Deutsche Nationalbibliothek

nestor

INS INNOVATION MIT NORMEN UND STANDARDS

Bericht

DIN-Workshop: Ermittlung des Standardisierungsbedarfs in den Bereichen „Langzeitarchivierung“ und „Digitalisierung“

2. u. 3. Nov. 2006 in der Deutschen Nationalbibliothek

Frankfurt am Main

Version 0.9

Inhaltsverzeichnis

1. Einführung S. 3

2. Standardisierungsanforderungen für die Langzeitarchivierung und Digitalisierung S. 4

    2.1 Klassifikation der Thesenpapiere S. 6

    2.2 Handlungsempfehlungen des DIN S. 7

3. Ermittlung des Standardisierungsbedarfs S. 10

    3.1. Persistent Identifier S. 10

    3.2. Vertrauenswürdige Archive S. 14

    3.3. Metadaten zur Langzeitarchivierung und Objektformat S. 16

    3.4. Qualität digitalisierter Kulturgüter S. 21

    3.5. Dateiformate und Dokumentenmodelle von Digitalisaten S. 24

    3.6. Metadaten zur Digitalisierung S. 26

4. Workshop-Ergebnisse und Ausblick S. 30

5. Anhang: Teilnehmer, Programm und Thesenpapiere S. 35

1. Einführung

Die unmittelbare Notwendigkeit einer nationalen Standardisierungsaktivität auf den Gebieten der Langzeitarchivierung und Digitalisierung setzt neue Formen der Zusammenarbeit zwischen den großen Einrichtungen der Bibliotheken, Archive, Museen und weiteren Organisationen aus dem Umfeld der Fachwissenschaften und der Medien voraus.[1] Aus diesem Grund veranstalteten die *Deutsche Nationalbibliothek* (DNB), *nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland* und die *Innovation mit Normen und Standards* (INS)-*Initiative* des Bundesministeriums für Wirtschaft (BMWi) und des Deutschen Instituts für Normung e. V. (DIN) am 02./03. November 2006 in Frankfurt am Main einen Workshop, der sich dem Normierungs- bzw. Standardisierungsbedarf in den genannten Gebieten widmete.[2] Das Ziel der Veranstalter war es, dass die Diskussionen der Workshop-Teilnehmer zu möglichst konkreten Empfehlungen führten.

Die erzielten Workshop-Ergebnisse liefern eine geeignete Grundlage, den digitalen Herausforderungen der kommenden Jahre mit tragfähigen Konzepten und konsensgestützten Vereinbarungen zu begegnen. Die vorgeschlagenen Handlungsschritte und Maßnahmen der ca. 50 Fachleute aus dem deutschsprachigen Raum und den beiden Vertretern des DIN gelten gleichermaßen für die Langzeitarchivierung digitaler Ressourcen wie für die Bereiche der Digitalisierung und digitalen Sammlungen.[3] Auf der Basis des

---

[1] An dieser Stelle sei auf die Definition des Standards vom „British Standards Institute“ hingewiesen: „Ein Standard ist ein öffentlich zugängliches technisches Dokument, das unter Beteiligung aller interessierter Parteien entwickelt wird und deren Zustimmung findet. Der Standard beruht auf Ergebnissen aus Wissenschaft und Technik und zielt darauf ab, das Gemeinwohl zu fördern.“.

[2] Das Bundeswirtschaftsministerium finanziert im Rahmen der Innovationsförderung ein langfristig angelegtes Projekt "Innovation mit Normen und Standards". Ziel ist es, Innovationen der Zukunft optimale Rahmenbedingungen zu bieten und ihre Marktfähigkeit zu fördern. Die für 2006 vorgesehenen Maßnahmen dienen als Vorbereitung und Wegbereitung, um der Normung in Deutschland eine Innovationsstruktur zu geben. Dabei geht es auch darum, nachhaltige Strukturen in der Forschung über Standardisierung und Normung zu entwickeln sowie Normung und Standardisierung im Forschungsprozess selbst zu integrieren. Hierzu werden zunächst drei Forschungsprojekte vom DIN vergeben. Diese wissenschaftlichen Basisuntersuchungen werden durch 27 konkrete Normungsprojekte zu einzelnen Themen ergänzt.

[3] Eine Liste mit den Teilnehmern und den dazugehörigen Institutionen befindet sich im Anhang des vorliegenden Berichts. Die Vertreter des DIN und Mitarbeiter des INS waren Herr Dr. Ingo Wende und Herr Roman Grahle.

vorliegenden Standardisierungsbedarfs werden mit Unterstützung der INS-Initiative bis Ende des Jahres 2006 erste Vorhaben beschrieben und weitere Anstrengungen in Richtung konkreter Normierungsverfahren umgesetzt.[4] Die zweite Projektstufe des Normierungsprozesses befindet sich derzeit noch im Genehmigungsverfahren beim INS.

2. Standardisierungsanforderungen an die Langzeitarchivierung und Digitalisierung

Die Ermittlung des Standardisierungsbedarfs für die Langzeitarchivierung und Digitalisierung bedeutet sowohl für die beteiligten Bibliotheken, Archive und Museen als auch für die INS-Initiative des DIN das Betreten von Neuland. Das DIN hat aufgrund des dringenden Normierungsbedarfs die vorliegende Initiative aus rund 150 eingegangenen Anträgen ausgewählt und in eine Reihe von rund 30 weiteren Aktivitätsfeldern eingeordnet. Auf dem Gebiet der Langzeitarchivierung elektronischer Ressourcen ist der Bedarf an einheitlichen Standards besonders groß. Alle Gedächtnisorganisationen sind sich seit Jahren darüber einig, dass keine Institution allein alle digitalen Ressourcen dauerhaft archivieren kann, sondern dass geeignete nationale Kooperations- und Austauschmodelle greifen müssen. Eine entscheidende Rolle kommt der Verständigung auf gemeinsame Standards und Schnittstellen zu, um damit ein hohes Maß an Interoperabilität zu garantieren.

Wie bei der Langzeitarchivierung besteht auch im Falle der Digitalisierung ein akuter Standardisierungsbedarf. Zahlreiche Bibliotheken und Archive planen derzeit die Einrichtung von online-gestützten Benutzerportalen, um auf verteilte und umfangreiche digitale Bestände zugreifen zu können. Solche länderübergreifende bzw. europäische Internetportale werden den Schwerpunkt der Benutzung auf digitale Sammlungen entscheidend fördern. Darüber hinaus können die ca. 12.000 kleineren kulturellen Einrichtungen aus dem öffentlich-rechtlichen Bereich miteinander vernetzt werden. Der zentrale Nachweis digitaler Sammlungen wird einen entscheidenden Einfluss auf die Entwicklung der nationalen und

[4] Eine „Norm“ bezeichnet eine allseits rechtlich anerkannte und durch ein Normungsverfahren beschlossene, allgemeingültige sowie veröffentlichte Regel zur Lösung eines Sachverhaltes.

internationalen Wissensgesellschaft(en) ausüben und ist gerade deshalb seit einigen Jahren ein wichtiger Bestandteil deutscher und europäischer Förderpolitik. Ein solches Vorhaben funktioniert nur auf der Basis eines standardisierten Informations- und Datenaustausches; eine Entwicklung, die den Standardisierungsprozess weiter dynamisieren wird.

