

# Qualitätssicherung für Metadaten

Maik Bunschowski, Marc Röser, Hartmut Schmeck, Djamshid Tavangarian, Denny Voigt

Lehrstuhl für Rechnerarchitektur  
Institut für Informatik  
Universität Rostock  
Albert - Einstein - Straße 21  
18059 Rostock  
maik.bunschowski@uni-rostock.de  
marc.roeser@uni-rostock.de  
djamshid.tavangarian@uni-rostock.de  
denny.voigt@uni-rostock.de

Institut für Angewandte Informatik und  
Formale Beschreibungsverfahren  
Universität Karlsruhe (TH)  
76128 Karlsruhe  
schmeck@aifb.uni-karlsruhe.de

**Abstract:** Metadaten für Lernobjekte stehen als Garant für Wiederverwendbarkeit und Austauschbarkeit und sind aus dem E-Learning-Umfeld nicht mehr wegzudenken. Doch das alleinige Vorhandensein eines standardisierten formalen Metadatenrahmenwerkes ist nicht genug. Die inhaltliche Ausgestaltung ist gerade bei pädagogisch anspruchsvollen Lehr- und Lernressourcen eine schwierige, arbeitsintensive und fehleranfällige Angelegenheit. Qualitätssichernde Maßnahmen sind die logische Konsequenz dieses Umstands. Die Autoren dieses Beitrags wurden mit der Problematik der Erstellung von qualitativ hochwertigen Metadaten konfrontiert und haben im Rahmen zweier Projekte eine praktikable Qualitätssicherungslösung erfolgreich umgesetzt.

## 1. Motivation

Ein Baustein zur Gewährleistung von Interoperabilität digitaler Ressourcen und Sicherstellung von Wiederverwendbarkeit, Auffindbarkeit, Austauschbarkeit und Nachhaltigkeit sind Metadaten ein vielbeschworenes Allheilmittel. Die konkreten Anwendungsfelder hierfür sind recht vielseitig und reichen vom Aufbau des Semantic Web [GR02] über die Entwicklung offener Archive elektronischer Verlage (GAP und DAREnet) bis hin zu Distributionsplattformen zur Verwaltung von Lernobjekten (ARIADNE, Learning Matrix, MERLOT, Learning-Alberta, iLumina etc. [ND02]). Die Verwendung standardisierter Metadatenformate, wie etwa des *Dublin Core Metadata Element Set* (DCMES) für Webdokumente oder *Learning Object Metadata* (LOM) [Lea02] für E-Learning-Ressourcen, ist von entscheidender Bedeutung zur Erreichung dieser Ziele. Das alleinige Vorhandensein von Metadaten gewährleistet nicht eo ipso die Zielerreichung, sondern hängt maßgeblich von der Qualität der Metadaten ab. Jedoch stellen Metadatenstandards lediglich den formalen Rahmen und das Beschreibungsformat zur Metadatenerstellung dar. Die inhaltliche Metadatenerstellung für pädagogisch anspruchsvolle Lehr- und Lernressourcen kann ein äußerst anspruchsvoller und komplexer Vorgang sein, der nicht nur hohe Anforderungen an die kognitive Konditionierung des Erstellers stellt, sondern auch

im hohem Maße Fachwissen, pädagogische Erfahrung, informationstechnische Grundfertigkeiten und ein bibliothekarisches Grundverständnis voraussetzt.

Wie aktuelle Studien belegen [CB03] ist nicht jedem Metadatenersteller die Bedeutung und Notwendigkeit von korrekten Metadaten bewusst, was dazu führen kann, dass Metadaten nicht mit der notwendigen Sorgfalt erstellt werden. Hinzu kommt, dass gerade bei der Metadatenerstellung für Lehr- und Lernressourcen auch didaktisch anspruchsvolle Metainformationen erhoben werden, was dazu führt, dass es äußerst diffizil ist, korrekte und vergleichbare Metadaten zu erstellen. Beispielsweise führen Baumgartner [Bau04] und Schulmeister [Sch02] die Schwierigkeiten bei der Erstellung des interpretativen LOM-Elements *interactivity* an. Der potentielle Dienst bzw. Mehrwert einer unzureichend mit Metadaten ausgezeichneten Ressource kann damit nicht auf die erhoffte Weise nutzbar gemacht werden und im schlimmsten Fall den Verlust im Datennirwana bedeuten. Um dies zu vermeiden, sind gerade im Anwendungsfall des E-Learning qualitätssichernde Maßnahmen für die Erstellung von Metadaten unumgänglich, aber weitestgehend Terra incognita.

Der vorliegende Beitrag fokussiert Qualitätssicherung von Metadaten aus dem E-Learning-Umfeld, stellt diese vor und zeigt anhand von Projektergebnissen aus den Projekten *Wissenswerkstatt Rechensysteme* (WWR) [wwr05] und *Automatic Manuscript Generation* (AMG) [amg05] die Zuverlässigkeit entwickelter Qualitätssicherungsmaßnahmen. Bevor die Konzeption und praktische Umsetzung dieser Maßnahmen im Workflow für Erstellung und Veröffentlichung von Lernobjekten vorgestellt wird, bildet der folgende Abschnitt mit einer kurzen theoretischen Vorbetrachtung die Grundlage für diesen Artikel, ergänzt durch eine Darstellung des momentanen Forschungsstandes in diesem Bereich.

## 2. Theoretische Vorbetrachtung

In diesem Abschnitt soll zur Schaffung eines klaren Grundlagenverständnisses kurz erläutert werden, was unter Metadaten zu verstehen ist, wie sie sich inhaltsbasiert unterscheiden lassen und welche Möglichkeiten zur Charakterisierung von Qualität es gibt.

