

Sonja Fiala

Schritt-für-Schritt- Anleitung: Metadatenmapping

Handbuch Repositorienmanagement, Hg. v. Blumesberger et al., 2024, S. 185–195
<https://doi.org/10.25364/978390337423211>



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz, ausgenommen von dieser Lizenz sind Abbildungen, Screenshots und Logos.

Sonja Fiala, Universität Wien, Fachbereichsbibliothek Philosophie und Psychologie, sonja.fiala@univie.ac.at |
ORCID iD: 0000-0002-5492-8934

Zusammenfassung

Metadaten sind die Basis der Informationsgesellschaft. Nur wenn sie vollständig und korrekt angegeben werden, ist ein präzises Wiederauffinden der Objekte, die sie beschreiben, möglich. In dem folgenden Beitrag werden die verschiedenen Vorgehensweisen beim Metadatenmapping erörtert. Es wird auf die wichtigsten Punkte beim Mappingvorgang hingewiesen.

Schlagwörter: Metadatenmapping; Phaidra; MODS; RDF

Abstract

A Step-By-Step Tutorial: Metadata Mapping

Metadata form the fundament of the information society. The objects they describe can only be retrieved if the data are complete and precise. In the following paper different possible procedures in metadata mapping will be discussed, highlighting the most important points in the mapping process.

Keywords: Metadata mapping; phaidra; MODS; RDF

1. Einleitung

Die folgende Anleitung schildert die Erfahrungen, die während eines gemeinsamen Projekts der Universität Wien mit der Universität Padua gewonnen wurden. Seit einiger Zeit werden im „Universitätslehrgang Library and Information Studies“ das Wahlfach „Data Librarian“ und der Teilbereich Metadatenmapping unterrichtet. Als Beispiel der Ausführungen dient das Repositorium der Universität Wien (Phaidra¹). Die Datenstruktur von Phaidra wurde über die letzten zehn Jahre durch verschiedene Fremdelemente, wie z. B. um die Angabe der ORCID iD, konstant erweitert und adaptiert. Um für die Zukunft gerüstet zu sein, kann es notwendig sein, bestehende Datensätze in international übliche Datenformate zu konvertieren und neue Datensätze nach internationalen Standards einzupflegen. Besonders wichtig wird dies, um verlässliche Suchergebnisse zu erhalten.

Metadaten spielen seit jeher im Bibliothekswesen eine wichtige Rolle. Früher waren es die Zettelkataloge, heute sind es die Daten über Daten, die im Zentrum der Arbeit stehen. Es ist von größter Bedeutung, dass die Erschließung der Daten nach Richtlinien erfolgt, um die Qualität der Suchergebnisse garantieren zu können. So wird es notwendig, zwischen den unterschiedlichen Beschreib- und Erschließungssystemen Parallelen herzustellen und die einzelnen Kategorien in Verbindung zu bringen. Die Metadaten müssen operabel sein, um in Zukunft von Nutzen sein zu können.

Mapping is an important operation for data migration from legacy systems, a database model system, to other modern web technology or semantic system that computers can understand meaning of data. These are in many application domains, such as semantic web, schema or ontology integration, data integration etc. The methodology of mapping can distinguish three ways are: (1) database schema mapping, (2) Ontology mapping, and (3) database to RDF mapping [...] ² [sic!]

Metadaten werden zunehmend zum entscheidenden Faktor in der Informationsgesellschaft. Nur wenn die Metadaten eines Objektes exakt erfasst sind, werden Datensätze auffindbar und präzise recherchierbar. Man kann somit Metadaten als die Basis der Informationsgesellschaft bezeichnen³. Neben den Kernelementen (z. B. Metadata Object Description Schema (MODS): titleInfo, name, typeOfResource,

1 <https://phaidra.univie.ac.at/>

2 Thangsupachai, N. et al. (2014), S. 123.

3 Als allgemeine Einführung in das Thema Metadaten ist folgendes Buch sehr empfehlenswert: Pomerantz, J. (2015).

genre, originInfo, language, physicalDescription, abstract, identifier, ...)⁴ enthalten Metadaten viele weitere Informationen, die auch immer im Auge behalten werden müssen. Alle Ebenen müssen korrekt verwaltet und gewartet werden – so beispielsweise die deskriptiven Metadaten, strukturellen Metadaten, Erhaltungsmetadaten, Herkunftsmetadaten, Nutzungsmetadaten und administrativen Metadaten.

