



e·infra
austria

AP 5 Metadaten

Metadatenmapping Die Gegenüberstellung verschiedener Metadatenschemata am Beispiel UWMETADATA>>MODS 3.6

Sonja Fiala , Christina Huggle 

Universität Wien

September 2019



Der Inhalt dieser Veröffentlichung steht unter einer Creative Commons
Namensnennung 4.0 Lizenz (<http://creativecommons.org/licenses/by/4.0/>)

[DOI 10.25651/1.2019.0013](https://doi.org/10.25651/1.2019.0013)

Über e-Infrastructures Austria Plus

Das Projekt „e-Infrastructures Austria Plus“ (2017-2019) ist ein vom österreichischen Bundesministerium für Bildung, Wissenschaft und Forschung (<https://bildung.bmbwf.gv.at/>) gefördertes Projekt von neun österreichischen Universitäten. Ziel des Projekts ist der koordinierte Aufbau eines österreichischen Netzwerks zur Einrichtung und Weiterentwicklung gemeinsamer e-Infrastrukturen durch Bündelung von Ressourcen und vorhandenem Wissen.

Ziel

Ziel des Arbeitspakets „*Metadaten*“ ist die Erarbeitung eines Metadatenframeworks für Repositorien in Österreich und von Guidelines, die mit FAIR-Richtlinien (FAIR Guiding Principles for scientific data management and stewardship) arbeiten. Die Suche von Forschungsdaten und anschließende Projektkooperationen können repositorienübergreifend nur funktionieren, wenn ein gemeinsamer Satz von Metadaten beim Einbringen der Daten erstellt wird und wenn die eingebrachten Daten mit kontrolliertem Vokabular beschrieben werden. Das Projekt will diese beiden Voraussetzungen schaffen.

Einleitung

Die in Phaidra verwendeten UWMETADATA sind ein Metadatenchema der Universität Wien, das über die letzten zehn Jahre durch verschiedenste Fremdelemente konstant erweitert und adaptiert wurde. Um einen zukünftigen Datenaustausch zu erleichtern, soll diese Form der Metadatenbeschreibung in einem ersten Schritt auf MODS (Metadata Object Description Schema) gemappt werden. MODS ist ein XML-Schema für bibliografische Metadaten, das vom Network Development and MARC Standards Office der Library of Congress entwickelt und verwaltet wird.¹

Dabei soll eine Dokumentation über die Verwendung der Metadaten ausgearbeitet werden. Es geht hierbei um die konkrete Anwendung der Metadaten an der Universität Wien. Ein gutes Beispiel für so eine Dokumentation ist die „Princeton University Library MODS Documentation“.² In einem weiteren Schritt sollen dann die konvertierten MODS Metadaten zu RDF gemappt werden.

Der vorliegende Bericht gibt einen Einblick in die Vorgehensweise, die Schwerpunkte und etwaigen Probleme des Mappings und informiert über dessen Fortschritt.

1. Schritt: „Meaning is use“

UWMETADATA

Zunächst geht es darum, einen Überblick über die historisch gewachsenen Metadatenabschnitte und deren Verwendung zu erhalten. Alle Listen, Codes und Auswahlmöglichkeiten müssen gesichtet sowie deren Verwendung verstanden und hinterfragt werden. Um die Verwendung der verschiedenen Kategorien nachvollziehen zu können steht vor allem die **Darstellung** der UWMETADATA in Phaidra im Zentrum der Aufmerksamkeit. Die Häufigkeit der Verwendung wird anhand einer Statistik überprüft.

MODS 3.6

Auf der Seite der Library of Congress befinden sich die **MODS User Guidelines** mit Beispielen ihrer Anwendung.³ Ein Schlüsselsatz für die weitere Vorgangsweise ist der Punkt **Cataloging Rules**: „Any set of cataloging rules may be used with MODS, as in the case with MARC 21.“

Anhand von Beispielen der Library of Congress und der Universität Heidelberg wird die Verwendung von MODS sichtbar und nachvollziehbar.

¹ https://de.wikipedia.org/wiki/Metadata_Object_Description_Schema (Stand: 04.09.2019)

² <http://library.princeton.edu/departments/tsd/metadoc/mods/index.html> (Stand: 04.09.2019)

³ <http://www.loc.gov/standards/mods/mods-guidance.html> (Stand: 04.09.2019)

Die Darstellung von MODS erfolgt analog zu UWMETADATA im XML-Format, was für ein webbasiertes Repositorium von Vorteil ist: „MODS is expressed using the XML (Extensible Markup Language) schema language of the World Wide Web Consortium.⁴ XML provides markup for documents and is expected to allow more flexibility and detail than HTML (Hypertext Markup Language). It serves well as a syntax for metadata.“

2. Schritt: Das Mapping

In einem weiteren Schritt stehen die Überlegungen der Darstellung und der Arbeitsrichtung des Mappings im Mittelpunkt. Es hat sich gezeigt, dass vom Zielmetadatenchema zum vorhandenen Metadatenchema hingearbeitet werden muss.

