



nestor Handbuch:
**Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung**

10.1.1 Metadata Encoding and Transmission Standard
Das METS Abstract Model
– Einführung und Nutzungsmöglichkeiten

Herausgeber:

Heike Neuroth
Hans Liegmann
Achim Oßwald
Regine Scheffel
Mathias Jehn

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Im Auftrag von:

nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
<http://www.langzeitarchivierung.de>

**Dieser Artikel ist ein Auszug aus dem
nestor Handbuch:
Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung**

Dieser Artikel ist verfügbar unter der URL:

http://nestor.sub.uni-goettingen.de/handbuch/artikel/text_84.pdf

Die Online Version des Handbuches unter der URL:

<http://nestor.sub.uni-goettingen.de/handbuch/>

Kontakt:

Niedersächsische Staats- und Universitätsbibliothek Göttingen

Dr. Heike Neuroth

Forschung und Entwicklung

Papendiek 14

37073 Göttingen

neuroth@sub.uni-goettingen.de

Tel. +49 (0) 55 1 39 38 66

Der Inhalt steht unter folgender Creative Commons Lizenz:

<http://creativecommons.org/licenses/by-nc-sa/2.0/de/>



10.1.1 Metadata Encoding and Transmission Standard Das METS Abstract Model – Einführung und Nutzungsmöglichkeiten

von Markus Enders

Einführung

Ausgehend von den Digitalisierungsaktivitäten der Bibliotheken Mitte der 90iger Jahre entstand die Notwendigkeit, die so entstandenen Dokumente umfassend zu beschreiben. Diese Beschreibung muß im Gegensatz zu den bis dahin üblichen Verfahrensweisen nicht nur einen Datensatz für das gesamte Dokument beinhalten, sondern außerdem einzelne Dokumentbestandteile und ihre Abhängigkeiten zueinander beschreiben. Nur so lassen sich gewohnte Nutzungsmöglichkeiten eines Buches in die digitale Welt übertragen. Inhaltsverzeichnisse, Seitennummern sowie Verweise auf einzelne Bilder müssen durch ein solches Format zusammengehalten werden.

Zu diesem Zweck wurde im Rahmen des „Making Of Amerika“ Projektes Ebind entworfen. Ebind selber war jedoch ausschließlich nur für Digitalisate von Büchern sinnvoll zu verwenden.

Um weitere Medientypen sowie unterschiedliche Metadatenformate einbinden zu können, haben sich Anforderungen an ein komplexes Objektformat ergeben. Dies setzt ein abstraktes Modell voraus mit Hilfe dessen sich Dokumente flexibel modellieren lassen und als Container Format verschiedene Standards eingebunden werden können. Ein solches abstraktes Modell bildet die Basis von METS und wird durch das METS-XML-Schema beschrieben. Daher wird METS derzeit auch fast ausschließlich als XML serialisiert in und Form von Dateien gespeichert. Als Container Format ist es in der Lage weitere XML-Schema (so genannte Extension Schemas) zu integrieren.

Das METS Abstract Model

Das METS „Abstract Model“ beinhaltet alle Objekte innerhalb eines METS Dokuments und beschreibt deren Verhältnis zueinander. Zentraler Bestandteil eines METS-Dokuments ist eine Struktur. Diese Struktur kann eine logische oder physische Struktur des zu beschreibenden Dokumentes (bspw. eines Textes) abbilden. Das bedeutet, daß eine Struktur aus mindestens einer Struktureinheit (bspw. einer Monographie) besteht, die weitere Einheiten beinhalten kann. Somit läßt sich eine Struktur als Baum modellieren. In METS wird diese Struktur in der <structMap>-Sektion gespeichert. Jedes METS-Dokument kann mehrere Strukturen in separaten Sektionen beinhalten. So lassen sich bspw. logische und physische Strukturen voneinander trennen. In einer Struktur läßt sich das Inhaltsverzeichnis eines Werkes dokumentieren; in der anderen Struktur kann das Buch (mit Seiten als unterliegende Struktureinheiten) beschrieben werden. Das „Abstract Model“ besitzt eine weitere Sektion – die <structLink> Sektion –, um Verweise zwischen unterschiedlichen Strukturen zu speichern.