Unter dem beschrieben Eindruck stand am Beginn des Frankfurter Workshops die Frage, wie eine spezifisch deutsche Normungsstrategie auf den Gebieten der Langzeitarchivierung und Digitalisierung aussehen könnte. Die Forderung nach einer Intensivierung der Normierungsanstrengungen deckt sich mit dem politischen Appell der Bundesregierung, dass nationale Anwendungen im internationalen Vergleich konkurrenzfähiger werden sollen. Genauso hat umgekehrt die Europäische Union sämtliche Mitgliedsstaaten zum Handeln aufgerufen: Nur im Falle einer aktiven Mitgestaltung bei der Einführung der europäischen Normen können nationale Interessen eingebracht und formuliert werden. Andernfalls wird die Norm aus der Europäischen Union (CEN) zu 100 Prozent in das Regelwerk eines Mitgliedstaates überführt.[5] Eine Mitwirkung der beteiligten Anwender, der Forschung und der betroffenen Organisationen am europäischen Normungsprozess wird umso mehr empfohlen, wenn am Normgegenstand ein offensichtlich nationales Interesse besteht. Durch DIN-zertifizierte Verfahrensweisen können daher eigene Interessen an der Normierung noch stärker international geltend gemacht werden.[6]

---

[5] Die Europäische Normung wird im Rahmen der drei Organisationen CEN, CENELEC und ETSI durchgeführt. CEN bezeichnet sich als ein "System formaler Prozesse zur Herstellung von Normen, das durch die 29 nationalen Mitgliedsorganisationen getragen wird". Die nationalen Mitgliedsorganisationen stimmen über Europäische Normen ab und implementieren diese. Die Normungsorganisationen haben – ausgenommen ETSI – je Land nur ein Mitglied, das die gesamten Normungsinteressen dieses Landes zu vertreten hat. Bei Abstimmungen haben die Mitglieder entsprechend ihrer Wirtschaftskraft unterschiedliche Stimmen. Deutsche Interessen werden durch das DIN in Europa vertreten. Das Ziel der europäischen Normung ist die Harmonisierung der nationalen Normen in den Mitgliedsländern durch einheitliche Einführung von europäischen Normen.

[6] In Deutschland entscheiden die einzelnen Normenausschüsse des DIN über die Mitarbeit an einem europäischen Normungsvorhaben. Die fachliche Betreuung wird einem Arbeitsausschuss, einem so genannten Spiegelausschuss zugewiesen. Dieser hat eine deutsche Meinungsbildung durchzuführen und sie im europäischen Gremium zu vertreten. Dies kann durch schriftliche Kommentare, Entsendung von Delegationen und/oder Benennung von Experten geschehen. Bei Vorliegen eines Normungsantrages in Deutschland hat das DIN zu überprüfen, ob zu diesem Thema ein europäischer Normungsprozess besteht bzw. der Normungsvorschlag für diese Ebene in Betracht zu ziehen ist.

Der Normungsprozess ist in der DIN 820-4 „Normungsarbeit – Geschäftsgang“ detailliert geregelt: Während das DIN die Rahmenbedingungen und die Austauschplattform bzw. -infrastruktur zur Verfügung stellt, sind die beteiligten Einrichtungen für die konkrete Initiierung eines entsprechenden Verfahrens zuständig. Das Einleiten von Normungsarbeiten kann von jedem Interessierten beantragt werden, indem ein begründeter Normungsantrag gestellt wird. Der zuständige Normenausschuss entscheidet über dessen Annahme oder Ablehnung. Vor der Entscheidung ist zu klären, ob für den Normungsgegenstand ein Bedarf besteht bzw. zu erwarten ist, und ob die interessierten Kreise bereit sind, am Normungsvorhaben mitzuarbeiten und zur Finanzierung beizutragen. Angenommene Normungsanträge werden im „DIN-Anzeiger für technische Regeln“ veröffentlicht. Im Ausschuss treffen die interessierten Kreise aufeinander, wobei die Anzahl der Experten nicht höher als 21 sein soll, und erarbeiten auf Basis einer Normvorlage einen Norm-Entwurf. Dieses Dokument entsteht im Konsens und wird anschließend veröffentlicht. Die Öffentlichkeit sowie Experten, die nicht persönlich am Normungsprozess beteiligt waren, haben vier Monate Zeit, um zum Norm-Entwurf Stellung zu beziehen. In weiteren drei Monaten berät der Arbeitsausschuss die Stellungnahmen. Ein Schlichtungsverfahren regelt strittige Fälle. Nach der Verabschiedung der endgültigen Fassung durch den Ausschuss und nach der Prüfung durch das DIN wird das Ergebnis als DIN-Norm veröffentlicht.

Am Ende sollten unter den beschriebenen Voraussetzungen nach Angaben der INS vier Kriterien in beschriebenen Bereichen der Langzeitarchivierung und Digitalisierung erfüllt werden:

1. Die beteiligten Institutionen nehmen den Normungsprozess selbst vor.
2. Die Ergebnisse von Forschung und Entwicklung mit Innovationspotenzial werden aufgegriffen und im Hinblick auf Normungsfähigkeit überprüft.
3. Die Anforderungen (markt/-anwendungsbezogen) an die Normung werden identifiziert und formuliert.
4. Es werden die deutschen Lösungen in die internationalen und europäischen Normen eingebracht (Förderung des Wirtschaftsstandorts Deutschland).

2.1. Klassifikation der Thesenpapiere

Die DNB hat im Rahmen ihrer Veranstaltungsorganisation von den ausgewählten Experten kurze Thesenpapiere erbeten, die vorab an alle Teilnehmer verschickt wurden. Auf dem Workshop präsentierte das DIN eine erste informelle Einteilung der eingereichten Papiere, um den jeweiligen potentiellen Handlungsbedarfs darzustellen. Die einzelnen Elemente bildeten: Gesamtprozess der Langzeitarchivierung und Digitalisierung, Dokumentidentifikation, Vertrauenswürdigkeit, Interoperabilität und langfristige Reproduzierbarkeit. Mit der vorliegenden Klassifikation sollte ausdrücklich keine inhaltliche Bewertung, sondern lediglich eine formale Aufteilung vorgenommen werden.

Nr.	Referent	Einrichtung	Titel	Klassifikation
1	Wolf-Klostermann	BSB	Digitalisierung an der BSB	- Gesamtprozess, allgemeine Anforderungen
2	Frodl	DNB	Digitalisierung und digitale Sammlungen	- Gesamtprozess, allgemeine Anforderungen
3	Wolf	BSZ-Koblenz	Digitalisierung und digitale Sammlungen	- Gesamtprozess, allgemeine Anforderungen
4	Wolf-Klostermann	BSB	Dokumente und Dokumentenmodelle	- Gesamtprozess, allgemeine Anforderungen - Effizienz und Qualität (Skalierbarkeit) der Digitalisierung
5	Flores-Herr	DNB	Güte der Digitalisierung von Bildern und Drucken	- Effizienz und Qualität (Skalierbarkeit) der Digitalisierung
6	Stockmann	SUB-Göttingen	Qualität digitalisierter Kulturgüter	- Effizienz und Qualität (Skalierbarkeit) der Digitalisierung
7	Wolf-Klostermann	BSB	Qualität digitalisierter Kulturgüter	- Effizienz und Qualität (Skalierbarkeit) der Digitalisierung
8	Enders	SUB-Göttingen	Persistent Identifier	- Dokumentidentifikation
9	Brase	TIB-Hannover	Einführung zum Stand DOI in Deutschland	- Dokumentidentifikation
10	Korb	DNB	Uniform Ressource Name (URN)	- Dokumentidentifikation
11	Brandt	SUB-Göttingen	METS – Metadata Encoding and Transmission Standard	- Metadaten
12	Neuroth	SUB-Göttingen	Metadatenformate	- Metadaten