### 2.1 E-Learning-Metadaten

Eine Hauptaufgabe von Metadaten für Lernobjekte ist es, den Einsatz und die Wiederverwendbarkeit von Lernobjekten zu unterstützen. Hierfür hat das *Learning Technology Standards Committee* des *Institute of Electrical and Electronics Engineers* ein formales Rahmenwerk entworfen, in Form eines abstrakten Verbalmodells [Lea02]. Dieses standardisierte verbale Datenschema enthält Kategorien und Elemente, welches ein Lernobjekt dergestalt beschreibt bzw. charakterisiert, dass die oben genannte Zielsetzung für alle denkbaren Lernobjekteinsatzszenarien Unterstützung erhält. Um diesem hochgesteckten Ziel gerecht zu werden, ist es nicht verwunderlich, dass für Version 1.0 des verbalen Basisschemas neun Kategorien mit über 57 Elementen notwendig sind, wobei erwähnt sei, dass einzelne Elemente sich noch weiter verschachteln lassen und dadurch

mehrfach vorkommen können [Lea02]. Ein Element ist durch einen Wertebereich und Typ definiert, wobei sich der Wertebereich wiederum auf einen bestimmten Standard (z.B. RFC2048:1996 für die MIME Typisierung) bezieht oder in Form eines Vokabulars angegeben werden kann. Aus Erfahrungen der Projekte WWR und AMG wird geschätzt, dass zwischen 200 und 500 LOM-Elemente zur Charakterisierung eines Lernobjektes benötigt werden. Es liegt auf der Hand, dass eine solche Charakterisierung einen nicht zu unterschätzenden Umfang und Aufwand besitzt. Wie bereits beispielhaft erwähnt, sind einzelne Elemente und deren Verwendungsweise zudem nicht unumstritten (Beispiel LOM-Element *interactivity*) und bedürfen eines breiten technischen und pädagogischen Grundverständnisses. Da für eine vollständige und adäquate Beschreibung eines Lernobjektes ein nur verbales Basisschema nicht ausreicht, muss eine syntaktische und semantische Spezifikation erfolgen. Dies geschieht durch das so genannte Binding, indem das verbale Basismodell technisch umgesetzt wird. Die technologische Realisation kann beispielsweise mittels XML, RDF oder jedem anderen geeigneten Sprachgerüst erfolgen.

Alleine der Umfang, die Komplexität, die Schwierigkeiten mit einzelnen Metadatenelementen und die Notwendigkeit zur technischen Implementierung, gerade im Hinblick auf technisch nicht versierte Personen, offenbaren das Fehlerpotential bei der Metadatenerstellung.

## 2.2 Inhaltsbasierte Metadatenklassifikation

Ungeachtet der Komplexität, des Umfangs und ihrer domänenspezifischen Provenienz lassen sich Metadaten bzw. einzelne Metadatenelemente inhaltsorientiert unterscheiden. Eine solche allgemeine Einteilung hilft nicht nur definitorische Klarheit zu erlangen, sondern wird nachfolgend zur Zuordnung von Maßnahmen, Prozessen und Qualitätskriterien benötigt. Nach Kashyap und Sheth [KS98] werden Metadaten zunächst zwischen *inhaltsunabhängigen* und *inhaltsabhängigen* Metadaten unterschieden. Erstere enthalten Informationen, die nicht vom Inhalt der beschriebenen Ressource abhängen, wie etwa den Ort der Ressource oder das Datum der letzten Modifizierung. Inhaltsabhängige Metadaten sind demgegenüber direkt vom Ressourceninhalt abhängig, z.B. Titel eines Dokuments, und lassen sich wiederum in die Kategorien *direkt inhaltsbasiert* und *inhaltsbeschreibend* einteilen. Während direkt inhaltsbasierte Metadaten den Inhalt zur Beschreibung der Ressource heranziehen, z.B. Volltextindizes, tun dies inhaltsbeschreibende Metadaten ohne direkte Nutzung des Ressourceninhalts. Ein Beispiel hierfür sind etwa Annotationen bestimmter Dokumentenabschnitte. Der letztgenannte Metadatentyp wird dann schließlich nochmals in *domänenunabhängige* und *domänenspezifische* Metadaten unterschieden, die entweder unabhängig von deren konkreter Anwendung oder Thematik sind, wie z.B. DTDs, oder davon abhängig sind (und zumeist mithilfe domänenspezifischer Vokabulare beschrieben werden). Die komplette Klassifikation wird in Abbildung 1 veranschaulicht. Basierend auf dieser Kategorisierung ergeben sich unterschiedliche Anforderungen an jede einzelne Kategorie hinsichtlich der Metadatenerstellung als auch die damit verbundene Qualitätssicherung.

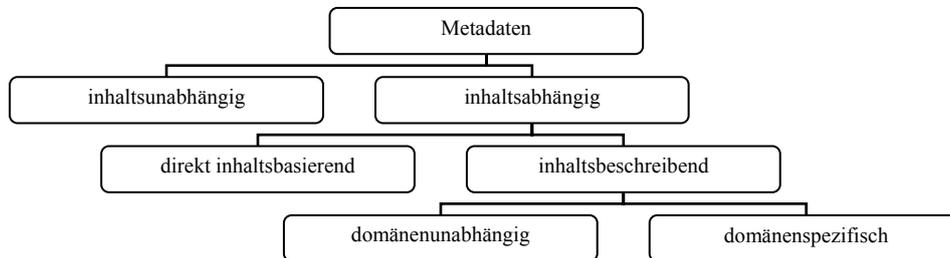


Abbildung 1: Metadatenklassifizierung nach Kashyap und Sheth [KS98].