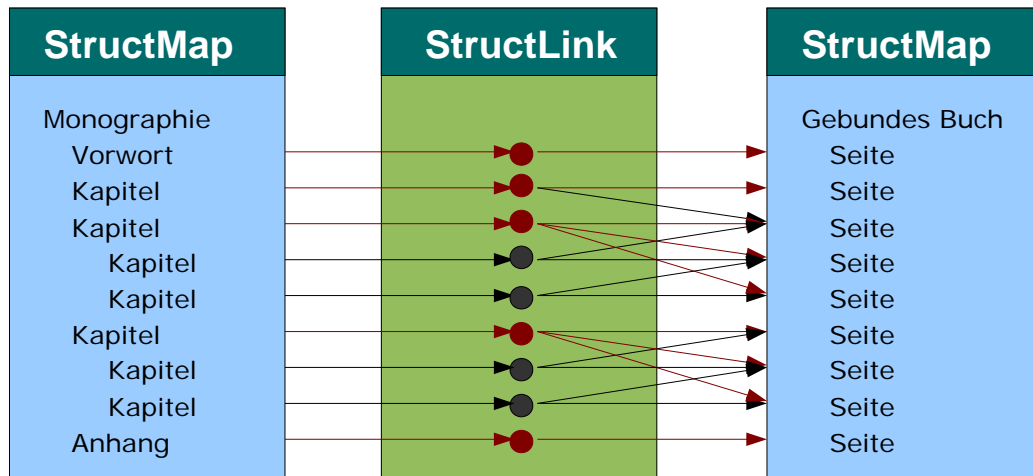


Abbildung 1: Verknüpfung von zwei Strukturen im Abstract-Model

Neben den Strukturen berücksichtigt das Modell auch Metadaten, wobei darunter nicht nur bibliographische Metadaten zu verstehen sind. Aus diesem Grund unterteilt das Modell die Metadaten in deskriptive Metadaten (in der Descriptive Metadata Section) und administrative Metadaten (in der Administrative Metadata Section). Während die deskriptiven Metadaten bibliographische Informationen enthalten, werden Informationen zu Rechteinhabern, Nutzungsrechte, technische Informationen zu einzelnen Dateien oder Langzeitarchivierungsmetadaten in den administrativen Metadaten gespeichert. Für beide Metadattentypen können beliebige Schema, so genannte „Extension Schema“ genutzt werden, die in der jeweiligen Sektion gespeichert werden. Auf diese Weise lassen sich sowohl XML-Metadatenschema (MARC XML, MODS, Dublin Core simple) als auch Text-/ bzw. Binärdaten einbinden (bspw. PICA-Datensätze).