13	Steinke	DNB	Universelles Objektformat und Langzeitarchivierungsmetadaten für elektronische Ressourcen	- Metadaten
14	Wolf-Klostermann	BSB	Metadaten zur Langzeitarchivierung	- Metadaten
15	Enders	SUB-Göttingen	Dokumentenmodelle	- Metadaten - Gesamtprozess, allgemeine Anforderungen
16	Dobratz	HUB	Zertifizierung vertrauenswürdiger Archive	- Vertrauenswürdigkeit
17	Menne-Haritz	Bundesarchiv	Präsentation von Digitalisaten aus Archivgut	- Langfristige Reproduzierbarkeit der digitalen Ressourcen

2.2. Handlungsempfehlung des DIN

Das DIN sprach für die Ermittlung des Standardisierungsbedarfs in den Bereichen Langzeitarchivierung und Digitalisierung folgende Empfehlungen aus:

Die DIN-Vertreter stellten grundsätzlich für alle vorgestellten Themenbereiche eine Übernahme als DIN-Norm und im Einzelfall eine internationale Norm in Aussicht. Im Einzelnen müssten die drei Beiträge im „Gesamtprozess, allgemeine Anforderungen“ noch stärker präzisiert werden. Zudem sollte im Rahmen des Klassifikationspunktes „Effizienz und Qualität (Skalierbarkeit) der Digitalisierung“ darüber nachgedacht werden, eine Metrik, eine Definition von Klassen und anschließend - gegebenenfalls - ein Zertifizierungsverfahren zu entwickeln. Für die drei Beiträge in „Dokumentidentifikation“ empfahlen die DIN-Vertreter eine stärkere Konsensbildung. Im Rahmen der „vertrauenswürdigen Archive“ müssten noch weitere Anforderungen und Lösungsmechanismen formuliert und sich intensiver mit den internationalen Normungsprozessen auseinandergesetzt werden. Die deutschen Positionen sollten in Form eines nationalen Spiegelgremiums eingebracht werden. Des Weiteren sprachen sich die DIN-Vertreter dafür aus, den Bereich „Interoperabilität“ in den Normierungsverfahren noch stärker zu berücksichtigen. Im Zuge der Beurteilung des Beitrags in „langfristige Reproduzierbarkeit der digitalen Ressourcen“ wurde angeregt, weitere Anforderungen zu definieren und Lösungsmechanismen bereitzustellen. In diesem Themenfeld müssten zudem weitere internationale Standards einbezogen werden. Auf der Basis

dieser Handlungsempfehlungen beschrieben die DIN-Vertreter den weiteren Aufgaben- und Fragenkatalog für die Workshop-Teilnehmer folgendermaßen:

- Für welche Aspekte der Digitalisierung und Langzeitarchivierung von Bibliotheksbeständen sieht Deutschland Handlungsbedarf?
- Was sind die deutschen Lösungsvorschläge und Ziele?
- Welche nationalen Spiegelgremien sind zu bilden bzw. zu stärken?
- In welchen Bereichen will sich Deutschland mit eigenen Lösungsvorschlägen und Zielen aktiv in der internationalen und/oder europäischen Standardisierung beteiligen?
- In welchen internationalen/europäischen Gremien soll die Mitarbeit erfolgen?
- Wie wird die notwendige Mitarbeit der Experten gewährleistet?
- Wie erfolgt die Finanzierung nach Auslaufen der INS-Förderung?

Für den Abschluss des Normierungsverfahrens stehen folgende Publikationsalternativen zur Auswahl:

1. Eine **DIN-Norm**. Eine DIN-Norm ist ein Dokument, das durch eine breite Beteiligung aller interessierten Kreise im Konsens erarbeitet wird. Alle Instanzen eines Normungsverfahrens werden durchlaufen. Voraussetzung für eine Norm ist, dass sie technisch ausgereift ist und einen Nutzen für den Anwender hat.
2. Ein **DIN-Fachbericht**. Der Fachbericht setzt weder einen Konsens noch eine präzise Festlegung der Sachverhalte voraus. Er ist nicht das Ergebnis eines normativen Prozesses, vielmehr werden darin Ergebnisse und Angaben als Vorstufe einer angestrebten Norm gesammelt.
3. Die **PAS** (Publicy Available Specification). Das PAS-Verfahren dient der schnellen Veröffentlichung von Konsortialergebnissen und findet häufig im Rahmen innovativer Vorhaben Anwendung. Wie der Fachbericht ist die PAS ebenfalls nicht konsensabhängig, kann aber der Ausgangspunkt für Normung sein.[7]

[7] Die Zeit bis zur Veröffentlichung einer PAS beträgt nur sechs bis acht Wochen ab dem Einreichen des PAS-Textes beim DIN. Das DIN ist für die Veröffentlichung, nicht aber für den Inhalt einer PAS verantwortlich. Die Auflagenhöhe wird vom Verfasser selbst bestimmt und der Text kann vom Verfasser selbstständig ausgearbeitet und beim DIN eingereicht werden. Die PAS wird vom DIN nach fünf Jahren überprüft.

3. Ermittlung des Standardisierungsbedarfs

In diesem Abschnitt werden erstens die einzelnen Themenbereiche des Workshops vorgestellt, zweitens die Kernsätze aus den eingereichten Thesenpapieren in verkürzter Form wiedergegeben und drittens der Diskussionsverlauf des Frankfurter Workshops ausführlich beschrieben.[8]

3.1. Persistent Identifier

Einführung:

Der Themenblock der persistenten Identifikatoren bzw. Persistent Identifier (PI) bildete den Auftakt.[9] Auf dem Workshop setzte man sich mit der Frage auseinander, welche Konsequenzen sich aus dem zunehmenden Bedarf für internationale bzw. generell übergreifende Resolving-Infrastrukturen ergeben. Da mit den URLs die Ressource der Netzpublikationen identifiziert wird, nicht aber die Netzpublikation selbst, und der langfristige Bestand einer URL im Allgemeinen nicht sichergestellt werden kann, stellt sich die Herausforderung, die Netzpublikation selbst (eindeutig und dauerhaft) zu identifizieren. PI bezeichnet eine Methode zur eindeutigen und persistenten Adressierung der im Internet auffindbaren Objekten oder Quellen und ist damit ein wichtiger Baustein für die langfristige Verfügbarkeit digitaler Ressourcen. In Deutschland werden derzeit verschiedene Arten von PIs parallel genutzt. Während der kommerzielle Bereich vor allem den „Digital Object Identifier“ (DOI), eine Anwendung des Handle-Systems, benutzt, verwendet der überwiegende Teil des deutschen universitären bzw. öffentlichen Sektors den URN-Service der Deutschen Nationalbibliothek. Die DOI-Foundation unternimmt mittlerweile Anstrengungen, die DOI als ISO-Standard zu etablieren.

Kernsätze der Vorträge:

Markus Enders (SUB-Göttingen) führte in die organisatorischen Rahmenbedingungen der PI ein. Er plädierte dafür, einen Resolver zu

[8] Die einzelnen Thesenpapiere mit den jeweiligen ausführlichen technischen Beschreibungen und den vorgeschlagenen Standardisierungsaktivitäten in den genannten Bereichen befinden sich im Anhang.

[9] Zur Definition siehe: http://www.persistent-identifier.de/.

entwickeln, der über dem PI-System angesiedelt ist und der die Weiterleitung an die jeweils zuständigen Resolver über den PI-Typ als „namespace“ gewährleistet. In diesem Zusammenhang präsentierte Enders den N2T-Resolver aus der amerikanischen „Name-to-Thing-Initiative“. Er sei besonders geeignet, um unabhängig des PI-Systems den Zugriff auf Dokumente zu ermöglichen.

