



Verfahren zur Produktion interoperabler Metadaten in digitalen Dokumentenverarbeitungsprozessen

Management Summary

Der Bericht untersucht Möglichkeiten zur Erhöhung der Auffindbarkeit digitaler Dokumente im Web, die sich durch die Anreicherung von Dokumenteigenschaften während der Dokumentenerstellung bei Nutzung eines Autorenwerkzeugs ergeben. Hierzu werden Metadaten verwendet, mit deren Hilfe maschinell auf den Dokumenteninhalt, den Autor sowie Themengebieten rückgeschlossen werden kann.

Der Einsatz weit verbreiteter und interoperabler Metadatenstandards für diesen Zweck ermöglicht die Interpretation der Dokumenteigenschaften durch eine Vielzahl von heterogenen Systembetreibern. Es liegt nahe, diese Metadaten bereits durch den Autor in den Dokumentenver- und -erarbeitungsprozess einfließen zu lassen, da er der entscheidende Wissensträger für die entstehende Publikation ist. Aus diesem Grund wurden drei weitverbreitete Dokumentenformate und die zugehörigen marktführenden Autorenwerkzeuge (Microsoft Office, OpenOffice.org, Adobe Acrobat) hinsichtlich ihrer Funktionalität und ihrer Gebrauchstauglichkeit zur Integration interoperabler Metadaten in Dokumenten untersucht.

Für die öffentliche Wahrnehmung einer Publikation im Web ist die Qualität der Metadaten entscheidend. Basierend auf den vorgenommenen Betrachtungen werden Qualitätskriterien für die Integration interoperabler Metadaten in Dokumenten definiert und Funktionen von Autorenwerkzeugen ermittelt, die für die Generierung dieser Metadaten durch einen Produzenten relevant und somit für die Auswahl des zu verwendenden Autorenwerkzeuges ausschlaggebend sind.

Neben einer ausführlichen Beschreibung der erhobenen Qualitätskriterien wurde eine Checkliste zur Evaluation von Dokumentenformaten und Autorenwerkzeugen in Hinsicht auf ihre Eignung für die Erfassung und den Transfer von Metadaten während des Dokumententstehungsprozesses entwickelt. Diese Checkliste kann bei Auswahl und Bewertung von Autorenwerkzeugen in Verbindung mit dem vorliegenden Bericht verwendet werden.

Inhaltsverzeichnis

1	Einführung	2
2	Der digitale Dokumentenverarbeitungsprozess	3
3	Analyse von Autorenwerkzeugen und zugrundeliegenden Dokumentformaten	4
3.1	Microsoft-Office-Dokumente	5
3.1.1	Office Open XML	5
3.1.2	Metadatenanreicherung durch Microsoft Office	7
3.2	OpenOffice.org-Dokumente	9
3.2.1	OASIS OpenDocument Format	9
3.2.2	Metadatenanreicherung durch OpenOffice.org	10
3.3	PDF-Dokumente	12
3.3.1	Portable Document Format	12
3.3.2	Metadatenanreicherung durch Adobe Acrobat	14
3.3.3	PDF-Generierung und Metadatenübernahme aus Dritt-Applikationen	17
4	Metadaten im Dokumentenverarbeitungsprozess	18
4.1	Metadatenübernahme aus einer digitalen Ressource	18
4.2	Metadatenanreicherung im Archivierungsprozess	20
5	Definition von Gütekriterien	21
5.1	Gütekriterien für interoperable Metadaten in Dokumentenformaten	21
5.2	Qualitätskriterien für Autorenwerkzeuge zur Erstellung interoperabler Metadaten	23
6	Schlussbetrachtung	24
A	Checkliste für Qualitätskriterien	25
A.1	Interoperabilität von Metadaten in Dokumentenformaten	25
A.2	Funktionen zur Generierung interoperabler Metadaten in Autorenwerkzeugen	26

1 Einführung

In einer stetig wachsenden digitalen Informationsgesellschaft rückt die Nachnutzung und gemeinsame Nutzung von digitalen Ressourcen immer mehr in den Vordergrund. Für die Auffindbarkeit digitaler Ressourcen im Web ist neben der inhaltlichen Aufbereitung auch eine semantische Anreicherung der Dokumente notwendig. Im Kontext von Dokumentenverarbeitungsprozessen versteht man unter semantischer Anreicherung, das Hinzufügen von Detailinformationen zu einer Datei (Dokumenteigenschaften), durch welche eine Datei und ihr Inhalt beschrieben werden. Zu den Dokumenteigenschaften gehören Informationen wie der Titel, der Name des Autors, das Thema und Stichwörter, mit denen das Thema oder der Inhalt des Dokuments in Beziehung stehen. Die semantische Anreicherung dient insbesondere dazu, Ressourcen anhand der vorliegenden Zusatzinformationen maschinell identifizieren und somit auf ihren Inhalt rückschließen zu können.

Eine semantische Anreicherung ermöglicht eine verbesserte Dateiverwaltung in Dokumentenrepositorien sowie in Content-Management-Prozessen. Für eine möglichst hohe Interoperabilität in heterogenen Dokumentenverarbeitungssystemen ist die Einhaltung von Metadatenstandards ausschlaggebend. Je früher in einem Publikationsprozess interoperable Informationen über ein Dokument in einer einfach nachnutzbaren Form entstehen, desto wahrscheinlicher und effizienter ist ihre (automatisierte) Nachnutzung.

Ziel dieser Arbeit ist es, ausgewählte Dokumentenformate auf ihre grundsätzliche Eignung für die Integration interoperabler Metadaten zu untersuchen. Aufbauend auf den Formatbetrachtungen werden für die Dateiformate vorgesehene Autorenwerkzeuge und deren Funktionen bezüglich ihrer Gebrauchstauglichkeit sowie der Qualität ihrer semantischen Anreicherung analysiert. Als Autorenwerkzeuge werden in dieser Arbeit Softwareprodukte zur Erstellung und Bearbeitung von Dokumenten bezeichnet. Weiterführend werden die Vorteile der Dokumentenanreicherung für den Dokumentenverarbeitungsprozess in Repositorien und eine (semi-)automatische Metadatennachnutzung diskutiert.

Die Untersuchungen konzentrieren sich auf weitverbreitete Dokumentenformate, die insbesondere typisch für den Einsatz in Büroanwendungen sind. Ausgenommen von der Untersuchung sind webbasierte Autorenwerkzeuge wie beispielsweise Google Docs, da eine Vorauswahl dieser nicht repräsentativ für die Heterogenität der auf dem Markt zur Verfügung stehenden Systeme wäre.

Basierend auf den vorgenommenen Betrachtungen werden abschließend Qualitätskriterien für die Integration interoperabler Metadaten definiert. Ebenso werden Qualitätskriterien und Funktionen von Autorenwerkzeugen ermittelt, die für die Generierung interoperabler Metadaten durch einen Produzenten relevant sind.

Diese Arbeit soll Autoren einen Überblick über existierende Softwareprodukte und ihre Anwendungsbereiche geben und ihnen ein pragmatisches Werkzeug zur Verfügung stellen, um ihre Publikationen mit interoperablen Metadaten zu versehen. Des Weiteren können die Ergebnisse Metadatenmanagern von Repositorien dabei helfen, adäquate Autorenwerkzeuge für Publizierende des entsprechenden Repositoriums vorzuschlagen.

2 Der digitale Dokumentenverarbeitungsprozess

Ein Dokumentenverarbeitungsprozess ist in eine Anzahl definierter Teilprozesse unterteilt. Für die Verdeutlichung der Teilprozesse in Archivierungssystemen und der zugrundeliegenden Workflows wird das Open Archival Information System (OAIS) Reference Model¹ hier angewandt. Die Verwendung von Metadaten wird jeweils im Kontext des zugehörigen Teilprozesses diskutiert.

Jeder Dokumentenverarbeitungsprozess beginnt mit der Produktion eines Dokumentes bzw. einer Ressource. Der für die Produktion verantwortliche Prozessteilnehmer ist in der Regel der Autor. Unter Umständen gibt es weitere Produzenten, beispielsweise Lektoren oder Designer. Für die Ressourcenproduktion wird ein Autorenwerkzeug verwendet. Bei der Produktion nicht-digitaler Ressourcen oder Ressourcen auf Medienträgern sind zusätzlich Prozessteilnehmer involviert, die für die physikalische Umsetzung der Ressource verantwortlich sind.

An die Produktion einer Ressource schließt i. d. R. die Veröffentlichung dieser an. Der Teilprozess der Veröffentlichung muss hinsichtlich der Zielgruppe der jeweiligen Publikation näher betrachtet werden. Diese kann ein abgeschlossener Nutzerkreis, beispielsweise bei einer organisationsinternen Distribution, oder die breite Öffentlichkeit, wie z. B. im Falle einer Verlagspublikation, sein. Im letztgenannten Fall sind neben dem Autor zusätzliche Akteure in den Teilprozess involviert.

Ein insbesondere für digitale Ressourcen an Bedeutung gewinnender Teilprozess ist die Archivierung von Publikationen. Digitale Repositorien als Dokumentenserver stehen in der Verantwortung, die Ressourcen für interessierte Nutzer dauerhaft auffindbar und zugänglich zu machen sowie die Ressourcen im Sinne einer Langzeitarchivierung aufzubewahren.

Der Teilprozess der Distribution ist die Schnittstelle zwischen dem Endnutzer und dem Archivierungssystem. Wenn ein Nutzer eine für ihn von Interesse erscheinende Ressource im Repository gefunden hat, bezieht er diese Ressource sowie gegebenenfalls Zusatzinformationen vom Bereitsteller dieser Ressource.

Der abschließende Teilprozess des Dokumentenverarbeitungsprozesses stellt die Nutzung der Ressource dar. Die Nutzung kann nach dem Umgang mit der Ressource differenziert werden. Entweder entschließt sich der Endnutzer ausschließlich zu einer Rezeption der Ressource oder zu einer Weiterbearbeitung dieser. Die Weiterbearbeitung ist ein typischer Anwendungsfall für abgeschlossene Systeme, wenn beispielsweise kollaborativ an einem unternehmensinternen Dokument gearbeitet wird.

Anhand des Open Archival Information System (OAIS) Reference Model lässt sich die Bewegung von Informationsobjekten (Information Packages) in Bezug auf ein Archivierungssystem verdeutlichen. In Archivierungssystemen werden zu jedem Information Package deskriptive Informationen vorgehalten, um die Pakete (sowie die einzelnen enthaltenen Ressourcen) such- und auffindbar zu machen.

Das OAIS Reference Model unterscheidet drei Arten von Information Packages:

- Submission Information Packages (SIP) werden vom Produzenten an das Archiv gesendet
- Archival Information Packages (AIP) werden im Archiv aufbewahrt

¹ http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683

- Dissemination Information Packages (DIP) werden vom Archiv zum Konsumenten gereicht

In Abbildung 1, die das Functional Model des OAIS Reference Model zeigt, wird der Fluss und die Umwandlung der einzelnen Pakete durch das Archivierungssystem verdeutlicht.