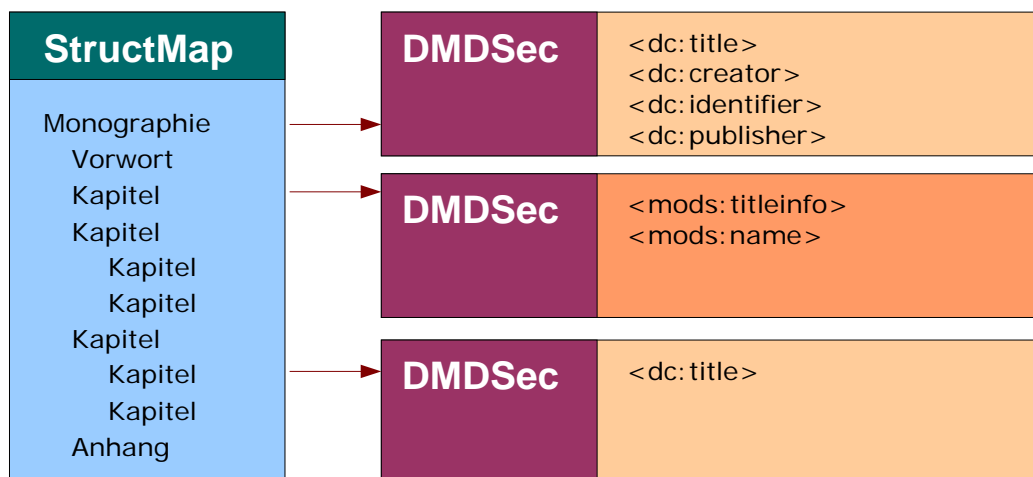


Abbildung 2: Verweis auf Metadatensektionen im METS-Abstract-Model

Neben den Struktureinheiten und ihren zugehörigen Metadaten spielen auch Dateien bzw. Streams eine wesentliche Rolle, da letztlich in ihnen die durch das METS-Dokument beschriebenen Inhalte gespeichert sind. Eine Datei kann bspw. den Volltext eines Buches, die Audioaufnahme einer Rede oder eine gescannte Buchseite als Image enthalten. Entsprechende Daten können in ein METS-

Dokument eingebunden werden (bspw. BASE64 encoded in die METS-XML Datei eingefügt werden) oder aber mittels xlink referenziert werden. Ein METS-Dokument kann also als Container alle für ein Dokument notwendigen Dateien enthalten oder referenzieren, unabhängig davon, ob die Dateien lokal oder auf entfernten Servern vorhanden sind.

Grundsätzlich müssen alle für ein METS-Dokument relevanten Dateien innerhalb der File-Sektion aufgeführt werden. Innerhalb der File-Sektion können Gruppen (File-Groups) von Dateien gebildet werden, wobei die Abgrenzungskriterien zwischen einzelnen Gruppen nicht in METS fest definiert sind. Je nach Modellierung lassen sich Dateien bspw. nach technischen Parametern (Auflösung oder Farbtiefe von Images), Anwendungszweck (Anzeige, Archivierung, Suche) oder sonstigen Eigenschaften (Durchlauf bestimmter Produktionsschritte) den einzelnen Gruppen zuordnen.

Das METS-Abstract-Model erlaubt das Speichern von administrativen Metadaten zu jeder Datei. Generelle, für jede Datei verfügbare technische Metadaten wie Dateigröße, Checksummen etc. lassen sich direkt in METS speichern. Für weiterführende Metadaten kann mit jeder Datei eine oder mehrere Administrative Metadatensektion(en) verknüpft werden, die bspw. Formatspezifische Metadaten enthalten (für Images könnten die Auflösungsinformationen, Informationen zur Farbtiefe etc. sein).