### 2.3 Qualitätskriterien

Auf die Qualität von Daten im Allgemeinen bezogen, hat sich mittlerweile die Phrase „fitness for use“ als prägnante und häufig verwendete Definition eingepreßt [PS04]. Die Forschungsbemühungen der letzten Jahre haben eine Reihe von Ansätzen und Sichtweisen hervorgebracht, wie man zur Identifizierung und Klassifizierung von Qualitätsmerkmalen vorzugehen hat. Price und Shanks [PS04] geben dazu einen Überblick über die verschiedenen derzeit existierenden Ansätze für Datenqualität im Allgemeinen und erarbeiteten aufgrund der verschiedenen Defizite existierender Ansätze einen auf der Semiotik basierenden Rahmen für das Verständnis und die Definition von Daten- bzw. Informationsqualität. Die vorliegende Arbeit greift dieses *Semiotic Information Quality Framework* [PS04] auf und leitet daraus Qualitätskriterien für den speziellen Anwendungsfall von Metadaten für Lernobjekte ab, in eine syntaktische, semantische und pragmatische Ebene. Ohne diese Kriterien hier im Einzelnen ausführlich zu erläutern, seien sie genannt. Im Rahmen der Syntaxebene wird das Kriterium der Konformität erhoben. Für die semantische Ebene sind es die Kriterien der Vollständigkeit, Eindeutigkeit, Korrektheit, Aussagekraft und Redundanzfreiheit. Regelkonformität, Zuverlässigkeit, Vollständigkeit, Verständlichkeit, Erreichbarkeit, Sicherheit, flexible Präsentationsform, angemessene Präsentation, Relevanz und Nützlichkeit sind Kriterien der pragmatischen Ebene. Nachdem diese Kriterien im Sinne von Price und Shanks [PS04] detailliert formuliert und für den jeweiligen Anwendungsfall ausgearbeitet worden sind, werden sie in einem Kriterienkatalog zusammengefasst. Dieser Kriterienkatalog hat primär zwei Aufgaben. Zum einen als Pflichtenheft für die Implementierung eines Arbeitsflusses für den Metadatenerstellungs- und Veröffentlichungsprozess, zum anderen als objektiver Maßstab bzw. Prüfungsgrundlage der qualitätssichernden Maßnahmen, die im Metadatenworkflow umgesetzt worden sind. Die Erstellung eines solchen Kriterienkatalogs ist nicht nur für das Management, die Entwickler und die Autoren eines Projektes wichtig, sondern stellt auch eine Art Qualitätssiegel für den Endanwender dar.

### 3. Aktuelle Forschung

Eine gewissenhafte Qualitätssicherung von Metadaten unter Anwendung der aufgezählten Qualitätskriterien darf nicht als einmalige punktuelle Obliegenheit aufgefasst werden, sondern kennzeichnet sich durch einen dauerhaften Prozess, der bereits bei der

initialen Erstellung der Metadaten beginnt und die Lebensdauer der zugrunde liegenden Ressource besitzt. A priori ist die Intensität dieses Prozesses nicht gleichverteilt, sondern bedarf besonderer Aufmerksamkeit bei der Metadatenerstellung. Aus diesem Grund werden im Folgenden für die Qualitätssicherung die Phasen während und nach der Metadatenerstellung unterschieden.

### 3.1 Erstellungsphase

In Anlehnung an Gilliland-Swetland [GS98] können Metadaten in manuell und automatisch generierte unterschieden werden. Wie bereits erwähnt, können standardisierte Metadaten, aufgrund von unterschiedlichen Metadatenelementen, mehreren Kategorien zugeordnet werden. Metadatenelemente eines Metadatensatzes können somit manuell oder automatisch erzeugt werden, wodurch ein Metadatensatz (als Ganzes) einen hybriden Charakter besitzen kann.

Bei der *manuellen* Metadatengenerierung zählen Currier und Barton [CB03] den Ersteller der Lernressource (auch *Inhaltsersteller* genannt) und den *domänenspezifischen Fachexperten* zu den am häufigsten vorkommenden Typen von Metadatenautoren. Gerade bei den Inhaltserstellern ist die Sorgfalt und Einsicht zur Angabe qualitativ hochwertiger Metadaten meist nicht vorhanden und kann per se nicht vorausgesetzt werden, insbesondere wenn sie aus nicht-technischen Bereichen stammen. Dieser Personenkreis muss deshalb zunächst für die Thematik „Metadaten“ sensibilisiert werden, um deren Notwendigkeit und Wichtigkeit zu verdeutlichen sowie den damit verbundenen Mehraufwand zu legitimieren. Unterstützung während des Erstellungsprozesses geben vor allem intuitiv zu bedienende Metadateneditoren mit entsprechender Online-Hilfe und anschaulichen Beispielen. Nichtsdestotrotz kann die Qualität der erzeugten Metadaten von Autor zu Autor sehr stark schwanken, wodurch die Vergleichbarkeit der Metadaten offensichtlich beeinträchtigt wird. Allerdings verfügt der Inhaltsersteller naturgemäß über die besten Kompetenzen, wenn es um die Erstellung inhaltsbeschreibender Metadaten geht. Im Gegensatz zu Ressourcenerstellern sind sich Fachexperten von domänenspezifischen Metadaten ihrer Bedeutung sehr wohl bewusst, was sich in ihrer Motivation und Sorgfaltspflicht zur Metadatenerstellung niederschlägt. Sie besitzen die notwendigen Kompetenzen zur Erstellung inhaltsunabhängiger und domänenspezifischer Metadaten, insbesondere was die korrekte thematische Klassifizierung von Ressourcen und die Verwendung festgelegter domänenspezifischer Vokabulare als Wertemengen bestimmter Metadaten anbelangt.