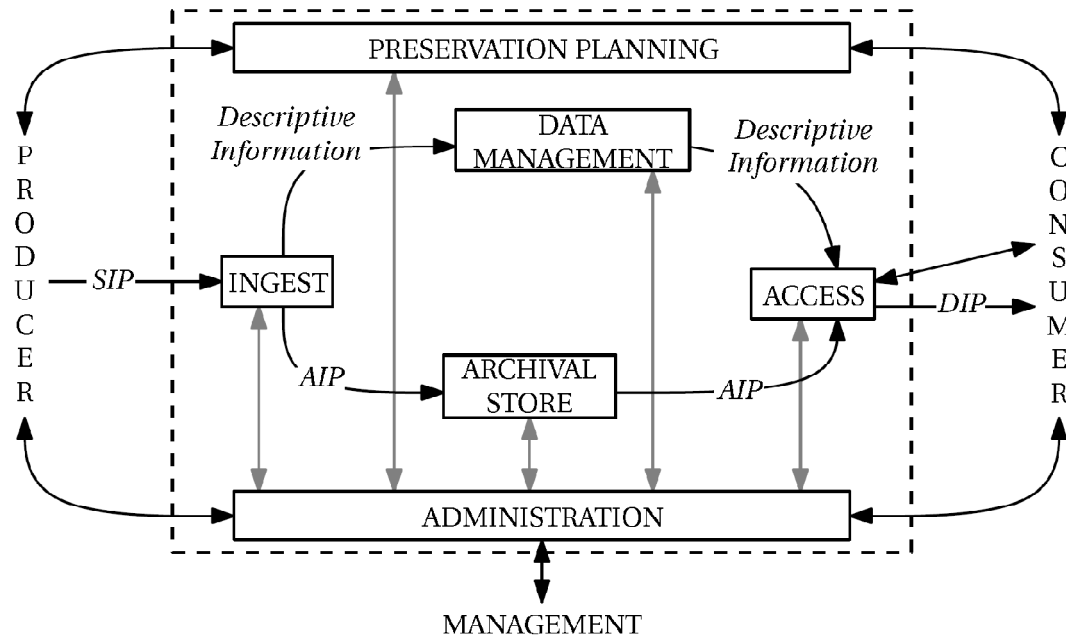


Abbildung 1: OAIS Functional Model

Für die frühzeitige Integration von Metadaten in eine Ressource, ist es von besonderer Wichtigkeit zu definieren, welche Metadaten bereits während der Ressourcenproduktion durch den Produzenten in die Ressource integriert werden können, um eine Übernahme in das SIP und eine Extraktion der Metadaten für die Descriptive Information des Archivierungssystems zu gewährleisten.

Der gegenläufige Ansatz wäre, die Metadaten während des Ingest-Prozesses zu generieren. Dies kann beispielsweise formularbasiert beim Hochladen des Dokumentes in das Archivierungssystem erfolgen. Allerdings ist es nicht sicher, dass das Archivierungssystem die Ressource später wieder mit diesen Metadaten anreichert.

In dieser Studie wird untersucht, wie beschreibende Metadaten eines Dokumentes bereits durch den Produzenten unter Zuhilfenahme eines Autorenwerkzeuges in einer digitalen Ressource integriert werden können.

Dabei gilt es, zu definieren, welche Metadaten in diesem frühen Dokumentenverarbeitungsteilprozess für die gesamte Verarbeitungskette von Relevanz sind. Anhand exemplarisch ausgewählter Autorenwerkzeuge wird aufgezeigt, welche Möglichkeiten das zugrundeliegende Format und ein auf dieses Format zugeschnittenes Autorenwerkzeug mit sich bringen müssen, um Produzenten bei der Metadatenanreicherung unterstützen zu können.

3 Analyse von Autorenwerkzeugen und zugrundeliegenden Dokumentformaten

Aufgrund ihres weitverbreiteten Einsatzes wurden die folgenden drei nicht-webbasierten Autorenwerkzeuge für die Betrachtungen ausgewählt:

1. Microsoft Office
2. OpenOffice.org
3. Adobe Acrobat

Ausgehend vom zugrundeliegenden Format des Autorenwerkzeugs wird analysiert, welche interoperablen Metadaten in Dokumente integriert werden können und welche Möglichkeiten das Autorenwerkzeug für eine gebrauchstaugliche Integration dieser für den Produzenten bietet. Die zugrundeliegenden Formate sind:

1. Microsoft Office: Office Open XML
2. OpenOffice.org: OASIS OpenDocument Format
3. Adobe Acrobat: Portable Document Format

3.1 Microsoft-Office-Dokumente

Die erste Analyse bezieht sich auf Microsoft Office 2007 und dem zugrundeliegenden Dateiformat Office Open XML.

3.1.1 Office Open XML

Das von Microsoft entwickelte, offene XML-basierte Dateiformat für Office-Dokumente wurde im Jahr 2006 als Ecma-Standard ECMA-376², im Jahr 2008 als ISO/IEC-Norm 29500:2008³ verabschiedet. Generell ist das Format unter dem Namen Office Open XML (OOXML) bekannt. OOXML wird in Microsoft Office Word, Microsoft Office Excel und Microsoft Office PowerPoint verwendet.

Grund für die Einführung eines neuen und vor allem offenen Formats durch Microsoft war, Kundenwünsche, die eine Dokumenterstellung aus unterschiedlichen Datenquellen und eine Verbesserung beim Daten- und Dateiaustausch in heterogenen Geschäftssystemen forderten, zu berücksichtigen. Durch den Einsatz eines offenen Formates wird der Umgang mit den erstellten Dateien jedem XML-verarbeitenden Programm und somit auch für Nicht-Microsoft-Produkte ermöglicht. Ein zusätzlicher Gewinn bezüglich der Interoperabilität in Verarbeitungsprozessen entsteht durch die Ausweisung von spezifizierten Dokumentenmetadaten in der XML-Dokumentstruktur. Weitere Vorteile gegenüber den zuvor verwendeten binären Dateiformaten sind die Größenreduzierung von Dateien um bis zu 75 Prozent sowie verbesserte Wiederherstellungsmechanismen für beschädigte Dateien⁴.

OOXML ist ein Containerformat, welches den Konventionen der Open Packaging Conventions (OPC) folgt. Jede OOXML-Datei setzt sich aus einer beliebigen Anzahl von Komponenten zusammen. Die enthaltenen Komponenten werden durch das ZIP-Kompressionsverfahren im Container komprimiert. Die Beziehungen zwischen Komponenten werden durch Beziehungskomponenten festgelegt. Aufgrund der dadurch gegebenen Modularität ist es relativ einfach, einzelne Komponenten zu identifizieren, anzupassen oder sogar auszutauschen.

Die einzelnen Office-Dokumenttypen basieren auf für ihren Dokumenttyp zugeschnittene Auszeichnungssprachen. Microsoft Office verwendet WordprocessingML, Microsoft Excel SpreadsheetML und Microsoft PowerPoint PresentationML. Zeichnungsdokumente werden auf Basis von DrawingML gespeichert und Zeichnungsobjekte innerhalb anderer Dokumente in VML. Weitere XML-basierte Auszeichnungssprachen existieren für Custom XML Data Properties (nicht-standardisiertes XML),

² <http://www.ecma-international.org/publications/standards/Ecma-376.htm>

³ <http://www.iso.org/iso/pressrelease.htm?refid=Ref1181>

⁴ <http://msdn.microsoft.com/de-de/library/aa338205%28office.12%29.aspx>

für Dateieigenschaften sowie für mathematische Formeln und bibliographische Daten. Ein OOXML-Container kann neben XML-basierten Dateien auch Binärdateien als Komponenten beinhalten. Beispielsweise werden Bilder und Audiodateien im Container in binären Dateiformaten hinterlegt. Für die Komponenten in einem Container muss jeweils ein Inhaltstyp deklariert werden, um einen akkuraten Umgang mit den Komponenten zu gewährleisten.

Im Folgenden werden nicht die Metadaten innerhalb eines OOXML-Containers für die Ausweisung von Beziehungen (Package Relationships) und der Repräsentation von Komponenten (Content Types) fokussiert, sondern insbesondere die Komponenten für die generelle Dokumenteigenchaftenauszeichnung. Alle Office-Dokumenttypen speichern Dokumenteigenschaften (Core Properties) nach demselben Schema.

Prinzipiell kann man Dokumenteigenschaften in OOXML-Dokumenten in drei Gruppen unterteilen:

- **Kern- oder Standardeigenschaften (Core Properties):** beschreibende Dokumenteigenschaften (z. B. Autor, Titel und Thema), welche durch den Benutzer in das Dokument integriert werden können und in vordefinierten Metadatenelementen abgelegt werden. Automatisch hinzugefügte Dateisystemeigenschaften (z. B. Dateigröße oder die Daten des Erstellens oder letzten Ändern einer Datei) und Statistiken, die aus den Office-Programmen generiert werden (z. B. die Anzahl von Wörtern oder Zeichen in einem Dokument)⁵. Auch automatisch hinzugefügte Dokumenteigenschaften werden in vordefinierten Metadatenelementen im Dokument hinterlegt.
- **Erweiterte und fortgeschrittene benutzerdefinierte Eigenschaften (Extended and Advanced Custom Properties)** sind zusätzliche Eigenschaften, die durch die Definition von Name-Type-Value-Tripeln formularbasiert in der Office-Applikation frei angegeben werden können. In Microsoft Office wird zwar eine Auswahlliste für vordefinierte Namen bereitgestellt, eine Interoperabilität erweiterter benutzerdefinierter Dokumenteigenschaften ist allerdings nur in abgestimmten Umgebungen zu erreichen.
- **Benutzerdefinierte Eigenschaften (Custom Document Properties)** können nur mithilfe eines Eigenschaftenverwaltungsservers oder einem Microsoft Office InfoPath 2007-Formulars zum Dokument hinzugefügt werden. Diese Eigenschaften dienen insbesondere der Workflowoptimierung in abgeschlossenen Umgebungen. Die Interoperabilität der Metadaten steht nicht im Vordergrund.

Für die Ausweisung der Kerneigenschaften wird auf das Dublin Core Metadata Element Set⁶, die DCMI Metadata Terms⁷ und das Microsoft Core Property Schema⁸ zurückgegriffen. In Tabelle 1 werden die Elemente der Kerneigenschaften und ihr jeweiliger Ursprung dargestellt.

Element	Namespace
category	Open Packaging Conventions
contentStatus	Open Packaging Conventions
created	Dublin Core
creator	Dublin Core

⁵ <http://office.microsoft.com/de-de/word/HA100475241031.aspx>

⁶ <http://dublincore.org/documents/dces/>

⁷ <http://dublincore.org/documents/dcmi-terms/>

⁸ <http://schemas.openxmlformats.org/package/2006/metadata/core-properties>

description	Dublin Core
identifier	Dublin Core
keywords	Open Packaging Conventions
language	Dublin Core
lastModifiedBy	Open Packaging Conventions
lastPrinted	Open Packaging Conventions
modified	Dublin Core
revision	Open Packaging Conventions
subject	Dublin Core
title	Dublin Core
version	Open Packaging Conventions

Tabelle 1: Kerneigenschaften in OOXML

3.1.2 Metadatenanreicherung durch Microsoft Office

Kern- oder Standardeigenschaften können durch den Benutzer In Office Word 2007, Office Excel 2007 oder Office PowerPoint 2007 im Dokumentinformationsbereich angezeigt oder geändert werden. Der Dokumentinformationsbereich kann durch Nutzung der Microsoft Office-Schaltfläche, mittels Auswahl des Menüpunktes „Vorbereiten“ und anschließendem Anklicken von „Eigenschaften“ angezeigt werden.

Abbildung 2: Dokumenteigenschaften in Microsoft Office

Wie bereits verdeutlicht, werden eine Reihe von Kern- oder Standardeigenschaften automatisiert durch das Autorenwerkzeug hinzugefügt. Auf die Integration dieser Metadaten hat der Anwender keine direkte Einflussnahme.

Erweiterte und fortgeschrittene benutzerdefinierte Eigenschaften können durch anklicken des Pfeils neben Dokumenteigenschaften im Dokumentinformationsbereich und Auswahl des Punktes „Erweiterte Eigenschaften“ eingesehen und bearbeitet werden. In diesem Bereich können Metadaten unter der Registerkarte „Zusammenfassung“ und „Anpassen“ integriert werden.