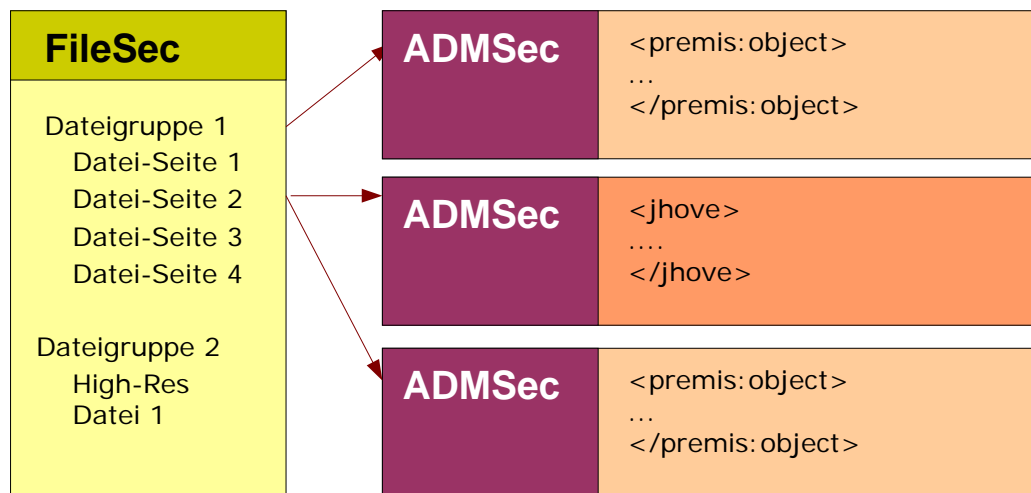
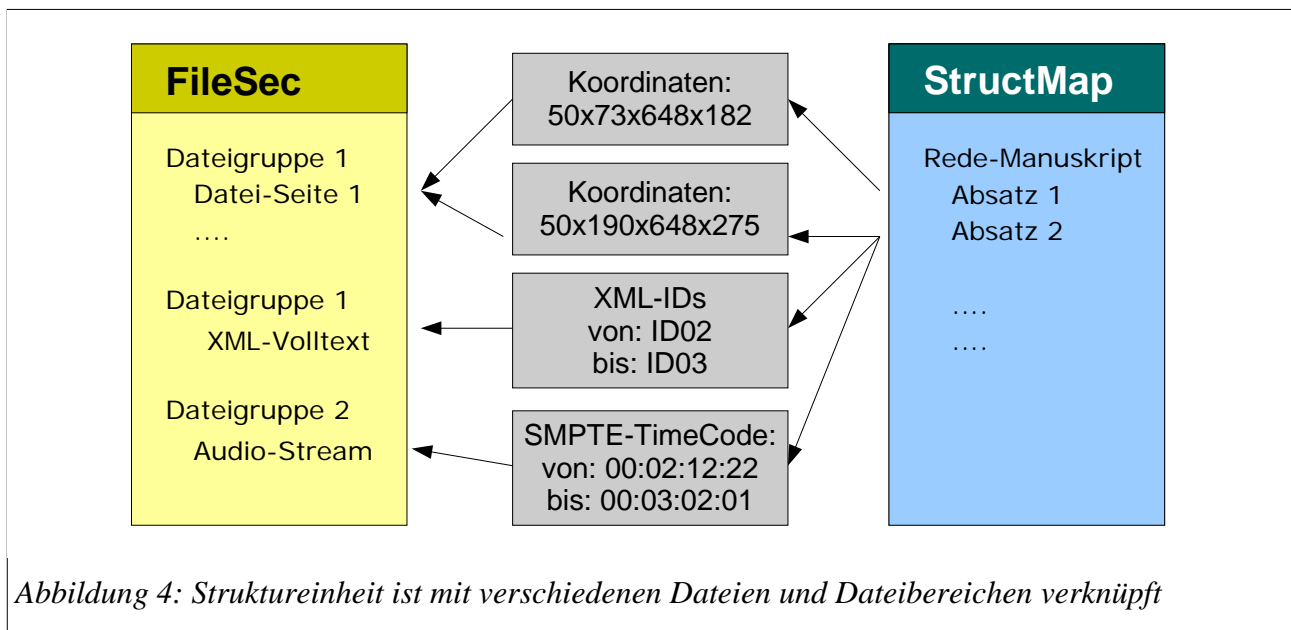