Im Anschluss daran berichtete Nikola Korb (DNB) über den URN-Einsatz in Deutschland. Die Referentin hob die zentrale Bedeutung der URN für die Langzeitarchivierung elektronischer Ressourcen hervor, weil die URN's in besonderer Weise geeignet seien, URLs in Nachweissystemen zu ersetzen und als stabile Referenzen in den Objekten selbst angewendet zu werden. Die Verwaltung der URN obliegt der Deutschen Nationalbibliothek.

Im letzten Beitrag dieses Themenblocks stellte Jan Brase (Technische Informationsbibliothek der Universität Hannover) den DOI-Einsatz in Deutschland vor. Danach sei im DFG-Projekt "Publikation und Zitierfähigkeit wissenschaftlicher Primärdaten" eine Infrastruktur zur Registrierung und Verwaltung von DOIs für wissenschaftliche Datensätze geschaffen und erfolgreich getestet worden. Des Weiteren werde an der TIB eine Non-profit DOI-Registrierungs-Agentur eingerichtet, die zum Selbstkostenpreis DOIs als stabile Referenzen für Primärdaten und andere Wissenschaftliche Inhalte aus Öffentlich geförderter Forschung vergeben soll.

Diskussion:

Nach der Ansicht von Ingo Wende (DIN) spiegeln die Vorträge der ersten Sektion die ganze Vielfalt der deutschen Diskussion um die PI wider. Es müssten sämtliche Möglichkeiten ausgeschöpft und sich hinsichtlich einer eventuellen Normierung für den besten Weg der drei vorgestellten PI-Systeme entschieden werden. Thomas Stäcker (Herzog August Bibliothek Wolfenbüttel) entgegnete, dass es für einen institutionalisierten PI keine Garantie für ein Funktionieren nach 50 oder 100 Jahren gebe. Daher sollte nicht lediglich auf eine einzige Strategie für ein PI-Verfahren gesetzt werden, sondern je nach Community das zugehörige PI-System zugelassen werden. Schließlich befinden sich DOI und URN bereits erfolgreich im Einsatz. Ähnliche Probleme gelten nach Stäcker auch für die Resolving-Server. Es sei dafür zu sorgen, dass der Resolver bei derjenigen Institution mit der größtmöglichsten Persistenz angesiedelt wird. Matthias Hemmje (Fernuniversität Hagen) führte

aus, dass die vorgestellten PI-Modelle sehr stark von der jeweiligen Community abhängig sind, wobei eine Reihe weiterer Communities, wie beispielsweise der ganze multimediale Bereich, bisher noch unerwähnt geblieben sind. Zudem gebe es im internationalen Vergleich weit mehr PI-Modelle als in Deutschland. Nach der Meinung Hemmjes liegt der Standardisierungsansatz deshalb im Mapping zwischen den unterschiedlichen PI-Systemen.

Reinhard Altenhöner (DNB) betonte in seinen Ausführungen den jeweiligen Nutzungszusammenhang des PI. Während DOI von den Verlagen verwendet wird, sei die URN vor allem in den Bibliotheken weit verbreitet. Aus diesem Grund müssten auch beide Anforderungen hinsichtlich einer Standardisierungsaktivität gleichwertig berücksichtigt werden. Jens Klump (Geoforschungszentrum Potsdam) ergänzte, dass Forschungsinstitutionen noch ganz andere Interessen als Gedächtnisorganisationen haben. Abgesehen davon sieht Klump einen bisher unberücksichtigten Problembereich für einen standardisierten Einsatz der PI: Eine Normierungsdiskussion könne nicht losgelöst vom weiteren Verlauf des „Kriterienkatalogs für vertrauenswürdige Archive“ geführt werden. Der Druck, der in Zukunft von den zertifizierten vertrauenswürdigen Archiven auf die Gedächtnisorganisationen ausgehe, werde in gleicher Weise den Druck auf die Geldgeber nach der Implementierung einheitlicher Repositories erhöhen und sich in der Folge auch unmittelbar auf die Weiterentwicklung der PI auswirken. Demnach sei es zum jetzigen Zeitpunkt für ein Standardisierungsverfahren noch zu früh, wie Klump (Geoforschungszentrum Potsdam) weiter ausführte.

Heike Neuroth (SUB-Göttingen) kam zu einem differenzierten Urteil. Sie spricht sich für eine variable Handhabung des PIs aus, um auf dem internationalen Markt mit allen potenziellen Institutionen zusammenzuarbeiten. Man müsse prinzipiell offen mit den in der Welt befindlichen PI sein, um sich keine Chancen zu verbauen. Aus dem gleichen Grund befürwortete Markus Enders (SUB-Göttingen) die Abkopplung des PI von den „vertrauenswürdigen Archiven“. Ein PI sollte von den beteiligten Institutionen und Repositories unabhängig eingesetzt werden und ein Objekt identifizieren. Diese Sichtweise bestätigte auch Tobias Steinke (DNB), der für eine übergeordnete (Meta-)Ebene oberhalb der eingesetzten IT-Systeme mit generellen Kriterien plädierte. Auf einer solchen könne man die URN und DOI einer Normung zuführen. Markus Enders (SUB-Göttingen) präzisierte daraufhin, dass sich der vorgestellte N2T-Resolver derzeit noch im Projektstatus befindet. Wichtig sei

daher, dass sich die einzelnen Projektpartner noch intensiver an der nationalen und internationalen Debatte beteiligten. Wie seine Vorredner hob auch Jan Brase (Technische Informationsbibliothek der Universität Hannover) hervor, DOI und URN gleichermaßen einzusetzen. Er forderte die Browserhersteller auf, die jeweilige DOI und URN als gültige Adressen zuzulassen. Nikola Korb (DNB) ergänzte, dass beide PI-Systeme im Browser auflösbar sein sollten.

Im Folgenden führte Matthias Hemmje (Fernuniversität Hagen) aus, dass im Kontext des PI eine Metaebene semantisch beschrieben werden müsse, weil sie grundsätzlich verschiedene Elemente in sich integriere. Im Zusammenhang des Dokuments sei der Aspekt der Authentizität nicht zu unterschätzen. Susanne Dobratz (Humboldt-Universität Berlin) wies darauf hin, dass eine vorgeschlagene Metaebene auch hinsichtlich der Langzeitarchivierung von großer Bedeutung sei. Sie ermunterte deshalb die SUB Göttingen, den eingeschlagenen Weg von Enders fortzuführen und stellte nestor als geeignete Plattform für die weitere inhaltliche Arbeit in Aussicht. Tobias Steinke (DNB) bekräftigte im weiteren Verlauf die Forderung, DOI und URN für ein Normierungsverfahren beim DIN vorzuschlagen. Gerade im DOI-Bereich läge ein internationaler Vorstoß vor, den man national unterfüttern müsse. Es besteht nach Ansicht Steinkes ein großer Bedarf, sich in Deutschland weiter zu engagieren. Reinhard Altenhöner (DNB) ergänzte, dass die ISO-Aktivitäten offen gelegt und in die Communities hineingetragen werden müssten. Darüber hinaus sprach sich Jan Brase (Technische Informationsbibliothek der Universität Hannover) für die Einbeziehung der IDF (internationaler DOI-Found) hinsichtlich des PI aus. Außerdem befürwortete er, sowohl DOI als auch URN einer Normierung zuzuführen und nicht von einer Standardisierungsaktivität beider Systeme auszugehen.