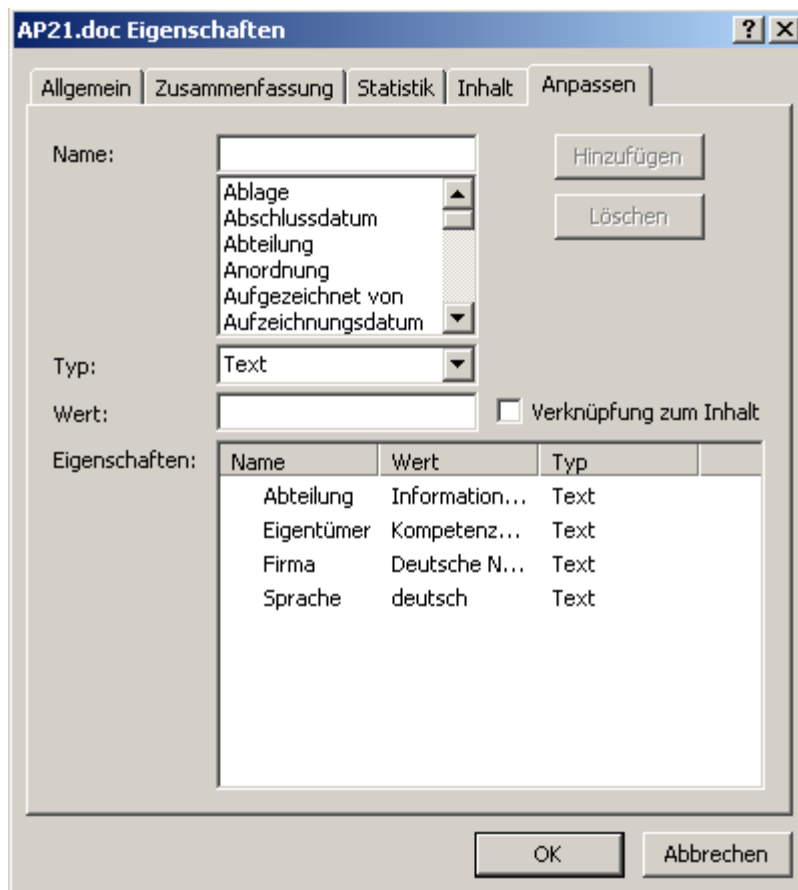


Abbildung 3: Erweiterte und fortgeschrittene benutzerdefinierte Eigenschaften in Microsoft Office

Durch Microsoft Office wird beim Anlegen von Werten für einzelne Metadatenelemente eine Validitätsprüfung mittels Auswahl vordefinierter Regeln durchgeführt. Beispielsweise kann eingestellt werden, dass eine Telefonnummer nur Zahlen in ihrem Wertebereich enthalten darf. Allerdings ist die Validierung eines definierten Wertebereiches die einzig mögliche Validitätsprüfung.

Die Integration von Kern- oder Standardeigenschaften sowie von erweiterten und fortgeschrittenen benutzerdefinierten Eigenschaften durch einen Produzenten kann relativ komfortabel vorgenommen werden. Die Integration von Metadaten kann mittels Microsoft Office durch einen Benutzer komfortabel und effizient vorgenommen werden. Dies lässt auf die gegebene Gebrauchstauglichkeit (Usability) des Autorenwerkzeuges rückschließen. Obgleich an dieser Stelle festgehalten werden muss, dass das Autorenwerkzeug keine integrierte Möglichkeit für den Import standardisierter Metadatenschemas zulässt. Außerdem ist als nachteilig festzuhalten, dass aufgezählte Inhalte (bspw. für Stichworte) nicht in wiederholten Feldern für Metadatenelemente, sondern alle als Wert innerhalb eines Feldes eines Metadatenelements im Dokument gespeichert werden.

Benutzerdefinierte Eigenschaften, zu denen u. a. die Anreicherung von unternehmensspezifischen Metadatenelementen in Dokumenten zählen, können im Verarbeitungsprozess ausschließlich durch externe Applikationen und nicht in Microsoft Office selbst integriert werden. Benutzerdefinierte Eigenschaften können nur mittels eines Eigenschaftenverwaltungsservers oder einem Microsoft Office InfoPath 2007 Formular zum Dokument hinzugefügt werden. Für die Implementierung entsprechen-

der Applikationen steht Entwicklern das Open XML SDK 2.0 for Microsoft Office⁹ zur Verfügung.

Eine Übernahme und Nachnutzung von Kern- oder Standardeigenschaften, die während einer Dokumentenproduktion in OpenOffice.org generiert wurden, ist ebenfalls möglich. Die einzige Ausnahme stellt die Übernahme der Autorenangabe dar, die durch Microsoft Office nicht erkannt wird. Die ursprüngliche Metadatenstruktur bei einer anschließenden Speicherung im Microsoft-Office-Dokumentenformat wird nicht erhalten. Stattdessen werden die Metadaten entsprechend der OOXML-Spezifikation im Dokument abgelegt. In OpenOffice.org erstellte, erweiterte und fortgeschrittene benutzerdefinierte Eigenschaften konnten vollständig in Microsoft Office übernommen werden und werden ebenfalls entsprechend den OOXML-Richtlinien in einem Microsoft-Dokument abgelegt.

3.2 OpenOffice.org-Dokumente

Als immer stärker genutztes Autorenwerkzeug wird nachfolgend OpenOffice.org in der Version 3.2.0 und sein Dokumentenformat OpenDocument Format v1.1 für die Eignung zur Metadatenintegration diskutiert.

3.2.1 OASIS OpenDocument Format

Das OASIS OpenDocument Format (ODF) ist ein offener und internationaler Standard für Office-Dateiformate. Das ursprüngliche Format wurde von der Firma Sun für eigene Office-Applikationen entwickelt. Die Spezifikation eines offenen XML-basierten Dateiformats wurde nachfolgend durch OASIS im Zeitraum 2002 bis 2005 umgesetzt. Auf Empfehlung der Europäischen Kommission wurde die Formatspezifikation 2006 in der internationalen Norm ISO/IEC 26300 veröffentlicht.

Wie auch Office Open XML bietet ODF aufgrund des XML-basierten Formats eine Vielzahl an Vorteilen bezüglich Nachnutzbarkeit, Austauschbarkeit und Erweiterbarkeit. ODF findet neben der Verwendung in OpenOffice.org und StarOffice eine stetig steigende Verbreitung in Produkten von Drittanbietern.

ODF ist ebenfalls ein Containerformat, welches sich an der Struktur des Java-Archive-File-Formats¹⁰ orientiert. Für die Komprimierung der im Container enthaltenen Dateien wird das ZIP-Kompressionsverfahren verwendet. Jede OpenDocument-Datei besitzt im Container eine unkomprimierte Manifest-Datei, in der alle im Container enthaltenen Dateien mit ihren MIME-Typen aufgeführt sind. Dateien, die Dokumentinhalte, Dokumentstile und Dokumenteigenschaften enthalten, werden im XML-Format serialisiert. In ODF-Dokumenten integrierte audiovisuelle Medien können im Binärformat hinterlegt werden. Prinzipiell sind alle Dateien außer dem Manifest im Container komprimiert.

In ODF 1.1 werden Metadaten noch für ein gesamtes Dokument und in einer eigens definierten XML-Struktur ausgewiesen. Die ODF 1.2 Metadata Specification¹¹ verdeutlicht bereits die Erneuerungen in ODF 1.2¹². Für die Serialisierung von Metadaten soll für eine bessere Kompatibilität zu den Semantic-Web-Entwicklungen einerseits auf das Resource Description Framework (RDF) zurückgegriffen werden. Andererseits wird es möglich sein, Metadaten nicht nur dokumentübergreifend, sondern

⁹ <http://www.microsoft.com/downloads/details.aspx?FamilyId=C6E744E5-36E9-45F5-8D8C-331DF206E0D0&displaylang=en>

¹⁰ <http://java.sun.com/j2se/1.5.0/docs/guide/jar/jar.html>

¹¹ <http://www.oasis-open.org/committees/download.php/26847/08-01-16-ODF-Metadata-Proposal-Committee-Draft.odt>

¹² http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office

auch für einzelne Inhaltsbausteine wie Paragraphen, Überschriften oder Bilder anzugeben.

In ODF werden grundlegend drei Arten von Metadaten unterschieden:

- Vordefinierte Metadaten (pre-defined metadata) besitzen eine definierte Semantik, die durch Dokumentenverarbeitungsprozesse einheitlich behandelt werden kann. Die eingesetzten Schemata setzen sich aus etablierten sowie durch OASIS neu spezifizierten Metadatenelementen zusammen.
- Benutzerdefinierte Metadaten (user-defined metadata), welche anhand einer Definition von Name-Type-Value-Tripeln ausgewiesen und in der Regel formularbasiert durch eine Applikation in das Dokument integriert werden. Benutzerdefinierte Metadaten können ausschließlich in Umgebungen kollaborativ verwendet werden, in denen man sich auf die Verwendung eines Triple-Sets verständigt hat.
- Kundenspezifische Metadaten (custom metadata) sind frei spezifizierbare Metadaten, die bei einer maschinellen Verarbeitung des ODF-Dokuments in den Metadaten integriert werden dürfen. Da kundenspezifische Metadaten keinen Standards unterliegen, sind sie aufgrund ihrer eingeschränkten Interoperabilität ausschließlich in abgestimmten Umgebungen einzusetzen.

Durch das OpenDocument Format werden folgende vordefinierte Metadatenelemente bereitgestellt. Tabelle 2 führt diese Elemente mit entsprechendem Ursprungsnamespace auf.

Element	Namespace
generator	Office Meta
title	Dublin Core
description	Dublin Core
subject	Dublin Core
keywords	Office Meta
initial-creator	Office Meta
creator	Dublin Core
Printed-by	Office Meta
creation-date	Office Meta
date	Dublin Core
print-date	Office Meta
template	Office Meta
auto-reload	Office Meta
hyperlink-behavior	Office Meta
language	Dublin Core
editing-cycles	Office Meta
editing-duration	Office Meta
document-statistics	Office Meta

Tabelle 2: Vordefinierte Metadaten in ODF

3.2.2 Metadatenanreicherung durch OpenOffice.org

Vordefinierte und benutzerdefinierte Metadaten können in OpenOffice.org unter dem Menüeintrag „Datei“ unter dem Menüpunkt „Eigenschaften“ angelegt und geändert werden. Die vordefinierten Metadaten finden sich unter der Registerkarte „Beschreibung“ (vgl. Abbildung 4). Benutzerdefinierte Metadaten können unter der Registerkarte „Benutzerdefinierte Eigenschaften“ (vgl. Abbildung 5) eingesehen und angepasst werden.

