht. Die Ausgabe soll auch in anderen domain-spezifischen Formaten wie EBUCore/XML oder FIAFcore/RDF ermöglicht werden.

Auch für den Metadatenimport, d.h. das Registrieren neuer *efi*, wird eine REST-Schnittstelle entwickelt und eine möglichst weitgehende Automatisierung angestrebt. Insbesondere wird versucht, automatisch zu erkennen, ob zu einem Satz von Werksmetadaten bereits ein *efi* existiert, um kein Duplikat anzulegen. Für dieses Matching von Werksmetadaten werden Kriterien in der Metadatenfachgruppe definiert, also beispielsweise welche Felder wie gut übereinstimmen müssen, um von ein und demselben Werk ausgehen zu können. Wurde ein Match gefunden, so erfolgt gegebenenfalls eine Ergänzung der Metadaten nach bestimmten Kriterien, die ebenfalls in der Metadatenfachgruppe festgelegt werden.

Bei diesen Abläufen ist mit Fehlern zu rechnen, weshalb eine redaktionelle Nachbearbeitung per WebUI ermöglicht werden soll. Außerdem sind eine Änderungshistorie und der Abruf früherer Bearbeitungszustände vorgesehen.

9. Kontakt / Projektteam

contact@av-efi.net

Projektmanagement:

TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek

Sabrina Silbernagel

Matti Stöhr

Datenmanagement und Schnittstellenentwicklung:

Stiftung Deutsche Kinemathek (SDK)

Jonas Pitz

Antje Dittmann

Annette Groschke

Filmmuseum Düsseldorf (FMD)

Thomas Ochs

Stefan Matijasevic

Backend-Entwicklung:

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)

Elias Oltmanns

Sven Bingert

UX-Design & Frontend Development:

TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek

Stefan Stretz

10. Weitere Informationen

[Anmeldung Mailingliste für Praxispartner und Interessierte \(Community-Kommunikation\)](#)

[Eintrag in GEPRIS-Datenbank mit Projektzusammenfassung](#)

[Projektwebsite](#)

[Metadatenschema](#)

[Zotero-Literatursammlung zum Projekt](#)

[Report zum Vorprojekt „Archivierungsstrategie für audiovisuelles Material“](#)