Metadaten entscheiden über die Wiederauffindbarkeit, die präzise Suchbarkeit und letztendlich über die Möglichkeit des korrekten wissenschaftlichen Arbeitens. Hierbei spielen verschiedene Komponenten zusammen: Es müssen die Metadaten in den richtigen Feldern eingetragen sein, sie müssen vollständig sein und sie müssen die Daten ausreichend beschreiben.

2. Metadatenmapping

Eine adäquate Herangehensweise beim Metadatenmapping ist, dass man sich zunächst die Datenstruktur ansieht: Es ist wichtig zu verstehen, wie ein Datensatz in sich aufgebaut ist. Oftmals sind die einzelnen Kategorien in den unterschiedlichen Beschreibssystemen verschieden benannt. Jeder Datensatz folgt einer Logik, die nach bestimmten Regeln aufgebaut ist. Die Bedeutung der einzelnen Kategorien erschließt sich im Gebrauch. Das strukturierte Verstehen der einzelnen Kategorien zeichnet den Mappingvorgang aus. Die Problematik, dass die Eingabe der Daten z. B. in einem Repositorium wie Phaidra von verschiedenen Personen und mit unterschiedlicher Erschließungsgenauigkeit erfolgt, führt zu einem Herantasten an die Verwendung der Kategorien. Hier ist in Zukunft auf genaue Anleitungen für Personen, die Daten hochladen, zu achten. Das Metadatenmapping erinnert an das Zitat von Ludwig Wittgenstein, der zum Gebrauch der Wörter Folgendes sagt: „So lerne ich nach und nach verstehen, welche Dinge die Wörter bezeichnen, die ich wieder und wieder, an ihren bestimmten Stellen in verschiedenen Sätzen, aussprechen hörte.“⁵ Und weiter „Was sie bezeichnen, wie soll sich das zeigen, es sei denn in der Art ihres Gebrauchs?“⁶ Es werden auch Angaben sichtbar, die nicht in ein neues Schema überführbar sind. Für sie muss dann nach einer anderen Lösung gesucht werden (z. B. Notizfeld etc.).

4 MODS Elements and Attributes: https://www.loc.gov/standards/mods/userguide/general-app.html#top_level

5 Wittgenstein, L. (2019), S. 237.

6 Ebd. S. 242.

3. Mappingvorgang

Das Mapping gleicht einer Drehbewegung, die an verschiedenen Stellen Halt sucht. Man beginnt mit jeweils einem Beispiel der unterschiedlichen Beschreibungsarten. Diese stellt man einander gegenüber und versucht Verbindungen herzustellen. Dann überprüft man die gewonnenen Erkenntnisse auch in den darauffolgenden Beispieldatensätzen. Hier kommt noch ein ganz wichtiger Aspekt dazu, nämlich die Frage, wie tief man die Datensätze in Zukunft erschließen möchte. Das heißt, an diesem Punkt muss man sich die Frage stellen, wie die Verbesserung in Zukunft aussehen soll. Man sieht sich das zugrundeliegende Schema genau an und entscheidet, welche Kategorien wichtig sind. Hier kann auch das probeweise Erfassen des Objekts in dem neuen Schema ein hilfreicher Vorgang sein. So werden die unterschiedlichen Kategorien schnell sichtbar und der Fokus wird auf die zu klärenden Unterschiede gelegt.