Eine Unterscheidung nach Medientypen ist sinnvoll um Musterbeispiele für zukünftige Mappings zu erstellen.

Computer file (Web document) Quelle: <http://www.loc.gov/standards/mods/mods-guidance.html> = Artikel, Text

MODS 3.6	UWMETADATA
TITELDATEN <titleInfo> <title>American ballroom companion</title> <subTitle>dance instruction manuals, ca. 1600-1920</subTitle> <titleInfo type="alternative"> <title>Dance instruction manuals, ca. 1600-1920</title>	<ns1:title language="de">Humanistischer Generalismus</ns1:title> <ns1:title language="en">Oriental Music And Migration</ns1:title> <ns2:irdata>yes</ns2:irdata> <ns2:subtitle language="de"> Kontrastierung und Überblendung. Strategien der Rauminzenierung in der Narration der Johannesoffenbarung </ns2:subtitle> <ns1:language>de</ns1:language> <ns2:alt_title language="de">Phaidra für ErstanwenderInnen</ns2:alt_title> <ns1:description language="en"> This brochure was created as a part of the seminar "Oriental Music and Current Migration" (summer term 2017, University of Vienna/ Institute of Musicology). Contributions written by the students. Tutors: Kerstin Koppi, Maria Legenstein </ns1:description>

Abb. 1: Obige Darstellung zeigt beispielsweise die Gegenüberstellung von verschiedenen Kategorien des Medientyps „Web document“ (Artikel, Text).

Ausgehend von den Kernkategorien arbeitet man sich zu den Randkategorien vor:
 Top-level Elements in MODS⁵

Top-level Elements in MODS

(listed in order, read down each column)

titleInfo	language	note	location
name	physicalDescription	subject	accessCondition
typeOfResource	abstract	classification	part
genre	tableOfContents	relatedItem	extension
originInfo	targetAudience	identifier	recordInfo

Abb. 2: Top-level Elements in MODS

⁴ <https://www.w3.org/> (Stand: 04.09.2019)

⁵ <http://www.loc.gov/standards/mods/userguide/generalapp.html> (Stand: 04.09.2019)

Es gilt, immer die Darstellung in Phaidra zu kontrollieren und die Kategorienkonkordanz zu hinterfragen. Oft werden von den EndanwenderInnen gleiche Eingaben in unterschiedlichen Feldern vorgenommen. Die größte Herausforderung ist deshalb, herauszufinden, wie im Laufe des zehnjährigen Bestehens von Phaidra die verschiedenen Felder von den Nutzerinnen und Nutzern interpretiert und verwendet wurden.

Beispiel o:685260

<pre><ns0:uwwmetadata 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.at/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/metadata/etheses/V1.0"xmlns:ns2="http://phaidra.univie.ac.at/XML/metadata/extended/V1.0" xmlns:ns3="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/entity" xmlns:ns4="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/requirement"xmlns:ns5="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/educational" xmlns:ns6="http://phaidra.univie.ac.at/XML/metadata/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/histkult/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 nachhaltige Disseminations- und Archivierungsstrategie für digitale Forschungsdaten zwingend nötig. Maßnahmen zur langanhaltenden Gewährleistung der freien Verfügbarkeit in Form von Parametern wie Zugänglichkeit, Auffindbarkeit, Verwertbarkeit sowie Nachprüfbarkeit der Forschungsdaten und entsprechende technische Vorkehrungen verhindern, dass die archäologischen Informationen mittelfristig auf</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="3.5"> <titleInfo> <title>Digitale Archäologie und <u>Molino</u> San Vincenzo: Kein "Digital Dark Age" in der Toskana</title> </titleInfo> <abstract> Um den verschiedenen Prinzipien offener Wissenschaftskommunikation – wie auch dem ganzen Konzept »open science« als solchem – gerecht zu werden, ist eine nachhaltige Disseminations- und Archivierungsstrategie für digitale Forschungsdaten zwingend nötig. Maßnahmen zur langanhaltenden Gewährleistung der freien Verfügbarkeit in Form von Parametern wie Zugänglichkeit, Auffindbarkeit, Verwertbarkeit sowie Nachprüfbarkeit der Forschungsdaten und entsprechende technische Vorkehrungen verhindern, dass die archäologischen Informationen mittelfristig auf verschiedene Art und Weise unbrauchbar werden. So kann durch</pre>
--	---