Da weder die Inhaltsersteller noch die Fachexperten in der Lage sind, einen Metadatensatz alleine in der geforderten Qualität zu erstellen, werden in zunehmendem Maße kollaborative Ansätze zur Metadatenerstellung propagiert [CB03], bei denen sowohl Ressourcenersteller als auch Fachexperten in gemeinschaftlicher Weise am Metadatensatz arbeiten und diesen erstellen. Hierbei bearbeiten die beteiligten Metadatenautoren die ihnen zugewiesenen Metadatenelemente parallel oder komplettieren den Metadatensatz in vorgeschriebener sequentieller Abfolge. Entsprechend den vorhandenen Kompetenzen würde der Ressourcenersteller vornehmlich inhaltsbasierte und domänenunabhängige Metadaten bearbeiten, während die beteiligten Fachexperten in erster Linie für die do-

mänenspezifischen und inhaltsunabhängigen Metadaten verantwortlich sind [MF99]. Ein konstruktiver Austausch zwischen den Metadatenautoren während des Erstellungsprozesses ist dabei der Qualität zweifelsohne zusätzlich förderlich. Im Rahmen des VIROR-Projekts wurde für die Verschlagwortung von Lernmodulen ein solcher Ansatz gewählt, bei dem der Inhaltsersteller durch die spezifischen Fachreferenten der Universitätsbibliotheken Unterstützung erhält [VHO02a]. Natürlich darf eine solche Zusammenarbeit nicht nur auf die Verschlagwortung beschränkt sein, sondern muss sich auf die zu den jeweiligen Kompetenzen aufgeteilten Personen verteilen. Gerade im Kontext pädagogisch anspruchsvoller Lehr- und Lernressourcen ist die Einbeziehung eines ausgebildeten Didakten unumgänglich, als Beispiel sei die Metainformation des LOM-Elements 5.6 *Educational Context* genannt.

*Automatische Verfahren* zur Metadatenextraktion ergänzen und unterstützen den manuellen Prozess und kommen gerade dann zum Einsatz, wenn aus der zugrunde liegenden Ressource mittels algorithmischer Operationen Daten für die Metadaten gewonnen werden können [LA02]. In Bunschowski et al. [BLRT03] werden Verfahren und Methoden zur Metadatengenerierung aufgezeigt, die auf einer XML-basierenden Ressourcenbasis bis zu 74% der Metadaten automatisch generieren. Dabei kommen Verfahren des Information Retrieval und der Linguistik-basierten Informationsextraktion zum Einsatz. Unter der Voraussetzung, dass die zugrunde liegenden Algorithmen zuverlässig arbeiten, liegen die Vorteile der automatischen Metadatengenerierung klar auf der Hand, entlasten sie doch ungemein bei Routine- und Fleißarbeiten, wodurch die Fehlerquelle Mensch zumindest für diesen Bereich minimiert werden kann.

Einige Autoren stehen den aktuellen technischen Möglichkeiten der automatischen Metadatengenerierung skeptisch gegenüber, weil in vielen Fällen die zugrunde liegende Ressource nur begrenzten textuellen Kontext besitzt und somit eine automatische Generierung sehr schwer bzw. unmöglich wird [CB03]. Zwei maßgebliche Faktoren, die diese Situation bestimmen, sind die Art der Ressource (handelt es sich um ein proprietäres Ressourcenformat und beinhaltet die Ressource bereits Metadaten) und die entsprechend der Klassifikation zu erzeugenden Metadattypen (siehe Abschnitt 2.2). In Abhängigkeit davon und unter Abwägung des Aufwandes der Entwicklung von Automatismen zur automatischen Generierung, lassen sich in vielen Fällen Metadaten automatisch generieren [GR02].

Für die automatische Berechnung inhaltsunabhängiger Metadaten eignen sich meist algorithmische Verfahren, die das Umfeld der zugrunde liegenden Ressource analysieren und auswerten. Beispielsweise kann das LOM-Element 4.2 (size of the digital learning object) durch Aufruf einer simplen Methode zur Bestimmung der Datei- bzw. Zeichenisgröße ermittelt werden. Die Extraktion inhaltsbasierender Metadaten erfordert i.d.R. Methoden der Informationsextraktion, deren Komplexität sich entsprechend dem Charakter der zugrunde liegenden Ressource bemisst. So kann im einfachsten Fall, wenn die Ressource z.B. bereits implizit Metadaten in sich trägt, ein einfaches Mapping stattfinden. Bei der Verarbeitung von natürlichsprachlichen Ressourcen müssen hingegen Werkzeuge aus der Computerlinguistik Einsatz finden [NP02]. Da diese jedoch noch fehleranfällig sind, bedürfen so ermittelte Metadaten einer manuellen Qualitätssicherung und dienen in der Praxis nur als Vorschlag für den Metadatenautor. Auf diese Weise

kann z.B. das LOM-Element 1.5 (keyword) extrahiert werden. Die automatische Generierung inhaltsbeschreibender Metadaten funktioniert zum gegenwärtigen Zeitpunkt allerdings noch nicht zuverlässig und ist Gegenstand aktueller Forschungsbemühungen [CMD05, JJBW99].