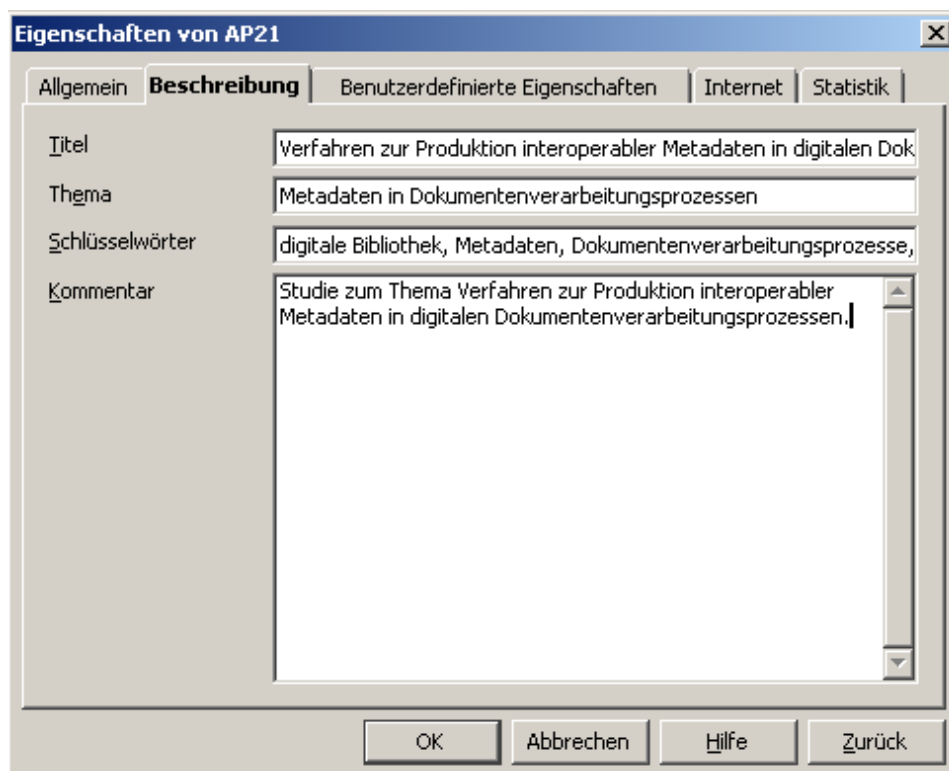


Abbildung 4: Vordefinierte Eigenschaften in OpenOffice.org

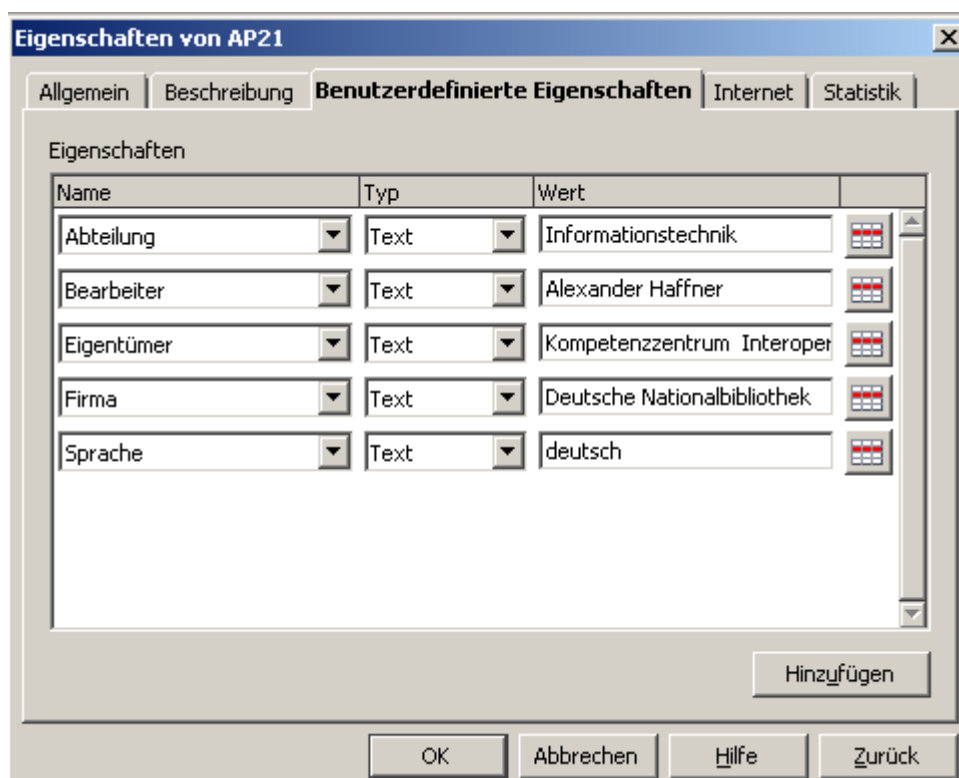


Abbildung 5: Benutzerdefinierte Eigenschaften in OpenOffice.org

In OpenOffice.org werden vergleichsweise wenig vordefinierte Metadaten durch den Anwender manuell als Dokumenteigenschaften erfasst. Es ist lediglich möglich, den Titel, das Thema, Schlüsselwörter und einen Kommentar anzugeben. Positiv anzumerken ist die Aufspaltung der durch Kommata abgetrennten Schlüsselwörter in der Eingabemaske in separate, aufeinanderfolgende Metadatenelemente bei der Spei-

cherung im Dokument. Alle weiteren Metadatenelemente der vordefinierten Eigenschaften werden durch das Autorenwerkzeug eigenständig belegt. Hierzu gehört auch die Autorenangabe, die aus den Anwendungseinstellungen übernommen wird.

Benutzerdefinierte Eigenschaften werden bei ihrer Angabe einer Validitätsprüfung hinsichtlich des ausgewählten (vordefinierten) Typs unterzogen. Wie in Microsoft Office ist auch in OpenOffice.org die Validierung des definierten Wertebereiches die einzig mögliche Validitätsprüfung.

Die Funktionalitäten zur formularbasierten Angabe von vordefinierten und benutzerdefinierten Metadaten weisen hinsichtlich der Gebrauchstauglichkeit keine gravierenden Mängel auf. Der Benutzer ist in der Lage, eine Reihe von Aussagen komfortabel zu treffen und wird durch die automatischen Metadatenanreicherungen unterstützt.

Kundenspezifische Metadaten können nicht durch OpenOffice.org in das Dokument integriert bzw. angepasst werden. Das Autorenwerkzeug stellt lediglich deren Aufbewahrung unabhängig von Inhaltsänderungen und Manipulationen der vordefinierten und benutzerdefinierten Metadaten sicher. Eine Anreicherung kundenspezifischer Metadaten kann ausschließlich durch die Verwendung zusätzlicher Applikationen oder OpenOffice.org Plug-ins vorgenommen werden. Typischerweise wird die Anreicherung serverseitig im Dokumentenverarbeitungsprozess vollzogen. Für die Erstellung entsprechender Werkzeuge steht Entwicklern die OpenOffice.org¹³ API zur Verfügung.

Eine Metadatenverwendung und -übernahme von in Microsoft Office erzeugten Metadaten ist in OpenOffice.org sehr leicht möglich. Es werden alle in der Schnittmenge der beiden Office-Applikationen befindlichen, vordefinierten Metadaten erkannt. Ebenso können sämtliche benutzerdefinierten Metadaten durch OpenOffice.org interpretiert werden. Beim Abspeichern des Dokuments in ODF werden die enthaltenen vordefinierten und benutzerdefinierten Metadaten entsprechend den Kriterien der ODF-Spezifikation abgelegt. Beispielsweise werden die in einem Metadatenelement gehaltenen - und durch Kommata getrennten - Schlüsselwörter aufgesplittet und in aufeinanderfolgenden Metadatenelementen abgelegt.

3.3 PDF-Dokumente

Abschließend werden Adobe Acrobat 8 Professional und das Portable Document Format (PDF) hinsichtlich einer Metadatenanreicherung während des Produktionsprozesses betrachtet.

3.3.1 Portable Document Format

Das durch Adobe entwickelte Portable Document Format (PDF) war in der Vergangenheit lediglich als proprietäres Format verfügbar. Adobe verfolgte mit PDF das Ziel, Dokumente plattformunabhängig zu erstellen und layoutgetreu in heterogenen Systemen nutzen und verbreiten zu können. Anfang 2007 wurde die PDF Version 1.7 als komplett offener Standard in der ISO 32000-1:2008 verabschiedet.

Grundsätzlich ist PDF eine vektorbasierte Seitenbeschreibungssprache, die die detailgetreue Wiedergabe eines Layouts sichern soll. Durch den Einsatz von Tagged PDF ist es allerdings auch möglich, Inhalte – ähnlich wie in HTML – durch Tags auszuzeichnen. Durch diese Auszeichnung kann PDF ebenso in einem angepassten Layout, beispielsweise auf portablen Endgeräten, personalisiert ausgegeben werden. Des Weiteren ermöglicht die Auszeichnung eine Inhaltsverarbeitung durch assistive

¹³ <http://api.openoffice.org/>

Technologien wie Screenreader. Auch wurde in PDF-Dokumente eine Reihe von Navigationshilfen wie Lesezeichen integriert, die es den Benutzern ermöglichen sollen, das digitale Dokument im jeweiligen Präsentationswerkzeug komfortabel zu verwenden.

Im Gegensatz zu Office-Dokumenten eignet sich PDF nur bedingt für eine Weiterverarbeitung. PDF ist für die Publikation von digitalen Textdokumenten und Grafiken sowie für den Druck vorgesehen.

Extensible Metadata Plattform (XMP) stellt ein durch Adobe publiziertes Standardformat dar, das für die Erstellung, Verarbeitung und den Austausch von Metadaten vorgesehen ist. Der Standard wurde entwickelt, um für verschiedene Applikationen einen gemeinsamen Zugriffspunkt auf den zu einer Datei zugehörigen Metadatenatz anzubieten. XMP kann neben dem PDF-Format auch in TIFF, JPEG, JPEG 2000, GIF, PNG, HTML, AI (Adobe Illustrator), SVG/XML, PSD (Adobe Photoshop), PostScript bzw. EPS und DNG eingebunden werden.

“For PDF files, the XMP Packet is embedded in a metadata stream contained in a PDF object (beginning with PDF 1.4).”¹⁴

Die Vorschriften zur Integration von XMP in PDF-Dokumenten sind im PDF Reference Document zu finden.¹⁵

XMP sieht ein Data Model und ein Storage Model für die Metadatenverarbeitung vor. Das zugrundeliegende Datenmodell ist ähnlich dem von Dublin Core. Es wird zur Beschreibung der Ressource (lediglich eine Datei, keine Aggregationen von Dateien) unter Zuhilfenahme von Property-Value-Paaren verwendet, die über Schemas in Beziehung stehen. Das Storage Model bestimmt, wie die zum Data Model konformen Daten physisch gehalten werden. Die XMP-Metadaten werden syntaktisch in einem XML-Stream serialisiert. Als Notation wird Resource Description Framework (RDF) verwendet. XMP umfasst nur ein Subset von RDF, für Detailinformationen sei die Leserschaft auf die XMP-Spezifikation verwiesen. Beispielsweise wird das about-Attribut nicht für die Identifikation der Ressource verwendet. Die serialisierten Daten werden mit Hilfe von Packet Wrappern eingehüllt, um im zugehörigen File integriert zu werden. Ein Wrapper kann prinzipiell mehrere Packets enthalten, wovon jedoch nur eines den aktuellen Stand des Files ausweist. Der XMP Packet Wrapper ermöglicht die Nutzung der XMP-Metadaten durch Software, die das Format der Datei selbst nicht versteht. Durch den Einsatz von Markern ist die Software in der Lage, den Metadatenatz zum Beispiel in Bytecode zu erkennen und ihn somit weiterzuverarbeiten oder anzupassen.

In der XMP-Spezifikation werden eine Reihe von Kernelementen (Core Schemas) u. a. zum Einsatz in PDF vorgegeben. Schemas sind vordefinierte Sets von Metadata-Property-Definitionen, die in XMP standardmäßig zugelassen sind und einen Großteil notwendiger Semantik abdecken. Diese Schemas sind erweiterbar, sollten jedoch aus Interoperabilitätsgründen in der empfohlenen Form verwendet werden.