Der Moderator Reinhard Altenhöner (DNB) fasste die Diskussion am Ende der Sektion zusammen. Hierbei bezeichnete er den N2T-Ansatz von Enders (SUB-Göttingen) auch wegen des Service- und Betriebsaspektes als einen interessanten und richtigen Ansatz. Der Moderator ermunterte das Plenum, die Initiative hinsichtlich eines angestrebten Normierungsverfahrens zu unterstützen. Auf die Frage, inwieweit man die Browser für die Auflösung von URN und DOI direkt einsetzen kann, äußerte sich Altenhöner skeptisch. Es gelte, die derzeitigen Marktmechanismen zu berücksichtigen und vorhandene

Normanstrengungen zu unterfüttern. Im Falle der URN sollten die Bemühungen auch in Richtung eines ISO-Standards weitergeführt werden.

Ingo Wende (DIN) erkannte in der zurückliegenden Diskussion einen sehr weit fortgeschrittenen Entwicklungsstand. Er empfahl deshalb, sich noch einmal grundsätzlich die Frage nach den Anforderungen eines PI zu stellen. Erst dann könne man auch die Fragen beantworten, welche PI-Systeme zu bevorzugen seien. Darüber hinaus müsse der jeweilige Nutzungszusammenhang des PI geklärt werden. Wende regte an, den Auftrag für ein Standardisierungsverfahren einer einzigen Organisation zu überlassen. Zum Schluss erinnerte Ingo Wende noch einmal daran, dass es keinen Zwang für eine Normierung gibt. Bei der DIN bestünden viele Möglichkeiten auf dem Weg der Normierung. Eine evtl. Norm dürfe daher nicht zu eng auf die Definition der Objekte ausgelegt werden.

3.2. Vertrauenswürdige Archive

Einführung:

Im Mittelpunkt der zweiten Sektion stand der „Kriterienkatalog für vertrauenswürdige Archive“ von nestor.[10] Die Workshop-Teilnehmer fragten nach gemeinsamen Qualitätsmaßstäben für eine sichere Langzeitarchivierung digitaler Ressourcen und nach den Möglichkeiten einer Standardisierungsaktivität. Neben dem Kriterienkatalog von nestor wurden die international verschiedenen Aktivitäten der Repositories diskutiert.

Kernsätze der Vorträge:

Susanne Dobratz (Humboldt Universität zu Berlin) führte mit einem Sachstandsbericht in das Thema ein. Für die nestor-AG vertrauenswürdige Archive gehe es nun darum, noch stärker auf eine internationale Übereinstimmung (Common Principles) zu setzen. Darüber hinaus sollten die Kriterien des Katalogs für eine Standardisierung überarbeitet und eine Metrik für den Kriterienkatalog entwickelt werden. Im Anschluss stellte Siegfried Hackel

[10] Siehe: http://nestor.cms.hu-berlin.de/tiki/tiki-index.php?page=ag-repositories.

(Physikalisch-Technische Bundesanstalt Braunschweig) das Projekt Archisafe vor.[11]

Diskussion:

In seiner Eingangsbemerkung warf der Moderator Frank M. Bischoff (Archivschule Marburg) zwei Fragen auf, die seiner Meinung nach im Zentrum der weiteren Diskussion um die vertrauenswürdigen Archive stehen: Sollte einerseits die weitere Arbeit der nestor-AG in eine Norm münden oder sei ein Fachbericht zu bevorzugen, der die nationalen Besonderheiten beinhalte, und soll sich andererseits Deutschland an einer internationalen Diskussion beteiligen oder nicht? Ingo Wende (DIN) wiederholte daraufhin seine Forderung, dass zunächst die genauen Anforderungen für eine Standardisierung genannt und die Folgekosten für eine normierte Sicherheit einkalkuliert werden sollten. Der hohe Aufwand einer Standardisierung müsse mit dem beigemessenen Wert für Sicherheit übereinstimmen. Wende empfahl darüber nachzudenken, welche Institution für ein Zertifizierungsverfahren für vertrauenswürdige Archive in Frage kommt. Eine Zertifizierung könne entweder eine unabhängige Einrichtung oder ein Hersteller eines digitalen Archivs ausstellen. Susanne Dobratz (Humboldt-Universität Berlin) erwiderte daraufhin, dass der Kriterienkatalog im Ergebnis keine absolute „Archivsicherheit“ garantiere, sondern wegen seiner abstrakten Ebene darüber hinausgehe. Die Regeln für Vertrauenswürdigkeit müssten in eine internationale Diskussion eingebettet und dann den nationalen Anforderungen angepasst werden. Udo Schäfer (Staatsarchiv Hamburg) hob ebenfalls das hohe abstrakte Niveau des Kriterienkatalogs als Anforderungsprofil für sämtliche Gedächtnisorganisationen hervor. Der Katalog sei nicht lediglich ein Konzept für die IT-Sicherheit, vielmehr ersetzte er aus archivischer Sicht die Nutzung des Signaturverfahrens. Darüber hinaus empfahl Schäfer eine Abstimmung einer eventuellen DIN-Norm mit der ISO-Norm. Tobias Steinke (DNB) stellte darüber hinaus fest, dass der Kriterienkatalog bereits der Schlusspunkt einer längeren Fachdebatte sei. Es gehe nun darum, genauere Anforderungen für bestimmte Bereiche abzugrenzen. Auch wenn man den Fokus des Kriterienkatalogs gegebenenfalls einschränken müsse, könne man ihn schon bald einer DIN-Norm zuführen. Darauf hin fragte Reinhard Altenhöner (DNB), was

[11] Zum Zeitpunkt der Erstellung des Berichts lag ein Thesenpapier nicht vor. Siehe nähere Informationen zu Archisafe in: http://nestor.sub.uni-goettingen.de/allg/detail.php?Show=689.

den Kern der vertrauenswürdigen Archive ausmache und ob es Sinn mache, in Richtung Qualitätsmanagement ISO 9000 zu gehen. Da es um Organisation und Betrieb geht, bekräftigte Matthias Hemmje (Fernuniversität Hagen) ein Vorgehen in Richtung ISO 9000. Des Weiteren kommen auch die beiden ISO-Standards Records Management ISO 15489:2001 und der elektronischen Archivierung ISO 14721:2003 in Frage.

In der weiteren Diskussion legte Ingo Wende (DIN) seinen Blick noch einmal auf die technischen Anforderungen für eine eventuelle Normierung des Kriterienkatalogs. Ein Vorgehen in Richtung ISO 9000 sei grundsätzlich immer hilfreich, weil das zu einer genauen Dokumentation der einzelnen Prozesse führe. Da Vertrauenswürdigkeit bereits beim Digitalisierungsprozess anfange, müssten alle bestehenden nationalen und internationalen Diskussionen berücksichtigt werden. Nach Ansicht Wendes sei die internationale Ebene die richtige Plattform, weil hier einige Standards bereits vorliegen. Günter Mühlberger (Universität Innsbruck) warnte dagegen vor zu großen Erwartungen hinsichtlich der internationalen Standards. Es wäre oftmals schon wichtig, wenn ein nationaler Standard vorläge. Ingo Wende (DIN) erwiderte darauf hin, dass man durchaus ein paralleles nationales und internationales Normierungsverfahren durchführen könne. Veränderungen müssten aber jeweils angepasst werden. Heike Neuroth (SUB-Göttingen) warnte in ihrem Redebeitrag vor unnötigem Zeitverlust bei einer Normierung. Die amerikanischen und englischen Initiativen der Repositories stünden bereits schon kurz vor Abschluss. Deutschland müsse in Richtung Normierungsprozess der vertrauenswürdigen Archive schnell vorankommen, um keine weiteren „Marktanteile“ zu verlieren.