4. Dokumentation

Der Dokumentation der Kategorienverwendung kommt eine große Bedeutung zu. Alle Überlegungen sollten dokumentiert und kommentiert werden. Je mehr Personen den Mappingvorgang begleiten und mitdenken, desto verlässlicher wird das Ergebnis. Auch die Kommunikation zwischen den Datenbankbetreiber:innen, den Nutzer:innen und den Wissenschaftler:innen bekommt hier eine besondere Bedeutung. Welche Suchanfragen werden von den Nutzer:innen benötigt und verwendet? Wie aufwendig soll die Dateneingabe gestaltet sein? Gibt es Anleitungen und Schulungen?

5. Metadatenmanager:innen

Die Aufgaben des/der Metadatenmanager:in⁷ umfassen nicht nur die korrekte Beschreibung der Metadatenätze, sondern auch die Kommunikation. Diese rückt ins Zentrum der Aufgabe, je weiter die Arbeit voranschreitet. Ebenso bedeutsam ist die Fähigkeit, auf Basis großer Erfahrung strukturiert zu arbeiten. Eine Kategorie nach der anderen muss bewertet und analysiert werden. Es sind viele unterschiedliche Darstellungsweisen möglich und potenziell sinnvoll, aber letztlich entscheidet der Erfolg darüber, welche Darstellungsweise man wählt.

Das Metadatenmapping wird somit zu einer spannenden und herausfordernden Arbeit, die wohl zu den gefragtesten bibliothekarischen Tätigkeiten der Zukunft werden wird. Viele Datenbanken entwickeln sich über viele Jahre und müssen ständig

7 Blumesberger, S.; Traub, I. D.; Schubert, B. et al. (2016)

gepflegt und gewartet werden. Neue Dateitypen entstehen und das kontrollierte Vokabular muss laufend gepflegt werden. Ein aktuelles Thema ist zum Beispiel die Beschreibung von barrierefreien Dateien⁸. Viele Normen müssen berücksichtigt und in der Dokumentation festgehalten werden. Diese Dokumentation ist wichtig für die nachkommende Generation, denn nur dann ist der Datenbankaufbau nachvollziehbar. Dieser Punkt wurde gerade in den Anfängen der Digitalisierung nicht ausreichend berücksichtigt. So sind viele Eingabemöglichkeiten historisch gewachsen und bei einem Mapping schwer zuordenbar. Die Kategorien von Dublin Core haben oft in der Praxis nicht ausgereicht und so wurden viele Angaben im Laufe der Zeit ergänzt.

As mentioned above, Dublin Core was developed to be a lowest common denominator metadata element set. The problem with the lowest common denominator, however, is that sometimes it's too low.⁹

Auch die neue Architektur der Linked Data macht eine Überarbeitung der Datenstruktur notwendig. Am Beispiel Phaidra lassen sich die unterschiedlichen Entwicklungsstadien gut ablesen. Es gibt UWMETADATA-Datensätze (Universität-Wien-Metadatenätze)¹⁰, Dublin-Core¹¹, JSON-LD¹² („JavaScript Object Notation for Linked Data“) und MODS-Datensätze¹³. MODS wird von der Library of Congress zur Verfügung gestellt.¹⁴ Von MODS soll dann der Weg zu Resource Description Framework (RDF) geebnet sein.¹⁵

Besonders erwähnenswert sind auf der Seite der Library of Congress die verschiedenen MODS-Beispiele für unterschiedliche Medienarten¹⁶ und die verschiedenen Versionen¹⁷, da MODS laufend aktualisiert wird: So kommen neue Eingaben hinzu, wie zum Beispiel ORCID iD¹⁸:

8 Siehe auch Fiala, S. (2019).

9 Pomerantz, J. (2015), S. 81.

10 Z. B. <https://fedora.phaidra.univie.ac.at/fedora/objects/o:740497/methods/bdef:Asset/getUWME-TADATA>

11 Z. B. <https://services.phaidra.univie.ac.at/api/object/o:740497/index/dc>

12 Z. B. <https://services.phaidra.univie.ac.at/api/object/o:1192264/jsonld>

13 Z. B. <https://services.phaidra.univie.ac.at/api/object/o:950749/mods?format=xml>

14 <http://www.loc.gov/standards/mods/>

15 <https://www.loc.gov/standards/mods/modsrdf/>

16 <https://www.loc.gov/standards/mods/userguide/examples.html>

17 <https://www.loc.gov/standards/mods/mods-schemas.html>

18 <https://orcid.org/>

Beispiel ORCID iD:

```
<name>
  <namePart type="given">Wilhelmina</namePart>
  <namePart type="family">Randtke</namePart>
  <nameIdentifier type="orcid">https://orcid.org/0000-0002-7439-
8205</nameIdentifier>
</name>
```