Der Anteil automatisch extrahierbarer Daten eines Metadatensatzes variiert entsprechend der Ausrichtung der zugrunde liegenden Metadatenpezifikation. So lässt sich ein vorwiegend aus inhaltsunabhängigen Metadaten bestehender DCMES-konformer Metadatenatz zu einem größeren Anteil automatisch generieren, als ein LOM-konformer Metadatenatz, der aufgrund seines vergleichsweise hohen Anteils inhaltsbeschreibender und inhaltsbezogener Metadaten mehr manuelles Zutun verlangt.

### 3.2 Qualitätssicherung nach der Erstellungsphase

Diese der Metadatenerstellung nachgelagerte Phase der Qualitätssicherung berücksichtigt sowohl syntaktische, semantische als auch pragmatische Qualitätskriterien. Maßgeblich hierfür sind die bewährten Verfahren Peer Reviewing, Peer Commentary und Automated Quality Assurance-Measures aus dem E-Print-Bereich zu erwähnen, die für die Qualitätssicherung von Metadaten zunehmend an Bedeutung gewinnen. Die Verwendung eines *recommender systems* für Metadaten kann weiterhin die Qualität und Akzeptanz nachhaltig sicherstellen und inaktuelle bzw. veraltete Metadaten kennzeichnen. Darüber hinaus finden die im letzten Abschnitt behandelten Verfahren, die während der manuellen und automatischen Metadatenerstellung zum Einsatz kommen, auch zur qualitätssichernden Überprüfung bereits erzeugter Metadaten Anwendung.

*Peer Reviewing* ist eine Qualitätssicherungsmaßnahme, die vor der eigentlichen Bereitstellung von Dokumenten zum Tragen kommt und sich häufig der Mithilfe freiwilliger Expertengutachter bedient [mer05]. Die Unterschiede der verschiedenen Varianten des Peer Reviewings sind genauer in Armstrong [Arm97] herausgestellt worden, wo unter anderem auch der Einfluss auf Fairness, Qualität sowie Zitations- und Publikationshäufigkeit der begutachteten Werke evaluiert wird. Die Begutachtung von Experten wurde im Rahmen des bereits genannten VIROR-Projekts umgesetzt, bei dem ein so genannter Validator die Überprüfung der Qualität der Metadaten zur Aufgabe hat [VHO02b]. Evident und als Qualitätssicherungsmaßnahme unumgänglich ist jedoch, dass der Peer Reviewing-Prozess in den technischen Workflow vollständig eingebunden wird und von Benutzern nicht umgangen werden kann.

*Peer Commentary* ist eine Variante der Begutachtung von Daten, die erst nach der eigentlichen Veröffentlichung zum Tragen kommt [Har96]. Dabei werden Kommentare, Bewertungen oder Änderungsvorschläge über verschiedene Medien (z.B. Foren, Wikis, E-Mail) zu den Autoren geleitet [Sca03]. Dieser Ansatz zeichnet sich im Metadatenkontext dadurch aus, den Anwendern die Entscheidung bezüglich der (Nicht-) Verwendung von Dokumenten zu erleichtern, sowie die Metadatenautoren mit Hinweisen für Folgeversionen zu versorgen. Nachteilig ist, dass zwischen der Bereitstellung der Metadatenätze und den ersten Anwenderkommentaren ein großer zeitlicher Abstand liegen kann und die nachträgliche Begutachtung nicht zu erzwingen ist.

*Automatisierte Verfahren zur Qualitätssicherung* lassen sich in der Form von Validatoren zur formalen Qualitätssicherung den oben genannten Gruppen zuordnen oder können die Datenaufbereitung im Vorfeld unterstützen. Für metadatenbezogene Arbeitsflüsse bietet sich die Informationsextraktion von formal exakt definierten Teilen der Datensätze (wie die Autorennamen) aus den Ausgangsdokumenten an. Dem gegenüber steht die mangelhafte Sicherstellung semantischer und pragmatischer Qualitätskriterien, die auch durch Maßnahmen aus der künstlichen Intelligenz nur unvollkommen zu erreichen ist. Ebenso wie bei der Metadatenerstellung sind automatisierte Verfahren zur Qualitätssicherung als Ergänzung zu den klassischen Qualitätssicherungsmaßnahmen Peer Reviewing und Peer Commentary zu sehen. Es muss im Einzelfall entschieden werden, welche Kombination von Maßnahmen sinnvoll ist.

#### **4. Praktische Anwendung**

Im vorangegangenen Abschnitt wurde evident, dass Qualitätssicherung ein durchgängiger Prozess ist, der bereits bei der Erstellung von Metadaten beginnt. Allerdings zeigen Studien, dass sich bestehende Ansätze zur Qualitätssicherung in Bezug auf Metadaten noch nicht durchsetzen konnten [RW03, Cur01]. Dieses Faktum impliziert das Fehlen eines integrierten, metadatenzentrierten Arbeitsflusses, der um eben die für Qualitätssicherung erforderlichen Prozeduren angereichert ist. Ein Ziel ist es, diese etablierten Maßnahmen zur Qualitätssicherung auf geeignete Weise in einem Arbeitsfluss zu verbinden, der in einem realen Anwendungsfall für Lehr- und Lerndokumente umgesetzt werden soll. Aus diesem Grunde und aus einer durch die praktischen Erfordernisse der Forschungsprojekte AMG und WWR motivierten Forderung heraus, wird im Folgenden einerseits ein solcher Arbeitsfluss definiert und andererseits seine Umsetzung im Rahmen der genannten Projekte vorgestellt. In diesem Abschnitt wird keine Alternative zu bewährten Qualitätssicherungsmaßnahmen wie Peer Reviewing, Peer Commentary und automatisierten Validierungsprozessen für Formalkriterien geschaffen. Die Absicht ist es vielmehr, diese etablierten Maßnahmen auf geeignete Weise in einem Arbeitsfluss zu verbinden, der in einem realen Anwendungsfall für Lehr- und Lerndokumente umgesetzt werden soll. Aufbauend auf diese Überlegungen wird anhand der im Abschnitt 3.1 vorgenommenen Charakterisierung von Maßnahmen zur Qualitätssicherung ein Inhaltserstellungs- und Veröffentlichungsprozess entwickelt, der insbesondere auf die Belange von Metadaten zugeschnitten ist.