Am Ende fasste der Moderator Frank M. Bischoff (Archivschule Marburg) die Diskussion zusammen: Die Anforderungen der vertrauenswürdigen Archive seien auf eine konkrete Ebene herunterzubrechen. Aus diesem Grund empfahl er unter Mitwirkung der internationalen Experten die Einberufung eines „runden Tisches“. Die nationale Ebene sei zwar für die Meinungsbildung wichtig, jedoch sollten Ergebnisse in die internationalen Gremien eingebettet werden. Für die internationalen Standards bilde die DIN-Normierung eine Vorarbeit, die im weiteren Verlauf auch in eine ISO-Zertifizierung münden müsse.

3.3. Metadaten zur Langzeitarchivierung und Objektformat

Einführung:

Die letzte Sektion des ersten Workshop-Tages widmete sich den Bereichen der Metadaten zur Langzeitarchivierung und Objektformat. Die Teilnehmer stellten fest, dass zu einem Archivobjekt mit der Perspektive einer langfristigen Verfügbarkeit neben den Dateien selbst auch technische und inhaltliche Metadaten gehören, die bei der Archivierung selbst aber auch beim Austausch zwischen Archivsystemen berücksichtigt werden sollten. Im BMBF-geförderten Projekt kopal entstand zu diesem Zweck das Paketformat Universelles Objektformat (UOF) das auf METS und LMER basiert, während in anderen Projekten PREMIS eingesetzt wird.

Kernsätze der Vorträge:

Zu Beginn stellte Tobias Steinke (DNB) UOF und LMER aus dem Projekt kopal vor.[12] Gemäß den Angaben sei die Gewinnung von Metadaten zur Langzeitarchivierung eine wichtige Schlüsselkomponente für die spätere Durchführung von Formatmigrationen zur Langzeitverfügbarkeit. Dazu gehören technische Informationen zu den einzelnen Dateien, aber auch eine Dokumentation der erfolgten Änderungen der Dateiformate. Ein Archivobjekt sollte zum Zeitpunkt der Archivierung diesen „Beipackzettel“ haben und auch beim Austausch mit anderen Langzeitarchiven mitführen. Steinke regte an, LMER 1.2 als genormter Kern von Metadaten zur Langzeitarchivierung und UOF als Basis für den genormten Objektaustausch zwischen Langzeitarchiven verbreitet zu verwenden.

Anschließend berichtete Olaf Brandt (SUB-Göttingen) über die Verwendung von METS und PREMIS im Projekt MathArc. Demnach können mit PREMIS digitale Objekte, Ereignisse, Rechte, „Agenten“ und ihre Verknüpfungen beschrieben werden. Zudem sei es möglich, die für die langfristige Verfügbarkeit wichtigen Kontextinformationen, die mit digitalen Objekten verbunden sind, zu beschreiben. Das MathARC-Projekt hat den Austausch von Daten zwischen zwei Langzeitarchiven zum Ziel. Die Implementierungen von uneinheitlichen Metadatensets (LMER und PREMIS) sei nach den Aussagen Brandts sehr schwierig, weshalb er den Austausch eines einheitlichen Datenformats auf der Basis von PREMIS anregte.

[12] Siehe: http://kopal.langzeitarchivierung.de.

Im Anschluss an diesen Bericht gab Thomas Wolf-Klostermann (Bayerische Staatsbibliothek München) einen Überblick über weitere Beispiele und Kriterien für die Verwendung von Metadaten bei der Langzeitarchivierung.

Diskussion:

Die Moderatorin Heike Neuroth (SUB-Göttingen) stellte folgende Leitfragen zur Diskussion: Welche Konsequenzen ergeben sich für die Gedächtnisorganisationen aus den unterschiedlichen nationalen und internationalen Metadaten und welche konkrete Maßnahmen müssten in Richtung Normierungsbestrebungen eingeleitet werden?

Thomas Stäcker (Herzog August Bibliothek Wolfenbüttel) wies auf die zurückhaltende Position Deutschlands in der internationalen Diskussion um die Metadaten hin. Außerdem sei der richtige Zeitpunkt für eine Normierung nur schwierig zu bestimmen. Im Bereich METS gebe es beispielsweise noch keine langjährigen Erfahrungswerte. Die Experten bewerten das Containerformat nach Stäcker sowohl als flexibel als auch als sehr komplex, weshalb es für eine Standardisierung von METS noch zu früh sei. Die Konsequenz sei nach der Ansicht von Tobias Steinke (DNB) aber ein langes Abwarten, bis sich eine Richtung am Markt durchgesetzt habe. Am Ende könnten daher noch mehr unterschiedliche Formate als heute vorliegen. Ein notwendiges Mapping zwischen den verschiedenen Metadaten sei nach Steinke immer ein schwieriges Unternehmen. Steinke hob stattdessen die Vorteile einer strukturellen Offenheit des UOF hervor. Markus Enders (SUB-Göttingen) befürwortete die Sichtweise Steinkes. Aufgrund des unterschiedlichen Blickwinkels müsse zunächst eine gemeinsame Basis geschaffen werden. Siegfried Hackel (Phys.-Techn. Bundesanstalt, Braunschweig) sprach sich für ein möglichst schnelles Handeln aus. Notfalls solle man auch Transformationsmechanismen weiterentwickeln, wie Hackel weiter ausführte. Karsten Huth (Bundesarchiv) ging in seinem Redebeitrag auf die Praxis des Bundesarchivs ein. Er empfahl, neben der Normierung von Langzeitarchivierungsmetadaten auch deren Wertemengen zu normieren. Dafür sollte möglichst schnell ein Regelwerk geschaffen werden. Bei einer eventuellen Einführung eines solchen Systems seien internationale Anwender einzubeziehen.

Reinhard Altenhöner (DNB) lenkte noch einmal den Blick auf den gemeinsamen Blickwinkel. Danach dürften im Ergebnis keine neuen Metadaten

im Sinne eines Drittsystems dazukommen. Markus Enders (SUB-Göttingen) wies darauf hin, dass man noch eine geeignete Software entwickeln müsse, um den gemeinsamen Blickwinkel technisch zu implementieren. Zudem forderte Enders ebenfalls zum sofortigen Handeln auf, weil De-facto-Standards erst durch konkrete Implementierungen geschaffen werden.[13] Mit der Umsetzung sollte eine Arbeitsgruppe bald beginnen. Für Frank M. Bischoff (Archivschule Marburg) müsse man in der Diskussion um die Metadaten auch die spezifische Sichtweise der Archive berücksichtigen. Hierbei seien die prozesshaften Metadaten gerade hinsichtlich einer langfristigen Implementierung von großer Bedeutung. Nach der Meinung Bischoffs müsse man zunächst die genauen Einsatzbereiche der Metadaten identifizieren. Karsten Huth (Bundesarchiv) stellte in diesem Zusammenhang die Verwendungspraxis von PREMIS im ‚Archival Information Package' (AIP) vor, die von einer Arbeitsgruppe des Bundesarchivs entwickelt wurde. Auf diese Vorarbeiten könnte im Rahmen des Normierungsverfahrens aufgebaut werden. Nach Ansicht Huths sei noch zu prüfen, wie man LMER im AIP einbeziehe. Im Anschluss daran machte Tobias Steinke (DNB) auf die notwendige Unterscheidung der inhaltlichen von den technischen Metadaten aufmerksam. Inhaltliche Metadaten seien für die Langzeitarchivierung im Vergleich von geringerer Bedeutung. Dieser Ansicht widersprach Torsten Schaßan (Foto Marburg), der für eine großen Spielraum bei der Unterscheidung zwischen den inhaltlichen und technischen Metadaten plädierte. Im Hinblick der Standards sei in erster Linie das Mapping zwischen LMER und PREMIS herzustellen.