Spezifizierte Schemas sind:

- Dublin-Core-Schema (DCMI Element Set)
- XMP-Basic-Schema: Basisinformationen (u. a. URL, Datumsangaben der Erstellung und Änderung des Dokumentes sowie des Metadatenatzes, Identifier, Autorenwerkzeug)

¹⁴ http://www.adobe.com/devnet/xmp/pdfs/xmp_specification.pdf

¹⁵ http://www.adobe.com/devnet/acrobat/pdfs/pdf_reference_1-7.pdf

- XMP-Rights-Management-Schema: Rechtemanagement (u. a. Rechte-Existenz, Online-Zertifikat, Besitzer, Nutzungsbedingungen)
- XMP-Media-Management-Schema: für Digital-Asset-Management-Systeme, kurz DAM (u. a. Referenz zum Original, Historie, Version, Angaben zum Managementsystem und dem Management der Ressource)
- XMP-Basic-Job-Ticket-Schema: Workflowinformationen (Verweis zu Jobangaben)
- XMP-Paged-Text-Schema: Erscheinungsbild (von Text) im Dokument (u. a. Seitengröße und -anzahl, verwendete Schriftarten und Farben, Angaben für Platzhalter)
- XMP-Dynamic-Media-Schema: spezifiziert Properties, die von der Adobe Dynamic Media Group genutzt werden (u. a. Videoframerate und -größe, Pixel- und Farbinformationen, Audioinformationen, Kompression, Zeitangaben – z. B. zur Aufnahme sowie Modifikation, Genre, Copyright)
- Adobe-PDF-Schema: speziell in PDFs verwendete Properties (Stichworte, PDF-Version, Produzent)
- Photoshop-Schema: Metadaten für in Photoshop erzeugte Grafiken
- Camera-Raw-Schema: spezifiziert Properties für Bilder im Camera Raw Mode (basiert auf DNG - Digital Negative Format)
- EXIF-Schemas: Metadatenstandard für die Beschreibung von Bilddateien; wird in vielen Digitalkameras eingesetzt

Die aufgeführten Schemas sind Kernschemas, für Metadatenelemente sollte immer zuerst versucht werden, entsprechende Einträge in diesen Schemas zu finden. Falls dies nicht der Fall sein sollte und die Anwendung auf zusätzliche Informationen angewiesen ist, kann man eigene Custom-Schemas definieren. Beim Einsatz von Custom-Schemas empfiehlt es sich, auf standardisierte Metadatenschemas für eine Integration in XMP zurückzugreifen. Auch für das Autorenwerkzeug Adobe Acrobat 8 Professional sowie andere Werkzeuge zur PDF-Generierung müssen für eine entsprechende Metadateninteroperabilität im gesamten Dokumentenverarbeitungsprozess Absprachen mit allen betroffenen Akteuren getroffen werden.

3.3.2 Metadatenanreicherung durch Adobe Acrobat

In Adobe Acrobat 8 Professional können vordefinierte Metadaten im Datei-Menü unter dem Menüpunkt „Eigenschaften“ eingesehen und teilweise manuell angepasst werden. Unter der Registerkarte „Beschreibung“ (vgl. Abbildung 6) besteht die Möglichkeit, grundlegende Dokumenteigenschaften anzupassen.

Zum Einsehen und Anpassen weiterführender Dokumenteigenschaften muss der Menüpunkt „Zusätzliche Metadaten“ ausgewählt werden. Im darauffolgenden Dialogschritt (vgl. Abbildung 7) ist es möglich, ergänzend beschreibende Angaben für das Dokument vorzunehmen. Unter dem Menüpunkt „Erweitert“ (vgl. Abbildung 8) können alle in XMP-Stream enthaltenen Metadaten eingesehen werden. Darüber hinaus kann eine XMP-Datei aus der PDF-Datei extrahiert und lokal für eine manuelle Anpassung abgespeichert werden. Nach einer Anpassung können die Datei und der darin enthaltene Metadatenatz wieder in das PDF-Dokument integriert werden und den bislang enthaltenen Inhalt ersetzen. Auf diesem Weg ist es beispielsweise möglich, Metadatenelemente aus der FOAF-¹⁶ oder der RDA-Ontologie¹⁷ einzubinden. Die Ontologien sind in diesem Kontext als Custom-Schemas zu sehen. Die Anpas-

¹⁶ <http://xmlns.com/foaf/spec/>

¹⁷ <http://metadataregistry.org/rdbrowse.htm>

sung der XMP-Datei kann mit einem beliebigen XML-Editor vorgenommen werden. Von Adobe wird auch ein XMP Toolkit¹⁸ für die Bearbeitung zur Verfügung gestellt. An dieser Stelle soll darauf hingewiesen werden, dass Autoren in der Regel auf diesem Wege keine Integration von Custom Metadata vornehmen. Die Einbindung solcher Metadaten sollte durch Repositorien serverseitig unterstützt werden.

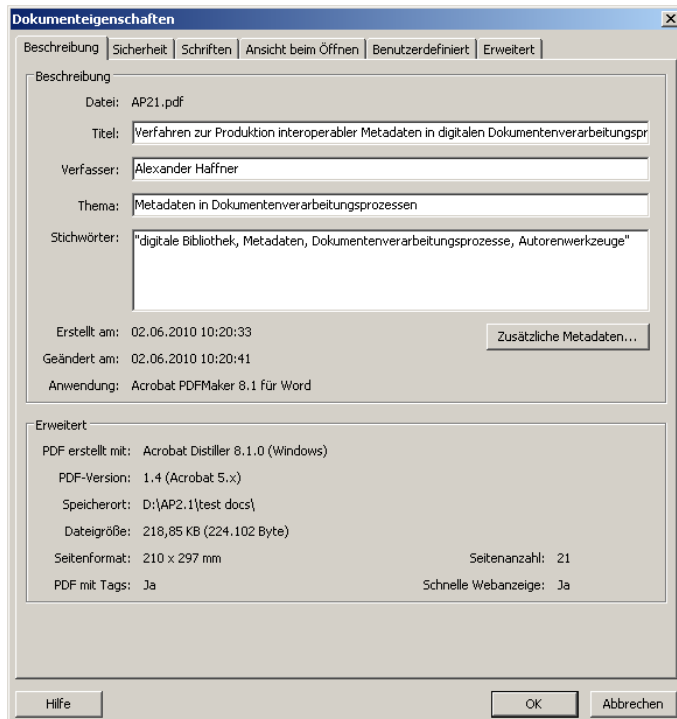


Abbildung 6: Dokumenteigenschaften in Adobe Acrobat 8 Professional

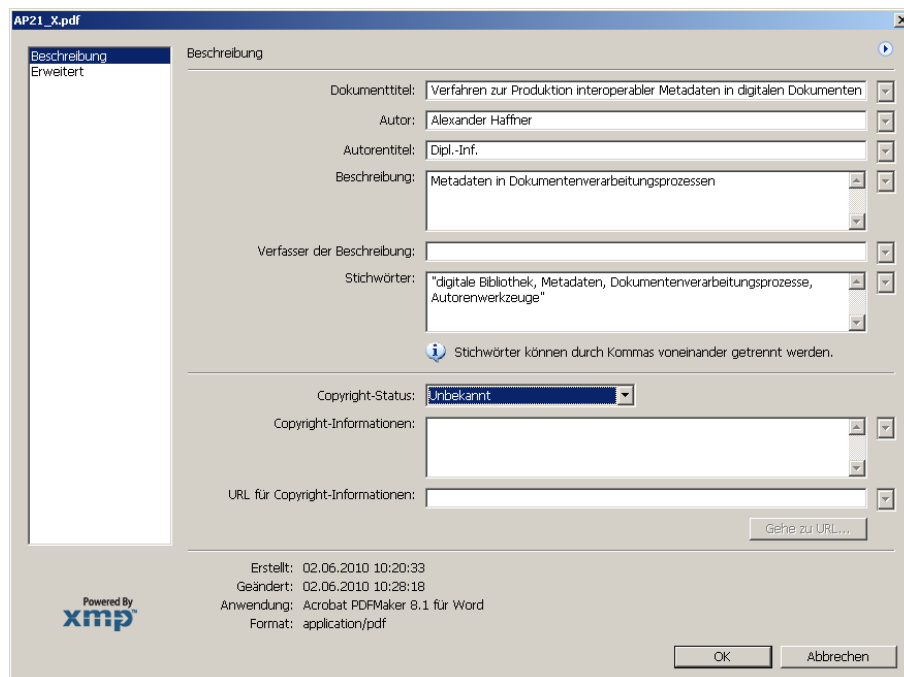


Abbildung 7: Erweiterte Dokumentenbeschreibung in Adobe Acrobat 8 Professional

¹⁸ <http://www.adobe.com/devnet/xmp/>

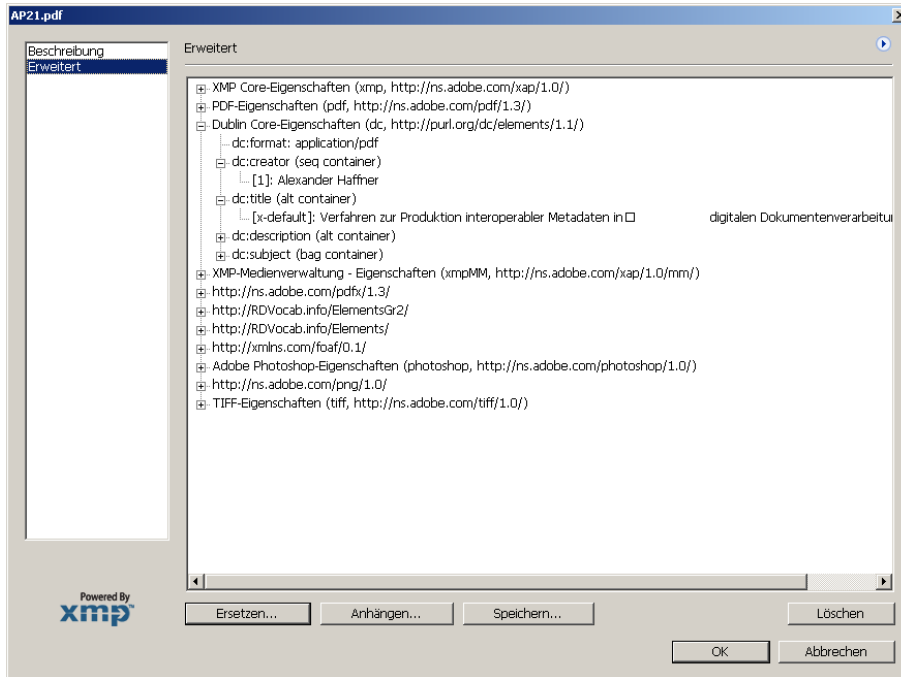


Abbildung 8: Metadatschema in Adobe Acrobat 8 Professional

Im Dialogfeld „Dokumenteigenschaften“ unter der Registerkarte „Benutzerdefiniert“ (vgl. Abbildung 9) können Anwender benutzerdefinierte Dokumenteigenschaften anpassen. Der angegebene Wert eines Metadatenelements muss valide zum vordefinierten Typ sein, um im PDF-Dokument durch das Autorenwerkzeug abgespeichert zu werden.

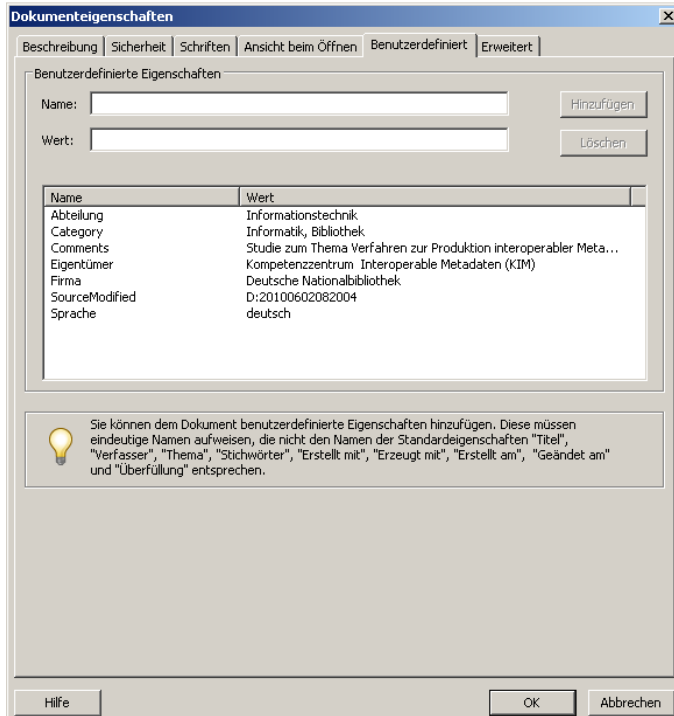


Abbildung 9: Benutzerdefinierte Dokumenteigenschaften in Adobe Acrobat 8 Professional

Auf Komponentenebene erlaubt es Adobe Acrobat zusätzlich, durch das TouchUp-Werkzeug, Metadaten für einzelne Komponenten des Dokumentes anzulegen: beispielsweise lassen sich in Container-Elementen enthaltene Textabschnitte mit einer Beschreibung versehen, die Sprache kann festgelegt oder Informationen zur Schrift hinterlegt werden.