Abbildung 3: Administrative Metadaten zu Dateien

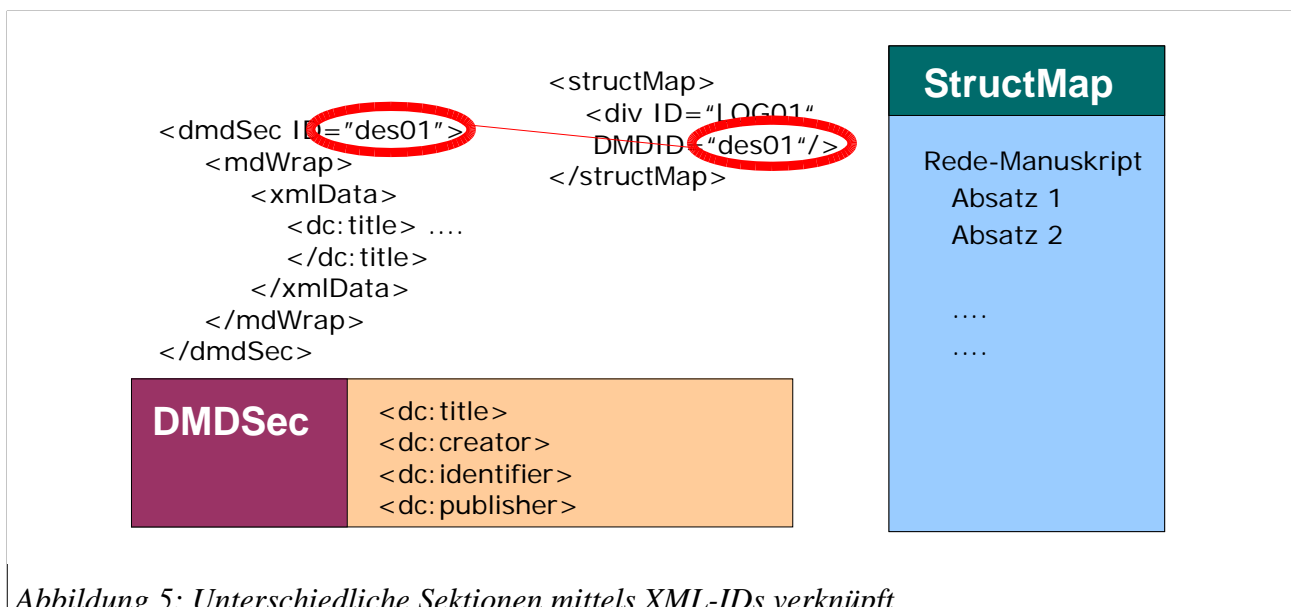
Dateien sind darüber hinaus mit Struktureinheiten verknüpft. Die Struktureinheit, die eine einzelne Buchseite repräsentiert, kann somit mit einer einzelnen Datei, die ein Image dieser Seite beinhaltet, verknüpft werden. Das „METS-Abstract-Model“ stellt hierzu eine N:M Verknüpfung bereit. Das bedeutet, daß eine Datei von mehreren Struktureinheiten (auch aus unterschiedlichen Struktursektionen) aus verknüpft werden kann, genauso wie eine Struktureinheit mehrere Dateien verknüpfen kann. Im Ergebnis heißt das, daß der Struktureinheit vom Typ „Monographie“ sämtliche Imagedateien ein gescanntes Werken aus direkt unterstellt wird.

Für die Verknüpfung von Dateien sieht das „METS-Abstract-Model“ noch weitere Möglichkeiten vor. So lassen sich mehrere Verknüpfungen hinsichtlich ihrer Reihenfolge beim abspielen bzw. anzeigen bewerten. Dateien können entweder sequentiell angezeigt (Images eines digitalisierten Buches) als auch parallel abgespielt (Audio- und Videodateien gleichen Inhalts) werden. Darüber hinaus kann nicht nur auf Dateien, sondern auch in Dateiobjekte hinein verlinkt werden. Diese Verlinkungen sind u.a. dann sinnvoll, wenn Einheiten beschrieben werden, die aus technischen Gründen nicht aus der Datei herausgetrennt werden können. Das können bestimmte Teile eines Images sein (bspw. einzelne Textspalten) oder aber konkrete zeitliche Abschnitte einer Audioaufnahme. In der Praxis lassen sich so einzelne Zeitabschnitte eines Streams markieren und

bspw. mit inhaltlich identischen Abschnitten eines Rede-Manuskriptes taggen. Das METS-Dokument würde über die Struktureinheit eine Verbindung zwischen den unterschiedlichen Dateien herstellen.