Für den Rundfunkbereich sieht Dietrich Sauter (Institut für Rundfunktechnik GmbH, München) große Probleme, Metadaten zu standardisieren. Eine Standardisierungsaktivität im Rundfunkbereich sei aber gerade wegen der Metadatenvielfalt unbedingt notwendig. Nach Ansicht von Tobias Steinke (DNB) zeige der Redebeitrag Sauters, dass nicht alle Communities in die Diskussion mit einbezogen werden könnten. Man müsse vielmehr Standards in einigen Teilbereichen festsetzen. Die Diskussion um LMER und PREMIS belege doch gerade die starke Fokussierung der Thematik auf die Belange der Bibliotheken. Darauf hin warnte Matthias Hemmje (Fernuniversität Hagen), dass man übergreifende Bereiche trotz der spezifischen Problembereiche nicht aus den Augen verlieren dürfe. Die Arbeitsprozesse könnten nicht völlig

[13] Setzen sich Methoden oder Regeln nicht infolge von Vereinbarungen, Gesetzen, Verordnungen o. Ä., sondern in der Praxis durch und etablieren sich auf diesem Wege als Standards, spricht man auch von De-facto- oder Quasi-Standards.

von den Produktionsprozessen abgekoppelt werden. Olaf Brandt (SUB-Göttingen) betonte noch einmal, dass nicht längst alle vorliegenden Metadaten für die Langzeitarchivierung berücksichtigt seien.

Siegfried Hackel (Physikalisch-Technische Bundesanstalt Braunschweig) sprach sich für eine deutliche Reduzierung der Metadaten für eine Standardisierung aus. Es müssten die bereits bestehenden, quasi standardisierten Formate für eine Normierung weiterentwickelt werden. Auch Frank M. Bischoff (Archivschule Marburg) fragte grundsätzlich, welchen Nutzen standardisierte Metadaten für die Langzeitarchivierung haben. Die Metadaten seien nach seiner Meinung vor allem für die Protokollierung von Ereignissen bei der Aufbewahrung von Originalen (Historie) wichtig. Aus der Sicht von Matthias Hemmje (Fernuniversität Hagen) fehle für die Prüfung der Metadaten ein nationaler State-of-the-Art-Bericht. Erst durch eine solche Lückenanalyse könne der Standardisierungsbedarf substanziell ermittelt werden. In der Beschreibung der Rundfunkarchive sei der große Aufwand deutlich geworden, da die vielen internationalen Filme sehr viele unterschiedliche Metadaten erzeugen. Reinhard Altenhöner (DNB) machte deutlich, dass mit METS bereits ein Standard für ein Containerformat vorliegt. In anderen Bereichen müsse man sich dagegen mehr um eine Standardisierung bemühen. Im Falle der Austauschformate sei nach Altenhöner der Common Sense für die existierenden Systeme ein zentrales Element, weshalb man prüfen müsse, wo sich die austauschrelevanten Bereiche befinden. Der ursprüngliche Ansatz von LMER sei gerade der Aspekt des Austauschformats, auch wenn derzeit die Implementierung im Vordergrund steht. Matthias Hemmje (Fernuniversität Hagen) sieht bei der Diskussion um Quellen- und Zielformate ebenfalls das Austauschformat auf der Grundlage von XML als entscheidenden Schritt an. Abschließend fragte Jens Klump (Geoforschungszentrum Potsdam), welche genauen maschinenlesbaren Informationen für eine Standardisierung notwendig seien?

Die Moderatorin Heike Neuroth (SUB-Göttingen) fasste die Diskussion dieser Sektion zusammen. Demnach verdeutliche die Aussprache, dass derzeit keine einheitlichen bzw. standardisierten Metadaten entwickelt werden könnten. Daher solle man sich stärker hinsichtlich der Austauschformate engagieren. Des Weiteren sei eine eindeutige Analyse über die Verwendung der Formate in Deutschland ein dringendes Desiderat. Genauso sei unklar, welche Standards bereits vorlägen. Dieses Resümee begründe die Notwendigkeit

eines State-of-the-Art-Berichts. Sie regte deshalb an, die bestehende Arbeitsgruppe in nestor zu beauftragen, eine Analyse der Metadatenstandards für die Langzeitarchivierung zu verfassen. Weitere Schritte seien die Übersetzung von PREMIS ins Deutsche, die Durchführung von Fallbeispielen und deren Verbreitung in die einzelnen Communities. Die AG-Ergebnisse von nestor sollten im nächsten Schritt in einen Normierungsprozess führen.

3.4. Qualität digitalisierter Kulturgüter

Einführung:

Den Auftakt des zweiten Workshoptags bildeten die Qualitätskriterien von Digitalisaten. Die Nachnutzbarkeit und Langlebigkeit digitalisierter Kulturgüter hängt nach der Meinung der Workshop-Teilnehmer hauptsächlich von den bei der Digitalisierung zugrunde gelegten technischen Parametern beim Scanning und der Qualitätssicherung ab. Im Bereich der Audio- und Printdigitalisierung kultureller Bestände liegen derzeit zahlreiche Empfehlungen vor, die mitunter unscharf, verwirrend und widersprüchlich formuliert sind. Im Mittelpunkt der Sektion stand daher der Versuch, die Voraussetzungen einer Standardisierungsaktivität zu identifizieren.

Kernsätze der Vorträge:

Zu Beginn stellte Ralf Stockmann (SUB-Göttingen) die Digitalisierungspraxis des Göttinger Digitalisierungszentrums vor. Demnach sei für die bei der Digitalisierung anzulegende Qualität die Frage nach den späteren Nutzungsarten entscheidend. Stockmann empfahl eine DIN-Zertifizierung, um dadurch vergleichbare Qualitätsstandards herzustellen. Zudem seien je nach Nutzungsszenarium mehrere Qualitätsstufen notwendig.

Im Anschluss daran berichtete Thomas Wolf-Klostermann (Bayerische Staatsbibliothek München) über die Digitalisierungspraxis an der Bayerischen Staatsbibliothek. Demnach dürfe sich die Qualitätsdiskussion nicht nur auf Scan-Parameter beschränken, sondern müsse auch die Umgebungsbedingungen, wie Vorlage und Verwendungszweck, berücksichtigen.

Den letzten Beitrag lieferte Nicolas Flores-Herr (DNB), der über notwendige Gütekriterien bei der Digitalisierung von Bildern und Drucken referierte. Demnach sei der speicherintensive Scan-Parameter (hohe Scanauflösung, hohe Farbtiefe) nicht das einzige Qualitätskriterium, sondern bei unvorteilhafter Wahl ein großer Kostenfaktor bei der Archivierung der Digitalisate. Es müssten nach Ansicht von Flores-Herr solche Scan-Parameter gewählt werden, die die digitale Rekonstruktion des Originals ohne Informationsverlust ermöglichen. Nach dem Referenten gelte es, so zu digitalisieren, dass die im Bild der Vorlage enthaltene Information sowie sämtliche Details und der gesamte Farbumfang erfasst werden.