Abb. 3: Anhand dieses Mappings kann nun an neuen Beispielen das 1:1 Mapping vorgenommen werden

Die bearbeitende Person ist nun in der Lage nach den neuen Metadatenregeln zu arbeiten und ausgehend von UWMETADATA nach MODS zu mappen. Dieser Prozess muss ständig betreut und überwacht werden. Das kontrollierte Vokabular und die verschiedenen Möglichkeiten müssen überprüft und auf ihre Anwendbarkeit getestet werden. Es fallen immer wieder Kategorien an, die nicht gemappt werden können. Diese werden in einer Liste zusammengeführt und für sie wird im Anschluss, in Zusammenarbeit mit dem Datenbankentwickler, nach Lösungen gesucht.

Bei all diesen Schritten ist ein gutes **Qualitätsmanagement** von größter Bedeutung. Das heißt, es müssen genug Kontrollpersonen, die das Mapping kontrollieren und hinterfragen, zur Verfügung stehen. Empfehlenswert sind regelmäßige Treffen zum Informationsaustausch, aber auch zur gemeinsamen Bearbeitung von Beispielen. Das Arbeiten im Team ist vor allem in jenen Situationen sehr hilfreich, in denen man Unsicherheiten bezüglich einiger Konkordanzen zwischen den Schemata hat oder innovative Lösungen für das Übertragen von Informationen aus Kategorien sucht, die kein Äquivalent im Zielschema besitzen.

PDF, Datei ist barrierefrei/barrierearm

Ein wichtiger Punkt für die Zukunft ist zudem die Kennzeichnung von barrierefreien Dokumenten. Siehe hierzu: Leitlinien für zugängliche Information: IKT für barrierefreien Informationszugang in der Aus- und Weiterbildung (ICT4IAL)⁶ für die verschiedenen Medientypen.

⁶ http://www.ict4ial.eu/sites/default/files/Guidelines%20for%20Accessible%20Information_DE.pdf (Stand: 04.09.2019)⁴

Hier wird unter anderem das Versehen eines Videos mit Untertiteln als Beispiel für Barrierefreiheit genannt. Aber auch bereits kleinere Maßnahmen wie die Verstellbarkeit der Schriftgröße und Schriftart in Dateiformaten sind Merkmale eines barrierearmen Textdokuments.

Schlussfolgerung

Es bleibt die Erkenntnis, dass man gemäß dem alten Zitat von Horaz einfach beginnen muss und dann ist die halbe Arbeit auch schon getan. Wichtig dabei ist, dass man im Team arbeitet, die einzelnen Schritte abspricht und dokumentiert. Oder wie Rudolf Wimmer es formuliert: „Team ist verteilte Intelligenz“ und „Teamförmige Strukturen werden heute überall dort dringender denn je gebraucht, wo es besonders komplexe, konflikthafte oder mit Unsicherheit behaftete Probleme zu bearbeiten gilt [...]“⁷

Eine Qualitätskontrolle muss von Anfang an in den Prozess des Mappens eingebunden sein!

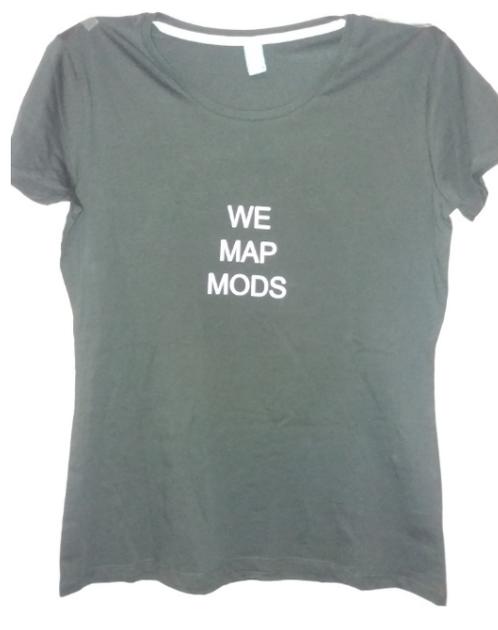


Abb. 4: Motivation

» *Dimidium facti, qui coepit, habet: sapere aude, / incipe.* «
Horaz

⁷ Die Presse vom 10./11. März 2018 S. K1