Das METS-Abstract-Model nutzt intensiv die Möglichkeit, einzelne Sektionen miteinander zu verknüpfen. Da METS überwiegend als XML realisiert ist, wird diese Verknüpfung über Identifier realisiert. Jede Sektionen verfügt über einen Identifier, der innerhalb des XML- Dokumentes eindeutig ist. Er dient als Ziel für die Verknüpfungen aus anderen Sektionen heraus. Aufgrund der XML-Serialisierung muß er weiteren Anforderungen genügen. Außerdem muß dieser Identifier mit den Regeln für XML-IDs verträglich sein. Ferner muß bei Verwendung von weiteren Extension Schemas darauf geachtet werden, daß die Eindeutigkeit der Identifier aus dem unterschiedlichen Schema nicht gefährdet wird, da diese üblicherweise alle im gleichen Namensraum existieren.



Wie deutlich geworden ist, stellt das METS-Abstract-Model sowie des XML-Serialisierung als METS-XML Schema lediglich ein grobes Modell da, welches auf den jeweiligen Anwendungsfall angepasst werden muß. Die Verwendung von Extension Schema sollte genauso dokumentiert

werden wie die Nutzung optionaler Elemente und Attribute in METS. Dabei sollte vor allem auch die Transformation realer, im zu beschreibenden Dokument vorhandene Objekte in entsprechende METS-Objekte bzw. METS-Sektionen im Vordergrund stehen. Eine Struktursektion kann bspw. lediglich logische Einheiten (bspw. das Inhaltsverzeichnis eines Buches) umfassen als auch bestimmte physische Einheiten (bspw. einzelne Seiten) enthalten. Eine weitere Option wäre es, bestimmte Einheiten in eine zweite separate Struktur auszugliedern. Jede dieser Optionen mag für bestimmte Arten von Dokumenten sinnvoll sein.

Dokumentation

Damit ein METS-Dokument von unterschiedlichen Personen verstanden werden kann, ist es notwendig, neben den formalisierten METS-Schemas auch eine weitere Dokumentation der konkreten Implementierung von METS zu erstellen. Das METS-Profile-Schema bietet daher eine standardisierte Möglichkeit, eine solche Dokumentation zu erstellen, in dem sie eine Grobstrukturierung vorgibt und sicherstellt, daß alle wesentlichen Bereiche eines METS-Dokuments in der Dokumentation berücksichtigt werden.

Um ein solches Profil auf der offiziellen METS-Homepage veröffentlichen zu können, wird es durch Mitglieder des METS-Editorial-Board verifiziert. Nur verifizierte METS-Profile werden veröffentlicht und stehen auf der Homepage zur Nachnutzung bereit. Sie können von anderen Institutionen adaptiert und modifiziert werden und somit erheblich zur Reduktion der Entwicklungszeit einer eigenen METS-Implementierung beitragen.

Fazit

Aufgrund der hohen Flexibilität des METS Abstract Models wird METS in einer großen Zahl unterschiedlicher Implementierungen für sehr verschiedene Dokumententypen genutzt. Neben der ursprünglichen Anwendung, digitalisierte Büchern zu beschreiben, gibt es heute sowohl METS-Profile für Webseitenbeschreibungen (aus dem Bereich der Webseitenarchivierung) sowie Audio- und Videodaten. Während in den ersten Jahren METS überwiegend zum Beschreiben komplexer Dokumente genutzt wurde, um diese dann mittels XSLTs oder DMS-Systeme verwalten und anzeigen zu können, kommt heute METS gerade im Bereich der Langzeitarchivierung wachsende Bedeutung zu. METS ist heute für viele Bereiche, in denen komplexe Dokumente beschrieben werden müssen ein De-facto-Standard und kann sowohl im universitären als auch im kommerziellen Umfeld eine große Zahl an Implementierungen vorweisen. Ein großer Teil derer sind im METS-Implementation Registry auf der METS-Homepage (<http://www.loc.gov/mets>) nachgewiesen.