Hinsichtlich der Gebrauchstauglichkeit für die dialogbasierte Integration von Metadaten im PDF-Dokument zeigt sich Adobe Acrobat als ein sehr komfortables Werkzeug. Der Benutzer wird durch automatische Metadatenanreicherungen unterstützt und bei der Angabe benutzerdefinierter Dokumenteigenschaften mittels Validierungsmechanismen auf Fehler hingewiesen. Ausschließlich die Validierung extern angepasster XMP-Dateien (außerhalb von Adobe Acrobat) wies Mängel beim Import dieser Dateien im PDF-Dokument auf. Der Nutzer kann durch die Anwendung nicht auf die tatsächlichen Fehler im XMP-File aufmerksam gemacht werden. Durch die Preflight-Funktion, die im Druckwesen die softwareseitige Interpretation und Prüfung von Dokumenten vornimmt, können anhand vordefinierter Profiles PDF-Dokumente weiterführend validiert werden. Diese Validierungsfunktionen erlauben jedoch auch in Adobe Acrobat 8 Professional keine Metadatenvalidierung.

3.3.3 PDF-Generierung und Metadatenübernahme aus Dritt-Applikationen

Microsoft Office wie auch OpenOffice.org erlauben ein direktes Speichern von Dokumenten in PDF. Durch die Offenlegung des Formates ist ein Export der Inhalte und Strukturen ebenso wie die Wiederverwendung der Metadaten in PDF möglich. Die Generierung des PDF-Dokuments setzt dabei auf dem XML-Format der jeweiligen Office-Applikation auf. Für die Erstellung des PDF-Dokumentes aus den XML-Daten sind zwei Schritte notwendig: die Transformation mit Hilfe eines XSLT-Stylesheets in das XSL-FO Format und das anschließende Erzeugen eines PDF-Dokumentes durch einen PDF-Prozessor.

Beim Abspeichern von OOXML-Dokumenten in PDF integriert Microsoft Office 2007 einen Großteil der Kern- oder Standardeigenschaften im XMP-Stream. Mittels Dublin-Core-Metadatenelementen werden der Titel, der Autor sowie das Thema hinterlegt. Einzelne Schlüsselwörter werden in PDF-Metadaten dargestellt, aber nicht in separaten Metadatenelementen aufgeführt. In XMP-Metadaten wird der Zeitpunkt der PDF-Dokumentenerstellung und dessen letzte Änderung sowie Microsoft Office als Autorenwerkzeug ausgewiesen. Des Weiteren werden in PDF spezifische Metadaten wie beispielsweise eine Dokumenten-ID eingebunden. Positiv anzumerken ist, dass durch das Autorenwerkzeug beim Speichern eines PDFs rudimentäre Tags (Struktur-Metadaten) erzeugt werden. Tags bringen neben einer erhöhten Barrierefreiheit des Dokumentes insbesondere Vorteile für zukünftige Formatmigrationen mit sich.

Bei der Verwendung des Adobe Acrobat Plug-Ins in Microsoft Office werden dem Anwender diverse Optionen unter dem Menüeintrag „Grundeinstellungen“ für die Transformation angeboten. So ist es beispielsweise zusätzlich möglich, Lesezeichen, Fußnoten, Kommentare und Verknüpfungen in das Zielformat zu übernehmen. Ebenso ist es möglich das PDF-Dokument den PDF/A-Richtlinien entsprechend und somit als Tagged-PDF abzuspeichern. Durch die Aktivierung des Auswahlpunktes „erweiterte Tag-Einstellungen aktivieren“ können qualitativ hochwertigere Tags im PDF-Dokument hinterlegt werden. Im Kontext dieser Untersuchung ist jedoch vor allem relevant, dass durch die Verwendung des Plug-Ins zusätzlich PDF/X-Metadaten in den XMP-Stream eingebunden werden. Hierzu gehören unter anderem „Firma“, „Manager“, „Kommentar“, „Kategorie“ und „Änderungszeitpunkt“ des Quelldokumentes (Word-Dokument).

Bei der Erzeugung von PDF-Dokumenten in Microsoft Office werden keine (erweiterten und fortgeschrittenen) benutzerdefinierten Eigenschaften übernommen.

Eine zum MS-Office-PDF-Plug-In vergleichbare Generierungsfunktion ist in OpenOffice.org schon länger standardmäßig implementiert. PDF-Dokumente können in OpenOffice.org unter dem Menüpunkt „Datei“ und unter Auswahl der Funktion

„Exportieren als PDF“ generiert werden. Durch Add-ons ist es auch möglich in OpenOffice.org (ab Version 3) PDF-Dokumente zu importieren und nachträglich zu bearbeiten. Dem Anwender stehen beim Export von PDF-Dokumenten eine Reihe von Transformationsoptionen zur Auswahl. Unter anderem kann ein Tagged-PDF nach dem PDF/A-Standard generiert sowie Lesezeichen, Dokumentreferenzen und Notizen exportiert werden. Bei sauberer Verwendung von Formatvorlagen im ODF-Dokument ist es möglich, eine sehr gute Tag-Qualität mit hierarchischer Auszeichnung zu erreichen.

Bezüglich einer Metadatenübernahme aus dem ODF-Dokument in den XMP-Stream wird allerdings nur eine bedingt zufriedenstellende Qualität erreicht. Beim Export werden ausschließlich ausgewählte Metadatenelemente der vordefinierten Metadaten überführt. Hierzu gehören „Titel“, „Thema“, „Schlüsselwörter“ und die Angabe von OpenOffice.org als Autorenwerkzeug. Alle Schlüsselwörter werden im XMP-Stream in einem einzigen Datenelement ausgewiesen.

Die Übernahme von benutzerdefinierten und kundenspezifischen Metadaten wird durch die in OpenOffice.org integrierte PDF-Exportfunktion nicht unterstützt.

4 Metadaten im Dokumentenverarbeitungsprozess

Dieses Kapitel diskutiert den Umgang mit in Autorenwerkzeugen generierten Metadaten in Dokumentenverarbeitungsprozessen. Hierfür wird zunächst die Übernahme der Metadaten in einen serverseitigen Verarbeitungsprozess und anschließend die praktische Relevanz weiterer Metadatenanreicherungen in einem Dokument unter verschiedenen Gesichtspunkten diskutiert.

4.1 Metadatenübernahme aus einer digitalen Ressource

Entsprechend dem OAIS Reference Model wird während des Ingest-Prozesses durch das Archivierungssystem ein Submission Information Package (SIP) vom Produzenten entgegen genommen. Das Archivierungssystem ist dabei in der Verantwortung, die Qualität des SIP's zu sichern. Eine Sicherung des SIP's umfasst dessen Ergänzung um deskriptive Metadaten und Preservation Description Information. Anschließend übergibt der Ingest-Prozess das SIP zum Archival Storage für eine Umwandlung des SIP in ein AIP sowie zur finalen Einpflege im Archivierungssystem. Für eine optimale Auffindung der Ressource werden die deskriptiven Informationen zusätzlich in einer Data-Management-Komponente registriert.

Die Administration des Archivierungssystems ist für die Aushandlung der Submission-Richtlinien mit den Produzenten verantwortlich. Die Submission-Richtlinien umfassen unter anderem die anzuwendenden Standards, wie beispielsweise akzeptierte Dateiformate einer Ressource, anzugebende deskriptive Metadaten oder die zu verwendenden Metadatenschemas. Bei der Einreichung einer Ressource und dem dadurch entstehenden SIP ist die Administration ebenfalls für die Überprüfung der Standardeinhaltung seitens des Produzenten/Einreichenden zuständig.

Die Erzeugung eines SIP kann auf unterschiedliche Weise ablaufen. Eine weitverbreitete Möglichkeit ist der kombinierte Upload der eigentlichen Publikation mit der manuellen Angabe ausgewählter deskriptiver Metadaten sowie Preservation Description Information. Eine manuelle Angabe solcher Informationen wird in der Regel formularbasiert durch den Ingest-Prozess unterstützt. Die angegebenen Metadaten werden nach Abschluss der Einpflege durch das Archivierungssystem in definierten

Metadatenformaten im SIP integriert. Die Metadatenintegration findet dabei nicht in der Publikation selbst statt. Ein ähnliches Vorgehen liegt bei der Abgabe von Publikationen mittels OAI-PMH¹⁹ vor. Metadaten werden in durch die Administration zuvor definierten Strukturen durch das Archivierungssystem (Rolle des Service Providers) vom Bereitsteller (Rolle des Data Providers) „eingesammelt“. Die zu publizierende Ressource ist nicht in den übersendeten Daten enthalten, sondern wird in der Regel aus diesen heraus referenziert. Durch die Referenzierung ist ein automatischer Bezug der Ressource für das Archivierungssystem möglich. Das SIP erhält folglich den gleichen Aufbau wie bei einer formularbasierten Ressourcenabgabe.

An dieser Stelle soll vorweggenommen werden, dass der Verzicht auf eine Integration der zusätzlich abgelieferten Metadaten in der Ressource selbst aus dem Bestreben der Archivbetreiber nach einer Trennung von Inhalt und Metadaten resultiert.

In den vorangegangenen Kapiteln wurden Verfahren für die Metadatenanreicherung in digitalen Ressourcen diskutiert. Nachfolgend sollen die dadurch entstehenden Vor- aber auch Nachteile bei der Einpflege in Archivierungssysteme betrachtet werden.

In der Vergangenheit war der Produzent einer Publikation meistens nicht am bibliothekarischen Katalogisierungsprozess beteiligt. Mit der zunehmenden Archivierung von elektronischen Ressourcen (Netzpublikationen) wird die Bereitstellung deskriptiver Metadaten vermehrt vom Verlag zum Autor verlagert. Ein populäres Beispiel stellt die Veröffentlichung von Hochschulschriften und wissenschaftlichen Publikationen dar. Der Produzent ist dazu verpflichtet bei der Abgabe (Upload) seiner Ressource die relevanten deskriptiven Metadaten für das Archivierungssystem anzugeben. Der Vorteil dabei liegt darin, dass der Verfasser der Publikation das umfassendste Wissen über den Inhalt der Publikation besitzt. Andererseits ist er nicht unbedingt damit vertraut, dieses Wissen in die Datenstrukturen einer Bibliothek zu integrieren. Formulargestützte Ansätze bei der Abgabe von Publikationen können den Anwender dabei unterstützen, die für die Katalogisierung notwendigen Daten unabhängig vom Datenformat akkurat in das Archivierungssystem einzupflegen. Durch Auswahllisten ist es möglich, der Publikation kontrollierte Werte aus standardisierten Klassifikationssystemen zuzuweisen. Des Weiteren besteht die Möglichkeit, z. B. eine Person mittels Normdaten eindeutig zu identifizieren und sie als Autor auszuwählen. Der bislang angewendete Ansatz weist ein hohes Potenzial bezüglich der Vernetzung von Wissen auf und fördert folglich die Auffindbarkeit von Ressourcen anhand der erhobenen Metadaten erheblich.