Grundsätzlich umfasst der Basisworkflow für Inhalts- und Metadatenveröffentlichung drei Prozesse, den der Inhaltserstellung, Metadatenerstellung und Veröffentlichung in einem Repository. Es wird verallgemeinernd davon ausgegangen, dass die genannten Prozesse sequentiell und beliebig oft durchlaufen werden können. Eine verlässliche Grundlage für die Erweiterung des Workflows um Qualitätssicherungsaspekte bedingt die Anwendung kontrollierender Maßnahmen. Wie bereits gezeigt worden ist, stellen Peer Review, Peer Commentary und automatische Qualitätssicherungsmaßnahmen geeignete Möglichkeiten zur Kontrolle von Qualitätskriterien bereit. Aus dem Vergleich aus Abschnitt 3.2 wurde deutlich, dass keine Maßnahme allein die Einhaltung aller Qualitätskriterien gewährleisten kann. Deshalb wird im Folgenden ebenso wie in [Sca03] ein Verbund aus Pre- und Post-Publishing-Ansätzen in Kombination mit automatischen

Verfahren zur Qualitätssicherung vorgeschlagen. Die Verbindung der optimalen Einsatzgebiete manueller und automatischer Verfahren führt zu einem Verbund-workflow, der in Abbildung 2 graphisch verdeutlicht worden ist. Aufgrund projektspezifischer Gegebenheiten fand eine kollaborative manuelle Metadatenerstellung bislang keine technische Umsetzung, sollte jedoch grundsätzlich Berücksichtigung finden.

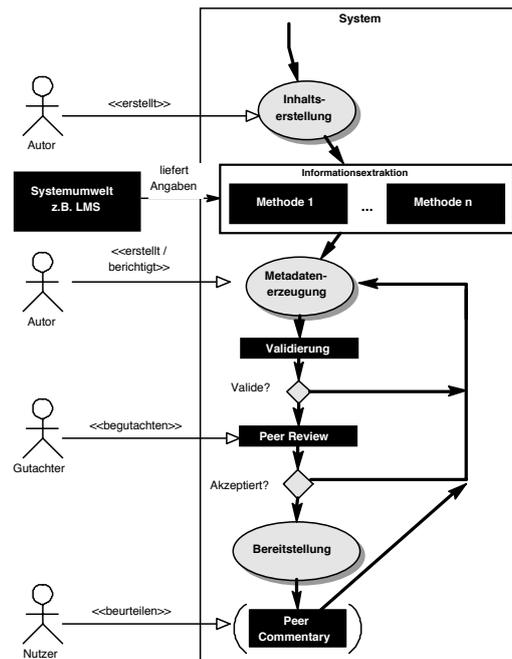


Abbildung 2: Vereinfachte Darstellung des erweiterten Workflow für Inhalts- und Metadatenveröffentlichung mit qualitätssichernden Maßnahmen.

In der sequentiellen Abarbeitung des Arbeitsablaufes erfolgt zunächst die Inhaltserstellung, die mit projektspezifischen Eingaberichtlinien erleichtert werden kann. Damit soll die optimale Anpassung der Materialien an spätere Automatismen gesichert und eine möglichst einheitliche Qualität der Rohdaten erzielt werden, die dem Metadateneingabeprozess zugrunde liegen. Im Anschluss daran erfolgt die Extraktion eines möglichst großen Anteils automatisch erhebbarer Metadatenwerte. Dieser Vorgang soll die Arbeit der Metadatenautoren erleichtern und häufig wiederkehrende Aufgaben in der Dateneingabe minimieren helfen. In der Implementierung wird eine automatische Datenextraktion unter Zuhilfenahme eines vordefinierten XML-Mappings [BVPK04], Verwendung externer Werkzeuge bzw. Systemfunktionalitäten, welche Aufschluss über das Lernobjekt, den Inhaltsteller und ggf. den Lerner geben, und erprobten Textextraktionswerkzeugen [NP02] vorgenommen. Daran anschließend werden die nicht automatisch extrahierbaren Metadatenwerte manuell eingegeben und die so entstandenen Metadatenätze automatisch auf Vollständigkeit und formale Korrektheit überprüft. Diese Überprüfung erfolgt in zwei Phasen. Im umgesetzten Client-Server-basierten Konzept erfolgt eine erste Prüfung auf Vollständigkeit der einzugebenden Pflichtdaten unmittelbar vor dem

Absenden an den Portalserver. Dabei wird durch ein clientseitiges Script geprüft, ob alle als Pflichtfelder markierten Datenbereiche ausgefüllt worden sind. Der zweite Prüfungsschritt erfolgt auf dem Server, wo eine Validierung der übertragenen Metadatenätze hinsichtlich Konformität zu der zugrunde liegenden Metadatenpezifikation erfolgt.