Wenn man die UWMETADATA nun auf MODS-Niveau bringen möchte, ist ein Metadatenmapping notwendig. In Zusammenarbeit mit der Universität Padua ist ein solches bereits vorbereitet worden¹⁹. Besonders dann, wenn es sich um eine gemeinsame Suchoberfläche (z. B. u:search der Universitätsbibliothek Wien) handelt, führt eine korrekte Datenstruktur zu qualitativ hochwertigeren Ergebnissen.

Die Bedeutung einer ausreichenden Dokumentation kann an dieser Stelle nicht genug betont werden. Die Festlegung der Reihenfolge der Kategorien, ihre Wiederholbarkeit, die zugrundeliegenden Standards und Normen müssen dokumentiert und erklärt werden.

6. Dokumentation der Verwendung

Es gibt bereits einige Beispiele für praxisnahe MODS-Dokumentationen. Hier seien z. B. die Dartmouth College Library MODS Documentation, die Princeton University MODS Documentation und die Swepub MODS format specification erwähnt.²⁰

Aufbauend auf der Kenntnis beider Metadatenschemata und der Gegenüberstellung verschiedener Beispiele rückt nun in einem nächsten Schritt die Datensatzkohärenz in den Mittelpunkt.

¹⁹ Fiala, S.; Huggle, C. (2019)

²⁰ Dartmouth College Library MODS Documentation: https://www.dartmouth.edu/library/catmet/metadata_nonmarc/mods_docs/; Princeton University Library MODS Documentation: <https://library.princeton.edu/departments/tsd/metadoc/mods/index.html>; National Library of Sweden: https://www.kb.se/namespace/swepub/mods-format-specification/SWP_MODS_3.pdf

<pre></mods:subject> <subject> <topic authority="gnd" authorityURI="http://d-nb.info/gnd/" valueURI="http://d-nb.info/gnd/4039902-3">Mönch</topic> <topic authority="gnd" authorityURI="http://d-nb.info/gnd/" valueURI="http://d-nb.info/gnd/4008690-2">Buddhismus</topic> <genre authority="gnd" authorityURI="http://d-nb.info/gnd/" valueURI="http://d-nb.info/gnd/4006568-6">Bild</genre> </subject></pre>	
<pre><classification authority="ddc" edition="13">930: Geschichte des Altertums bis ca. 499, Archäologie</classification> <classification authority="ÖFOS 2012">601010: Klassische Archäologie</classification> <classification authority="ÖFOS 2012">601026: Virtuelle Archäologie</classification></pre>	<pre><ns1:classification> <ns7:taxonpath> <ns7:source>13</ns7:source> <ns7:taxon seq="0">1072220</ns7:taxon> <ns7:taxon seq="1">1072233</ns7:taxon> <ns7:taxon seq="2">1068322</ns7:taxon> </ns7:taxonpath> <ns7:taxonpath> <ns7:source>16</ns7:source> <ns7:taxon seq="0">1072240</ns7:taxon> <ns7:taxon seq="1">1072397</ns7:taxon> <ns7:taxon seq="2">1073639</ns7:taxon> <ns7:taxon seq="3">1072400</ns7:taxon> </ns7:taxonpath> <ns7:taxonpath> <ns7:source>16</ns7:source> <ns7:taxon seq="0">1072240</ns7:taxon> <ns7:taxon seq="1">1072397</ns7:taxon> <ns7:taxon seq="2">1073639</ns7:taxon></pre>

Beispiel 1: Prüfung der gebräuchlichen Elemente (Metadatenmapping, Universitätsbibliothek Wien, Frühjahr 2018)

Eine Analyse der fehlenden Kategorien ist unumgänglich, um korrekte und vollständige Datensätze zu erhalten.