Die Einbindung von Metadaten in eine digitale Ressource hat hingegen den Vorteil, dass der Ingest-Prozess auf die in der Publikation integrierten Daten automatisiert zugreifen kann. Wenn Metadaten in einem standardisierten und offen zugänglichen Metadatenformat angeboten werden, sind diese im Dokumentenverarbeitungsprozess einfach auszulesen und nachnutzbar. Insbesondere XML-basierte Metadatenformate eignen sich für diese Art der Datenübernahme. Allerdings ist es für den Ressourcenproduzenten beim Anlegen der Metadaten in der Ressource nicht möglich, sich direkt mit dem bereits in Bibliotheken existierenden Wissensnetz (Normdaten wie in Schlagwortdateien, Klassifikationssystemen etc. enthalten) zu verlinken. Die Konsequenz daraus ist, dass die Daten zwar automatisch vom Ingest-Prozess validiert und angenommen werden, aber der Nutzer dazu gezwungen sein muss, diese Daten mit dem Datennetz der Bibliothek abzugleichen und zu verlinken. Angaben wie der Autorenname können vom System dazu verwendet werden, in Normdateien nach Personen mit diesem Namen zu suchen, um die Publikation gegebenenfalls nach

¹⁹ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Bestätigung des Einpflegenden mit einem der Personeneinträge zu verknüpfen. Ein ähnliches Bild zeigt sich bei der Angabe von Schlüsselwörtern. Für den Fall dass Schlüsselwörter, die im Metadatensatz der Publikation enthalten sind, im Klassifikationssystem noch nicht existieren, sollte der Anwender die Möglichkeit haben, diese Begriffe ähnlichen Begriffen oder einem Oberbegriff zuzuordnen. Derzeit entspricht dies jedoch nicht der typischen Praxis von Bibliotheken, da der Anwender als nicht autorisiert für derartige Aktivitäten eingestuft wird.

Resümierend ist festzustellen, dass die Übernahme von Metadaten aus einer digitalen Ressource in ein bibliothekarisches Archivierungssystem nur in Kombination mit einer manuellen Nachbearbeitung vollzogen werden sollte.

Ein weniger kritischer Nutzungskontext zeigt sich für die Übernahme von Metadaten in geschlossenen, nicht-öffentlichen Systemen wie beispielsweise Firmennetzwerken. Durch den ebenfalls abgeschlossenen Nutzerkreis sind Zuordnungen der Dokumente zu Personen ohne ein manuelles Eingreifen möglich. Unter Verwendung von Regeln zur Angabe zusätzlicher deskriptiver Informationen, wie Themen oder Schlüsselwörtern kann unter Umständen auch der Einsatz eines kontrollierten Vokabulars erreicht werden.

Nichtsdestotrotz empfiehlt es sich, auch für abgeschlossene Repositorien Formulare für die Dokumenteneinpflege anzubieten. Durch unterstützende Dialogsysteme wird gewährleistet, dass kontrollierte Vokabularien eingesetzt werden und außerdem zusätzliche kundenspezifische Metadaten in die Dokumente integriert werden. Die Integration von kundenspezifischen Metadaten kann in abgeschlossenen Systemen, wie bereits verdeutlicht, Vorteile für die Workflowoptimierung der verwaltenden Organisation mit sich bringen. Eine Konsistenzprüfung und Validierung bezüglich der Einhaltung organisationsspezifischer Vorgaben für die Angabe von Dokumenteigenschaften ist durch den Einpflegeprozess des eingesetzten Dokumentenservers bereitzustellen.

4.2 Metadatenanreicherung im Archivierungsprozess

Die in diesem Kapitel thematisierte Metadatenanreicherung im Archivierungssystem betrachtet ausschließlich die Übernahme von während der Archivierung neu gewonnenen Metadaten in die Ressource selbst. Der generell stattfindende Prozess des Metadatenmanagements im Archivierungssystem wird nicht berücksichtigt.

Wie bereits in Kapitel 4.1 soll für die bessere Unterscheidung der Anforderungen ein öffentlich zugängliches Repository (z. B. das einer Bibliothek) einem abgeschlossenen System (z. B. ein Firmennetzwerk) gegenübergestellt werden.

Große, öffentliche Archivierungssysteme, insbesondere die von Bibliotheken, zielen auf eine strikte Trennung von Inhalt und Metadaten ab. Grund dafür ist der extrem hohe Aufwand für die kontinuierliche Synchronisation von Metadaten und Inhalt (dem eigentlichen Dokument). Bei einer Anpassung von deskriptiven sowie administrativen Metadaten bzw. bei Migrationsverfahren von Dokumenten würde dies einen Mehraufwand bedeuten, der für den Archivierungsprozess in einer Bibliothek keinen entsprechenden Mehrwert hätte. Ein produziertes Dokument wird als „Transferdokument“ angesehen. Das bedeutet, sobald eine Ressource im Dokumentenverarbeitungsprozess des Archivierungssystems angekommen ist, stellen die deskriptiven Informationen des Archivierungssystems die primären Metadaten dar. Die Metadaten im eingereichten Dokument werden folglich nach der abgeschlossenen Ressourceneinpflege im Dokumentenverarbeitungsprozess gänzlich vernachlässigt.

Der sich daraus ergebende Nachteil zeigt sich bei der Betrachtung des gegenwärtigen Vorgehens von Suchmaschinen. Wenn Metadaten in einer Publikation enthalten

sind, können diese von Suchmaschinenbetreibern direkt ausgewertet und somit in die Suchergebnispräsentation einbezogen werden. Die deskriptiven Metadaten eines Archivierungssystems sind oft durch ihre HTML-Repräsentation nur bedingt der eigentlichen Ressource zuzuordnen. Ein erkennbares Indiz für die Beseitigung dieses Defizits sind die Bestrebungen, semantische Auszeichnungen in Form von Linked Data zu etablieren. Standardisierte, öffentlich zugängliche und maschinelle Metadaten können durch Linked Data eindeutig als bibliografische Daten einer bestimmten Ressource definiert werden. Dies wird Suchmaschinen zukünftig einen effizienten Umgang mit den durch Bibliotheken bereitgestellten Metadaten erlauben.

In einem abgeschlossenen System ist die Nutzung der Metadaten durch Dritte in der Regel nicht von Relevanz. Abhängigkeiten bestehen lediglich zum definierten internen Workflow. Um die Frage zu beantworten, ob eine kontinuierliche Anreicherung der Metadaten in den Ressourcen einen Mehrwert bringt, muss man der Frage nach dem eingesetzten Metadatenmanagementsystem nachgehen. Falls das Metadatenmanagement in einer eigenständigen Datenbank organisiert ist und Metadaten digitalen Ressourcen lediglich virtuell zuweist, kann auf eine erneute Anreicherung in Dokumenten verzichtet werden. Dies ist insbesondere legitim, da die Ressource im abgeschlossenen Netzwerk nur über das Metadatenmanagementsystem aufgefunden werden kann. Für den Fall, dass die Dokumentenorganisation im Repositorien über die in den Dokumenten existierenden Metadaten vorgenommen wird, ist es hingegen verpflichtend die Metadatenanreicherung stetig in den Ressourcen fortzuführen.

Resümierend kann festgehalten werden, dass die Entscheidung über eine Rückführung von während der Archivierung gewonnener Metadaten in die Ressourcen eine Fallentscheidung ist. Es gilt, eine ausführliche Abwägung der für die Organisation entstehenden Vor- und Nachteile vorzunehmen, um diese Entscheidung fällen zu können.

5 Definition von Gütekriterien

Basierend auf den vorgenommenen Betrachtungen werden in diesem Kapitel Qualitätskriterien für Dokumentenformate und die zu integrierenden Metadaten hinsichtlich ihrer Interoperabilität definiert. Des Weiteren werden Qualitätskriterien und Funktionen von Autorenwerkzeugen aufgeführt, die für die Generierung interoperabler Metadaten während der Dokumentenproduktion von Relevanz sind.

Die nachfolgenden Ausführungen wurden für die effiziente Evaluation von Dokumentenformaten und Autorenwerkzeugen bezüglich ihrer Eignung zur Integration interoperabler Metadaten in einer Checkliste (s. Anhang A) zusammengefasst. Es kann keine allgemeingültige Formel zur Auswertung der in der Checkliste aufgeführten Kriterien angeboten werden, da die Wichtung der Kriterien stark vom Einsatzszenario abhängig ist.

5.1 Gütekriterien für interoperable Metadaten in Dokumentenformaten

Die primäre Anforderung an das Dokumentenformat selbst ist die Gewährung der Zugänglichkeit zum serialisierten Metadatenatz. Hierfür ist die Verwendung offener und etablierter Dokumentenformatstandards unverzichtbar. Für die Zugänglichkeit zum Format und die damit verbundene Extraktionsfähigkeit und Nachnutzung des enthaltenen Metadatenstreams hat sich XML als Auszeichnungssprache bewährt. Dabei ist es nicht zwingend notwendig, wie am Beispiel von PDF ersichtlich, dass das gesamte Dokument in XML serialisiert ist. Es ist bereits ausreichend, wenn der

Metadatenstream im Bytecode in XML vorliegt. Ein unter Verwendung von XML ausgezeichneter Metadatenstream erlaubt es durch seine offene Zugänglichkeit unterschiedlichsten Produkten von verschiedensten Anbietern die Metadaten nachzunutzen und anzupassen. Allerdings soll darauf hingewiesen werden, dass Dokumentenformate, die vollständig in XML serialisiert sind, zusätzliche Vorteile mit sich bringen. Insbesondere ist dabei die Autorenwerkzeug-unabhängige Anpassung des Dokumenteninhaltes zu erwähnen, genauso wie die Möglichkeiten für eine Migration in aktuelle und zukünftig aufkommende Dokumentenformate.

Ein sich deutlich abzeichnender Trend ist die Ausweisung von semantischen Informationen im Web. Semantische Informationen können in Bezug auf erstellte Dokumente durch Autorenwerkzeuge direkt in das Dokument integriert oder durch Repositorien mit einem Dokument oder einem Dokumentenpaket in Verbindung gesetzt werden. Man spricht im Kontext von Semantic-Web-konformen Daten meist von Linked Data, da der Gedanke der Verlinkung von semantischen Informationen im Web im Vordergrund steht. Ein für die Auszeichnung von Linked Data etabliertes Format ist RDF. Neben der Angabe von Informationen in literaler Form (alphanumerischer String) ermöglicht RDF die direkte Verknüpfung mit bereits im Semantic Web existierenden Daten. Ein Beispiel hierfür kann die Verlinkung zu einer Personenbeschreibung als Autorenangabe sein.

Für die Interoperabilität der Metadaten eines Dokumentenformates ist insbesondere die Auswahl der Metadatenelemente im vordefinierten Kernset von Wichtigkeit. Aus pragmatischen Gründen werden Metadatenelemente für Dokumentenformate zwar meist nur als fakultativ gekennzeichnet, da man den Produzenten zur Angabe von Metadaten in einem Autorenwerkzeug nicht zwingen sollte, allerdings würde der Einsatz eines Minimalsets an obligatorischen Metadatenelementen eine Verbreitung von Dokumenten, beispielsweise im Web, erheblich fördern. Zu einem obligatorischen Minimalset innerhalb des Kernsets sollten Elemente wie Titel, Autor, Schlagwörter und eine Versionierung in Form des Zeitpunkts der letzten Änderung am Dokument gehören.

Allerdings entscheidet nicht allein die Auswahl von geeigneten Metadatenelementen über die Qualität der Interoperabilität. Von höherer Tragweite ist die Bekanntheit und Verbreitung des Schemas, denen die Metadatenelemente entstammen. Metadatenelemente die beispielsweise in den standardisierten Dublin-Core-Schemas vorkommen, sind weltweit etabliert und können folglich von einer Vielzahl von Systemen interpretiert werden. Somit gilt es bei der Auswahl eines Dokumentenformates, einen genauen Blick auf die unterstützten Schemas und deren Akzeptanz sowie deren Standardisierungsgrad zu werfen.

Eine sich weiterhin abzeichnende Tendenz ist die Möglichkeit vordefinierte Metadaten nicht ausschließlich auf Dokumentenebene angeben zu können, sondern ebenfalls Metadaten einzelnen Komponenten innerhalb des Dokumentes zuzuweisen. Prominente Vertreter für Formate, die dies unterstützen werden, sind HTML5 und ODF 1.2. Folglich wird dieser Aspekt zukünftig auch als ein relevantes Qualitätsmerkmal für interoperable Metadaten in Dokumentenformaten anzusehen sein.