An die Sicherung der formalen Qualitätskriterien schließt sich ein nicht anonymes, auf die Inhalte gestütztes, Peer Reviewing an, dass durch die manuelle Zuweisung kompetenter Gutachter durch einen oder mehrere Repositoryverwalter gekennzeichnet ist. Dieser Evaluationsvorgang kann nicht umgangen werden und soll gewährleisten, dass nur hochqualitative Inhalte zusammen mit angemessenen Metadaten publiziert werden können. Die Repositoryverwalter arbeiten vergleichbar mit den Editoren in wissenschaftlichen Veröffentlichungen, die den generellen Begutachtungs- und Publikationsprozess verwalten und überwachen. Diese Verwaltung der eingegebenen Inhalte und Metadaten erfolgt über eine webbasierte Administrationsschnittstelle, die ebenfalls Bestandteil einer umfassenden Portalarchitektur [BKR<sup>+</sup>04] ist. Zusätzlich zu diesem webbasierten Ansatz, der auch die Arbeit der Gutachter unterstützt, indem die zu evaluierenden Lehr- und Lernmodule zusammen mit den vorgeschlagenen Metadatenätzen innerhalb ihrer Arbeitsumgebung eingeblendet werden, findet das Medium „E-Mail“ Anwendung. Verwalter werden auf diesem Wege auf Änderungen des Begutachtungsstatus einzelner Module hingewiesen und Gutachter über neue Evaluationsaufgaben informiert. Der Versand dieser Nachrichten erfolgt automatisch vom System gesteuert und ohne zusätzliche Anwenderinteraktion. Nach der Veröffentlichung steht es den Anwendern frei, Kommentare zu der Qualität der Metadatenätze und Inhalte abzugeben. Diese können über verschiedene Medien an den Metadatenautor weitergeleitet und bei einer eventuellen Revision der Metadaten berücksichtigt werden.

## **5. Zusammenfassung**

Eine fehlende Umsetzung wirkungsvoller Mechanismen zur Sicherung der Metadatenqualität bewirkt, dass unzureichend oder fehlerhaft ausgezeichnete Ressourcen schlimmstenfalls spurlos in den Tiefen ihrer Repositories verschwinden und damit ihrer vorgesehenen Nutzung entgehen. Zwar zeichnet sich jüngst ein Trend ab, indem sich einzelne Bemühungen der automatischen Unterstützung bei der Metadatenerstellung annehmen, doch eine umfassende Betrachtungsweise für den Metadatenerstellungsprozess fehlt. Qualitätssicherung von Metadaten ist keine punktuelle Aufgabe, sondern ein nachhaltiger Prozess, der lediglich situativ bedingt durch automatische Maßnahmen unterstützt werden kann. Das Konzept, die Umsetzung und Schaffung entsprechender Rahmenbedingungen zur Erzeugung von qualitativ hochwertigen Metadaten ist gerade für E-Learning-Ressourcen eine nicht zu unterschätzende Managementaufgabe, die technische, administrative und pädagogische Aspekte zu berücksichtigen hat.

Der vorliegende Beitrag greift nun diese Problematik auf und leitet aus bestehenden zahlreichen Einzelmaßnahmen zur Sicherung der Qualität von Daten im Allgemeinen eine konzeptuelle Vorgehensweise zur Qualitätssicherung für Metadaten im Rahmen der genannten Projekte ab. Es wurde sowohl eine tiefgreifende Qualitätssicherung beim Erstellen von Metadaten durch das Metadatenautorenwerkzeug demonstriert, als auch

durch eine dem Erstellungsprozess nachgelagerte Qualitätsprüfung. Die Integration dieser Maßnahmen in den gesamten Ressourcenerstellungsworkflow trägt maßgeblich dazu bei, qualitativ hochwertige Metadaten zu erzielen. Die aufgezeigten Verfahren und Methoden stellen einen praktischen Lösungsansatz dar, welcher für unterschiedlichste Projekte als Anleitung dienen kann und ein erster wichtiger Schritt zur Sensibilisierung für qualitativ hochwertige Metadaten ist.

## 6. Literaturverzeichnis

Alle angegebenen Links wurden am 20.03.2005 auf Erreichbarkeit geprüft.

- [amg05] AMG - Automatic Manuscript Generation. URL <<http://www.amg-project.de>>, 2005.
- [Arm97] J. Scott Armstrong. Peer Review for Journals: Evidence on Quality Control, Fairness, and Innovation. *Science and Engineering Ethics*, 3(1):63–84, 1997.
- [Bau04] Peter Baumgartner. Didaktik und Reusable Learning Objects (RLOs). In *Campus 2004: kommen die digitalen Medien an den Hochschulen in die Jahre?* Doris Carstensen (Hrsg.), 2004.
- [BKR<sup>+</sup>04] Maik Bunschowski, Lars Kornelsen, Daniel Reichart, Marc Röser, Djamschid Tavangarian, Denny Voigt and Matthias Waldhauer. Harnessing Scalable and Future Proof Teachware for Existing Learning Management Systems by Providing a Service Based Middleware. *Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunication (ED-MEDIA)*, Lugano, Switzerland, 2004.
- [BLRT03] Maik Bunschowski, Ulrike Lucke, Marc Röser and Djamschid Tavangarian. The role of metadata for the automated generation of educational material. In *Proceedings on the 2nd European Conference on e-Learning ECEL*. Academic Conferences Limited, November 2003.
- [BVPK04] Aida Boukottaya, Christine Vanoirbeek, Federica Paganelli and Omar Abou Khaled. Automating XML documents transformations: a conceptual modelling based approach. In *CRPIT '31: Proceedings of the first Asian-Pacific conference on Conceptual modelling*, pages 81–90. Australian Computer Society, Inc., 2004.
- [CB03] Sarah Currier and Jane Barton. Quality Assurance for Digital Learning Object Repositories: How Should Metadata Be Created? In J. Cook and D. (Eds.) McConnell, editors, *Communities of Practice*, University of Sheffield & Sheffield Hallam University, September 2003. *ALT-C 2003 Research Proceedings*, ALT - the Association for Learning Technology.
- [CMD05] Kris Cardinaels, Michael Meire and Erik Duval. Automating Metadata Generation: the Simple Indexing Interface. Preprint of an Article accepted for publication in *ACM 1-59593-046-9/05/0005*. International World Wide Web Conference Committee, WWW 2005, Chiba, Japan, 2005.
- [Cur01] Sarah Currier. *SeSDL Taxonomy Evaluation Report*. University of Strathclyde, November 2001. Verfügbar im Internet: URL <[http://www.sesdl.scotcit.ac.uk:8082/taxon\\_eval/SeSDLTaxFinRep.doc](http://www.sesdl.scotcit.ac.uk:8082/taxon_eval/SeSDLTaxFinRep.doc)>.
- [GR02] Jane Greenberg and W. Davenport Robertson. Semantic Web Construction: An Inquiry of Authors. In *Proceedings of the International Conference on Dublin Core and Metadata for e-Communities 2002*, pages 45–52. Firenze University Press, October 2002.
- [GS98] Anne J. Gilliland-Swetland. *Introduction to metadata: pathways to digital information*. [Los Angeles, Calif.] : Getty Information Inst., 1998.
- [Har96] Stevan Harnad. *Scholarly Publication: The Electronic Frontier*, pages 103 - 108,