Man kann die bisherigen Schritte wie folgt zusammenfassen und im letzten Schritt gemeinsam anwenden:

- Gegenüberstellung von Beispieldatensätzen
- Kenntnis und Erlernen beider Metadatenschemata
- Analyse der fehlenden Kategorien
- Prüfung der Metadatenkohärenz
- Probemapping mit Anmerkungen und weiteren Bearbeitungsschritten (z. B. Festlegung des kontrollierten Vokabulars etc.)

Das Mapping erfolgt von den Kernelementen hin zu den Randelementen.

Beispiel 1 (Artikel, Text): o:685260

<pre> <ns0:uwmetadata xmlns:ns0="http://phaidra.univie.ac.at/XML /metadata/V1.0" xmlns:ns1="http://phaidra.univie.ac.at/XML /metadata/lom/V1.0" xmlns:ns10="http://phaidra.univie.ac.a t/XML/metadata/provenience/V1.0" xmlns:ns11="http://phaidra .univie.ac.at/XML/metadata/provenience/V1.0/entity" xmlns: ns12="http://phaidra.univie.ac.at/XML/metadata/digitalbook /V1.0" xmlns:ns13="http://phaidra.univie.ac.at/XML/metadat a/ethteses/V1.0" xmlns:ns2="http://phaidra.univie.ac.at/XML/ metadata/extended/V1.0" xmlns:ns3="http://phaidra.univie.a c.at/XML/metadata/lom/V1.0/entity" xmlns:ns4="http://phaid ra.univie.ac.at/XML/metadata/lom/V1.0/requirement" xmlns:ns 5="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/educa tional" xmlns:ns6="http://phaidra.univie.ac.at/XML/metadat a/lom/V1.0/annotation" xmlns:ns7="http://phaidra.univie.ac .at/XML/metadata/lom/V1.0/classification" xmlns:ns8="http:// phaidra.univie.ac.at/XML/metadata/lom/V1.0/organization" xmlns:ns9="http://phaidra.univie.ac.at/XML/metadata/histku lt/V1.0"> <ns1:general> <ns1:identifier>o:685260</ns1:identifier> <ns1:title language="de"> Digitale Archäologie und <u>Molino</u> San Vincenzo: Kein "Digital Dark Age" in der Toskana </ns1:title> <ns1:language>de</ns1:language> <ns1:description language="de"> Um den verschiedenen Prinzipien offener Wissenschaftskommunikation – wie auch dem ganzen Konzept «open science» als solchem – gerecht zu werden, ist eine </pre>	<pre> <?xml version="1.0" encoding="UTF-8"?> <mods xmlns="http://loc.gov/mods/v3" xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-6.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" version="3.6"> <titleInfo lang="de"> <title>Digitale Archäologie und <u>Molino</u> San Vincenzo: Kein "Digital Dark Age" in der Toskana</title> </titleInfo> <language> <languageTerm type="code" authority="iso639-2b">ger</languageTerm> </language> <abstract> Um den verschiedenen Prinzipien offener Wissenschaftskommunikation – wie auch dem ganzen Konzept «open science» als solchem – gerecht zu werden, ist eine nachhaltige Disseminations- und </pre>
---	---

Beispiel 2: Mapping der Datensätze (Metadatenmapping, Universitätsbibliothek Wien, Frühjahr 2018)

Hierfür kann eine Statistik hilfreich sein. Diese Statistik erfasst die Verwendungshäufigkeit einer Kategorie. Am Beispiel Phaidra kann diese Statistik dann für die Bewertung unterschiedlicher Kategorien herangezogen werden.

7. Gegenüberstellung

Für die weitere Vorgehensweise sind nun die Top-Level Elements²¹ in MODS wichtig. Am besten eignen sich zwei Spalten in einem Textverarbeitungs- oder Tabellenkalkulationsprogramm mit Kommentaren für die Analyse der Datensätze.