Für eine optimale Einbindung benutzerdefinierter Metadaten sollte das Format neben der Ausweisung der Autorenwerkzeug-spezifischen Serialisierung benutzerdefinierter Metadatenelemente auch in der Lage sein, zusätzliche Metadaten-Schemas und deren Elemente ergänzend im Format integrieren zu können. Durch die Integration standardisierter Schemas ließe sich die Interoperabilität der benutzerdefinierten Metadaten dahingehend erhöhen, dass es für Repositorien nicht zwingend notwendig wäre, jeden Autorenwerkzeug-spezifischen Standard unterstützen zu müssen.

Ein ähnliches Bild zeigt sich für die Ausweisung von kundenspezifischen Metadaten, da es somit auch Anbietern von Drittsystemen möglich wäre, die Metadaten ohne vorherige Konvertierung nachzunutzen.

5.2 Qualitätskriterien für Autorenwerkzeuge zur Erstellung interoperabler Metadaten

Anhand der Darstellung eines Soll-Stands werden nachfolgend Qualitätskriterien für Autorenwerkzeuge beziehungsweise relevante Funktionen zur Erstellung interoperabler Metadaten innerhalb des Autorenwerkzeuges aufgeführt.

Für Benutzer ohne Vorkenntnisse im Bereich Metadatenerfassung ist es am wichtigsten, dass ein gebrauchstauglicher Benutzerdialog für die Aufnahme und Anpassung von Metadaten im Dokument durch das Autorenwerkzeug angeboten wird. Ein Benutzer sollte durch das Autorenwerkzeug darauf aufmerksam gemacht werden, dass er die Qualität der Publikation durch die Metadatenanreicherung erhöhen kann und sollte den notwendigen Benutzerdialog im Autorenwerkzeug leicht und schnell erreichen können. Der Benutzerdialog zum Anlegen und Anpassen vordefinierter Metadatenelemente sollte formularbasiert und übersichtlich gestaltet sein. Die angebotenen vordefinierten Metadatenelemente müssen ein Kernset repräsentieren, das eine Dokumentenverarbeitung in Repositorien und die anschließende Dokumentenpublikation positiv unterstützt. Hierzu gehört insbesondere die Verwendung etablierter und standardisierter Schemas. Die angegebenen Metadaten müssen durch das Autorenwerkzeug in Bezug auf ihre Validität gemäß eines definierten Datentyps überprüft werden. Falls eine Angabe nicht dem zum Element definierten Datentyp entspricht, muss der Benutzer im Benutzerdialog darauf aufmerksam gemacht werden und ein Vorschlag zur Behebung der fehlerhaften Angabe gegeben werden.

Für die Angabe benutzerdefinierter Metadaten ist es wünschenswert, eine Import Funktion für definierte Schemas (bzw. Applikationsprofilen als Zusammenstellung von Metadatenelementen unterschiedlicher Schemas) im Autorenwerkzeug zur Verfügung zu haben. Die Metadatenelemente samt Datentyp sollten, analog zu bisherigen benutzerdefinierten Metadaten, im Benutzerdialogfenster in Auswahllisten für deren Integration im Metadatensatz angeboten werden. Durch den Import von zusätzlichen Schemas und der Ausweisung der Zugehörigkeit von Metadatenelementen zu einem bestimmten Schema, wird die Interoperabilität gegenüber herkömmlichen „Schema-losen“ benutzerdefinierten Metadaten erheblich verbessert. Auch bei der Angabe von benutzerdefinierten Metadaten sollte das Autorenwerkzeug eine Validitätsprüfung bezüglich des Datentyps vornehmen. Bezüglich kundenspezifischer Metadaten ist die einzige Anforderung an das Autorenwerkzeug, dass es diese im Metadatenstream unverändert beibehält.

Es empfiehlt sich, dass das Autorenwerkzeug integrierte Metadaten in einem offenen, standardisierten und Semantic-Web-konformen Format im Dokument serialisiert. In diesem Kontext wäre es wünschenswert, dass das Autorenwerkzeug eine Funktionalität bereitstellt, mittels derer bereits im Semantic Web befindliche Daten (beispielsweise Personenbeschreibungen oder Schlagwörter als normierte Daten) gesucht, ausgewählt und als Verlinkung in den Metadatensatz des Dokumentes aufgenommen werden könnten. Bislang wird dies in keinem klassischen, desktopbasierten Autorenwerkzeug angeboten. Durch die wachsende Verbreitung des Semantic Webs und von Linked Data ist eine solche Funktionalität jedoch sehr gut vorstellbar. Ein Benutzer könnte eine als Linked Date existierende Datenbasis mit Normdaten auswählen, das Autorenwerkzeug kann eine Suche auf dieser ausführen und der

Benutzer ausgewählte Normdaten über Elemente des Metadatensatzes im Dokument verknüpfen.

Wie bereits im Kapitel 5.1 verdeutlicht, werden zukünftige Dokumentenformate es ebenfalls erlauben, Metadaten auf Komponentenebene einzubinden. Folglich muss ein Autorenwerkzeug Funktionalitäten anbieten, die eine Metadatenzuweisung, beispielsweise für einzelne Textabschnitte, Bilder etc. unterstützt.

6 Schlussbetrachtung

Die Ausweisung qualitativ hochwertiger Metadaten ist der Schlüssel für eine erfolgreiche Dokumentenverbreitung im Web und institutionellen Repositorien. In dieser Arbeit wurden Möglichkeiten vorgestellt, wie Autoren bzw. Produzenten von Dokumenten einen ersten Schritt für eine erfolgreiche Metadatenanreicherung im Dokumentenverarbeitungsprozess vornehmen können. Die Erstellung von Metadaten durch Autorenwerkzeuge nimmt zunehmend eine größere Rolle beim Publizieren ein, da es auf diesem Wege Archivierungssystemen im Web ermöglicht wird, Metadaten automatisiert - und nicht wie herkömmlich manuell - in ihr Data Management aufzunehmen.

Die vorliegende Ausarbeitung bietet Autoren bzw. Produzenten von Dokumenten einen Leitfaden für die Auswahl adäquater Dokumentenformate sowie Autorenwerkzeuge, die eine optimale Metadatenanreicherung im Dokument und eine Übernahme in ein Archivierungssystem erlauben.

Die Qualität der Metadaten ist der ausschlaggebende Punkt für die öffentliche Wahrnehmung einer Publikation im Web. Im Zeitalter elektronischer Publikationen ist das Katalogisieren im klassischen Sinne, wie es Jahrzehnte lang von Bibliotheken betrieben wurde, nicht mehr mit dem gleichen Abdeckungsgrad realisierbar, da der Aufwand mit steigender Publikationsanzahl exponentiell ansteigt. Die Konsequenz daraus ist die Notwendigkeit zur automatischen Erschließung von Publikationen. Die vorliegenden Betrachtungen zeigen, dass der Autor bei gegebener Qualität von Dokumentenformaten sowie entsprechender Funktionen im Autorenwerkzeug eine Metadatenanreicherung für die automatisierte Erschließung vorbereiten kann.

Die Entwicklung von neuen Dokumentenformaten, Autorenwerkzeugen sowie Schnittstellen zur Ablieferung von elektronischen Publikationen in Archivierungssystemen müssen der Nachfrage nach hochwertigen Metadaten folglich zukünftig noch stärker gerecht werden.

A Checkliste für Qualitätskriterien

A.1 Interoperabilität von Metadaten in Dokumentenformaten

Nr.	Relevanz	Kriterium	Bewertung	
1	Dokumentenformat			
1.1	2	Offenheit des Dokumentenformats	gegeben	
			nicht gegeben	
1.2	1	Standardisierung des Dokumentenformats	gegeben	
			nicht gegeben	
2	Metadatenstream			
2.1	1	Zugänglichkeit des Metadatenstreams	gegeben	
			nicht gegeben	
2.2	2	Einbindung benutzerdefinierter Metadaten	möglich	
			nicht möglich	
2.3	3	Einbindung kundenspezifischer Metadaten	möglich	
			nicht möglich	
2.4	2	Einsatz von XML-Metadaten	möglich	
			nicht möglich	
2.5	2	Einsatz von RDF-Metadaten	möglich	
			nicht möglich	
3	Verwendung von Metadatenschemas			
3.1	2	Verbreitung der standardmäßig verwendeten Metadatenschemas des Dokumentenformats (sollte für jedes vorkommende Metadatenschema beantwortet werden)	hoch	
			mittel	
			nur in diesem Dokumentformat	
3.2	2	Standardisierung der standardmäßig verwendeten Metadatenschemas des Dokumentenformats (sollte für jedes vorkommende Metadatenschema beantwortet werden)	gegeben	
			nicht gegeben	
3.2	3	Möglichkeit der Integration zusätzlicher Metadatenschemas	möglich	
			nicht möglich	
4	Vordefinierte Metadatenelemente im Metadatenstream des Dokuments			
4.1	1	Titel	vorhanden	
4.2	1	Autor	vorhanden	
4.3	2	Schlüsselwörter	vorhanden	
4.4	2	Beschreibung	vorhanden	
4.5	3	Kategorie	vorhanden	
4.6	3	Status	vorhanden	
4.7	3	Rechte	vorhanden	

4.8	3	Unternehmen	vorhanden	
4.9	1	Datum der Erstellung	vorhanden	
4.10	1	Datum der letzten Änderung	vorhanden	
5	Objekt-Metadaten			
5.1	3	Zuweisung von Metadaten auf Dokumentkomponentenebene	möglich	
			nicht möglich	

Relevanz:

1 = hohe Relevanz

2 = mittlere Relevanz

3 = geringe Relevanz

geringe Relevanz

A.2 Funktionen zur Generierung interoperabler Metadaten in Autorenwerkzeugen

Nr.	Relevanz	Kriterium	Bewertung	
1	Angabe vordefinierter Metadaten			
1.1	1	Erreichbarkeit und Auffindung des Benutzerdialogs zur Angabe vordefinierter Metadaten	sehr gut	
			gut	
			befriedigend	
			ausreichend	
			mangelhaft	
1.2	1	Formulargestützte Angabe von vordefinierten Metadaten	möglich	
			nicht möglich	
1.3	1	Umfang der zur Verfügung stehenden vordefinierten Metadatenelemente	sehr gut	
			gut	
			befriedigend	
			ausreichend	
			mangelhaft	
2	Angabe benutzerdefinierter Metadaten			
2.1	2	Formulargestützte Angabe von benutzerdefinierten Metadaten	möglich	
			nicht möglich	
2.2	1	Festlegung von Datentypen zu einzelnen vordefinierten Metadatenelementen	möglich	
			nicht möglich	
2.3	3	Importfunktion für Metadatenschemas	vorhanden	
2.3.1	3	Aufnahme von Metadatenelementen aus importierten Metadatenschemas in die Auswahlliste der zur Verfügung stehenden benutzerdefinierten Metadaten	möglich	
			nicht möglich	
3	Validitätsprüfung			
3.1	1	Validierung der Angaben innerhalb eines vordefinierten und benutzerdefinierten Metadatenelementes bezüglich des vorgeschriebenen Datentyps	möglich	
			nicht möglich	
3.2	2	Vorschlagssystem/Hilfesystem für die Behebung aufgetretener Konflikte	vorhanden	

4	Umgang mit kundenspezifischen Metadaten			
4.1	1	Unveränderte Aufrechterhaltung der integrierten kundenspezifischen Metadaten im Dokument	ja	
			nein	
5	Verlinkung von Informationen im Semantic Web			
5.1	3	Einbindung von Datenbasen im Semantic Web (Normdateien, Thesauri, Klassifikationssysteme etc.) für die Recherche	möglich	
			nicht möglich	
5.2	3	Suche und Auswahl von Informationen in Datenbasen des Semantic Webs und Verknüpfung dieser als Werte vordefinierter und benutzerdefinierter Metadatenelemente	möglich	
			nicht möglich	
6	Angabe von Objekt-Metadaten			
6.1	3	Funktionalität für die Zuweisung von Metadaten auf Dokumentkomponentenebene	vorhanden	

Relevanz:

1 = hohe Relevanz

2 = mittlere Relevanz

3 = geringe Relevanz