1996. Verfügbar im Internet: URL  
<<http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad96.peer.review.html>>.
- [JJBW99] Charlotte Jenkins, Mike Jackson, Peter Burden and Jon Wallis. Automatic RDF Metadata Generation for Resource Discovery. In WWW '99: Proceeding of the eighth international conference on World Wide Web, pages 1305–1320. Elsevier North-Holland, Inc., 1999.
- [KS98] Vipul Kashyap and Amit Sheth. Semantic heterogeneity in global information systems: The role of metadata, context and ontologies. In Michael P. Papazoglou [Hrsg.], editor, Cooperative information systems: trends and directions, pages 139–178, Academic-Press New York, 1998.
- [LA02] Andreas D. Lattner and René Apitz. A Metadata Generation Framework for Heterogeneous Information Sources. In Proceedings of the 2nd International Conference on Knowledge Management (I-KNOW 02), Graz, Austria, July 11-12, pages 164–169, 2002.
- [Lea02] Learning Technology Standards Committee of the IEEE, 3 Park Avenue, New York, NY 10016-5997, USA. Draft Standard for Learning Object Metadata, IEEE Standards Department edition, July 2002. IEEE 1484.12.1-2002.
- [mer05] MERLOT - Multimedia Educational Resource for Learning and Online Teaching. URL <<http://www.merlot.org/home/PeerReview.po>>, 2005.
- [MF99] Jessica Milstead and Susan Feldman. Metadata: Cataloging by any other name. Online Magazine, January 1999. Verfügbar im Internet: URL <<http://www.onlinemag.net/OL1999/milstead1.html>>.
- [ND02] Filip Neven and Erik Duval. Reusable Learning Objects: a Survey of LOM-Based Repositories. In MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia, pages 291–294. ACM Press, 2002.
- [NP02] Günter Neumann and Jakub Piskorski. A Shallow Text Processing Core Engine. Computational Intelligence, 18(3):451–476, 2002.
- [PS04] Rosanne J. Price and Graeme Shanks. A Semiotic Information Quality Framework. In Proceedings of the 2004 IFIP International Conference on Decision Support Systems (DSS2004): Decision Support in an Uncertain and Complex World, Prato, July 2004.
- [RW03] Ben Ryan and Steve Walmsley. Implementing metadata collection: a projects problems and solutions. Learning Technology newsletter, 5(1), January 2003. Learning Technology publication of IEEE Computer Society Learning Technology Task Force (LTF). Verfügbar im Internet: URL <[http://lutf.ieee.org/learn\\_tech/issues/january2003/index.html#3](http://lutf.ieee.org/learn_tech/issues/january2003/index.html#3)>.
- [Sca03] Vinod Scaria. Peer Review of Scholarly communication in health: perspectives in the Internet Age. Online Journal of Health and Allied Science (jhas), 2(1), January - March 2003. Verfügbar im Internet: URL <<http://virtualmed.netfirms.com/internethealth/proposal0103.html>>.
- [Sch02] Rolf Schulmeister. Taxonomie der Interaktivität von Multimedia - Ein Beitrag zur aktuellen Metadaten-Diskussion. it+ti - Informationstechnik und Technische Informatik, 4:193–199, 2002. Oldenbourg Verlag.
- [VHO02a] Virtuelle Hochschule Oberrhein: VIROR Local Knowledge Pool. Aufgaben des Fachreferenten, December 2002. Verfügbar im Internet: URL <[http://www.viror.de/entwickler/knowledge\\_pool/fachreferent/index.php](http://www.viror.de/entwickler/knowledge_pool/fachreferent/index.php)>.
- [VHO02b] Virtuelle Hochschule Oberrhein: VIROR Local Knowledge Pool. Validierung, December 2002. Verfügbar im Internet: URL <[http://www.viror.de/entwickler/knowledge\\_pool/validierung/index.php](http://www.viror.de/entwickler/knowledge_pool/validierung/index.php)>.
- [Wan98] Richard Y. Wang. A product perspective on total data quality management. Commun. ACM, 41(2): 58–65, 1998.
- [wwr05] WWR - Wissenswerkstatt Rechensysteme. URL <<http://wwr-project.de>>, 2005.