In den Kommentaren werden die Sachverhalte, die noch geklärt werden müssen, erläutert. Um eine bessere Vorstellung des Mappingvorgangs zu bekommen, empfiehlt es sich, ein Probemapping durchzuführen. Alle Probleme und Fragestellungen werden in Kommentaren festgehalten. Ins Zentrum der Aufmerksamkeit rücken die Punkte, die noch weiter diskutiert und besprochen werden müssen.

Seit dem Sommersemester 2020 wird der Fachbereich Metadatenmapping im Universitätslehrgang Library and Information Studies unterrichtet. Die Studierenden

²¹ https://www.loc.gov/standards/mods/userguide/generalapp.html#top_level

erhalten die Aufgabe, ein Mapping (UWMETADATA >> MODS) durchzuführen, gemeinsam zu diskutieren und alle Überlegungen ausführlich zu dokumentieren und zu kommentieren.

Die Schritt-für-Schritt-Anleitung dient als Ausgangspunkt für die Studierenden, die dann gemeinsam ein breites Spektrum an Herangehensweisen erarbeiten. Diese praxisnahen Erkenntnisse sollen die Studierenden auf ihre zukünftigen Aufgaben als Metadatenpezialist:innen vorbereiten.

Bibliografie

- Blumesberger, Susanne; Traub, Imola Dora; Schubert, Bernhard; Hudak, Rastislav; Gründhammer, Veronika (2016): Jobprofil von MetadatenmanagerInnen. <https://phaidra.univie.ac.at/o:441513>
- Fiala, Sonja (2019): Kennzeichnung barrierefreier Dateien – eine Zusammenstellung am Beispiel MARC21 und MODS. <https://resolver.obvsg.at/urn:nbn:at:at-ubi:3-6007>
- Fiala, Sonja; Huggle, Christina (2019): Metadatenmapping. Die Gegenüberstellung verschiedener Metadatenschemata am Beispiel UWMETADATA >> MODS 3.6. <https://resolver.obvsg.at/urn:nbn:at:at-ubi:3-6020>
- Pomerantz, Jeffrey (2015): Metadata. (The MIT Press essential knowledge series). Cambridge, Mass.: The MIT Press.
- Thangsupachai, Noppol; Niwattanakul, Suphakit; Chamnongsri, Nisachol (2014): Learning Object Metadata Mapping for Linked Open Data. In: Kulthida Tuamsuk, Adam Jatowt, Edie Rasmussen (eds.): The Emergence of Digital Libraries – Research and Practices. 16th International Conference on Asia-Pacific Digital Libraries, ICADL 2014, Chiang Mai, Thailand, November 5-7, 2014 Proceedings. (Lecture Notes in Computer Science 8839). Cham: Springer.
- Tuamsuk, Kulthida; Jatowt, Adam; Rasmussen, Edie (eds.) (2014): The Emergence of Digital Libraries – Research and Practices. 16th International Conference on Asia-Pacific Digital Libraries, ICADL 2014, Chiang Mai, Thailand, November 5-7, 2014, Proceedings. (Lecture Notes in Computer Science 8839). Cham: Springer.
- Wittgenstein, Ludwig (2019): Tractatus logico-philosophicus, Tagebücher 1914-1916, Philosophische Untersuchungen, 23. Auflage. (Ludwig Wittgenstein: Werkausgabe 1) Frankfurt am Main: Suhrkamp.

Sonja Fiala ist Leiterin der Fachbereichsbibliothek Philosophie und Psychologie der Universität Wien. Sie unterrichtet das Wahlfach Informationsethik im Universitätslehrgang Library and Information Studies und den Fachbereich Metadatenmapping im Wahlfach „Data Librarian“ bzw. im gleichnamigen Zertifikatskurs. Sie veröffentlichte zum Thema Informationsethik und Bibliotheken sowie zum Thema

Metadatenmapping. Zuletzt: Sonja Fiala (2019): Kennzeichnung barrierefreier Dateien – eine Zusammenstellung am Beispiel MARC21 und MODS, <https://resolver.obvsg.at/urn:nbn:at:at-ubi:3-6007>.