Diskussion:

Der Moderator Günter Mühlberger (Universität Innsbruck) wies darauf hin, dass es sich bei der Themenvorstellung um sehr heterogene Vorträge handelte. In dem ersten Diskussionsbeitrag führte Frieder Schmidt (DNB) aus, dass das genaue Verständnis über die einzelnen Komponenten der Digitalisierung sehr wichtig sei. Außerdem könne man eine Standardisierung befördern, indem man beispielsweise einem marktüblichen Anbieter die Herstellung eines proprietären Formats überlasse. Günter Mühlberger erwiderte daraufhin, dass viele Bibliotheken bei ihrer Kaufentscheidung eines Scanners auf detaillierte Information angewiesen seien. Bereits im Jahr 1997 seien mit der Einrichtung des Göttinger Digitalisierungszentrums klare Qualitätsvorgaben aufgestellt worden, jedoch stelle man seit dieser Zeit kaum ein Fortschritt fest. Mühlberger regte deshalb die Zertifizierung eines Digitalisierungszentrums an. Nicolas Flores-Herr (DNB) bekräftigte seine Einschätzung, dass man den Blickpunkt nicht nur auf Mindeststandards beim Scannprozess und dem jeweiligen dpi-Wert legen sollte, sondern auch auf die genaue Dokumentation der Übertragung. Nach dem Digitalisieren liegen meist keine Informationen über den Digitalisierungsprozess (z. B. Scanparameter) in Form von Metadaten vor.

Im Anschluss daran plädierte Angelika Menne-Haritz (Bundesarchiv) für das Abschiednehmen einer Entweder-oder-Lösung. Man müsse vielmehr das augenblicklich beste Verfahren umsetzen. Die Praxis des Bundesarchivs sei es, den Mikrofilm als obere Sicherungsebene einzusetzen, um der Nutzerorientierung gerecht zu werden. Mit einem Mikrofilm könne das Dokument im Magazin verbleiben, ohne ständig ausgehoben zu werden. Frau Menne-Haritz

sprach die Empfehlung aus, mit den heute vorhandenen technischen Möglichkeiten die optimale Lösung herbeizuführen. Die Auflösung 75dpi eines Digitalisats sei heute als Minimum für die Internetanzeige bereits zu leisten, mehr sollte es nach Ansicht Menne-Haritz' aber auch nicht werden. Durch eine solche Strategie seien nicht zuletzt die Kosten beherrschbar. In diesem Zusammenhang sagte Olaf Brandt (SUB-Göttingen), dass Wirtschaftlichkeit zwar durchaus eine wichtige Komponente darstelle, jedoch seien vorab definierte Qualitätsstandards zu bevorzugen. Manfred Thaller (Universität Köln) betonte in seinem Beitrag, dass das oberste Ziel der Nutzungsaspekt der Digitalisate sei. Des Weiteren müsse man sich von dem Gedanken verabschieden, nur einmal ein Digitalisat herzustellen. Die aktuelle Praxis beweise, dass der technische Fortschritt alle zehn Jahre zu einer besseren Qualität führe. Im Anschluss daran unterstützte Olaf Brandt (SUB-Göttingen) noch einmal die Aussagen von Flores-Herr (DNB). Nach der Ansicht von Günter Mühlberger (Universität Innsbruck) fehlten für die Digitalisierungszentren ähnliche Kriterien, wie sie für vertrauenswürdige Archive bereits gelten. Nicolas Flores-Herr (DNB) betonte daraufhin noch einmal, dass Qualitätsstandards von der jeweiligen Vorlage abhängen. Unter dieser Voraussetzung könne man nicht von einer einheitlichen Normierung sprechen. Auch Ralf Stockmann (SUB-Göttingen) legte in seinem Statement den Fokus auf die Dokumentation der einzelnen Digitalisierungsebenen. Nach seiner Ansicht müsse man sich im Klaren sein, welche Kompromisse man beim Scannen in Kauf nehmen möchte. Darüber hinaus empfahl er, sofort in die Standardisierungsdiskussion im Sinne des Nutzers einzusteigen. Frieder Schmitt (DNB) machte am Ende der Diskussion darauf aufmerksam, dass man für die normierten Digitalisate weitere Partner in der Druckindustrie oder bei den optischen Anbietern gewinnen müsse. Reinhard Altenhöner (DNB) erkannte in der zurückliegenden Diskussion, dass man sich bei der Digitalisierung langfristig auf eine Normierungsstrategie einigen sollte, um damit auch die eingetretene technische Weiterentwicklung aufzufangen. Die notwendigen Bereiche seien Dokumentationsrichtlinien, Metadaten und Qualitätsmanagement. Der derzeitige Diskussionsstand der Digitalisierung deute aber darauf hin, dass man die Möglichkeit eines DIN-Fachberichts prüfen sollte. Zudem regte Altenhöner an, nestor für die notwendigen Vorbereitungsarbeiten zu beauftragen.

Der Moderator Günter Mühlberger (Universität Innsbruck) hielt als Ergebnis fest, dass in diesem Bereich noch keine breite Einigung für ein Nor-

mierungsverfahren beim DIN vorhanden sei. Die Schaffung eines runden Tisches in nestor biete sich jedoch als Möglichkeit an, vorab die unterschiedlichen Positionen abzugleichen.

3.5. Dateiformate und Dokumentenmodelle von Digitalisaten

Einführung:
Die Sektion nahm eine neue Entwicklung in Deutschland in den Blickpunkt. Unter den Gesichtspunkten Austauschbarkeit, Authentizität, Nachnutzbarkeit wurden unterschiedliche Speicherformate und Verfahren zur Nachberechnung von Images diskutiert. Hinsichtlich der Langzeitarchivierung von Digitalisaten setzten sich die Workshop-Teilnehmer auch mit dem Aspekt der Nachnutzbarkeit und Standardisierung von Dokumentenmodellen und deren Dokumentation (z. B. von METS-Profiles) auseinander.

Kernsätze der Vorträge:
Markus Enders (SUB-Göttingen) stellte in seinem Kurzvortrag die Verwendung von Dateiformaten und Dokumentenmodellen am Göttinger Digitalisierungszentrum vor. Demnach lassen sich für die Digitalisierung sehr unterschiedliche Dokumentenmodelle finden: globale Sicht auf das Dokument, einfach strukturierte Sicht auf das Dokument sowie zweifach strukturierte Sicht auf das Dokument. Für METS-Dateien stellte Herr Enders (SUB-Göttingen) die „METS-Profiles“ vor, die nach seiner Ansicht eine Möglichkeit der Dokumentation über die Nutzung und Bedeutung der einzelnen Elemente darstelle.

Im Anschluss daran berichtete Thomas Wolf-Klostermann (Bayerische Staatsbibliothek München) über die Praxis an der Bayerischen Staatsbibliothek. Laut den Angaben des Referenten werden nur ganze Werke in der Digitalen Bibliothek digitalisiert, erschlossen und publiziert. Eine fragmentarische Digitalisierung findet dagegen nicht statt. Daneben gibt es einheitliche Workflows zur Erschließung für alle Digitalisate, wobei verschiedene Erschließungstiefen möglich sind. Hinsichtlich der Langzeitarchivierung empfahl Wolf-Klostermann eine Archivierung eines digitalen Masters zusammen mit der XML-Strukturdatei.

Am Ende stellte Frau Menne-Haritz (Bundesarchiv) das „daofind“ dar, das zur Präsentation von Digitalisaten dient. Auf der Website des Projektes ist die prototypische Beständeübersicht in EAD und EAC mit Links zu den drei in EAD kodierten Findbüchern enthalten. Die Findbücher führen zu den digitalisierten Beständen mit 30.000 Images. Die Präsentation der Digitalisate wird über METS gesteuert.

Diskussion:

Die Moderatorin Susanne Dobratz (Humboldt-Universität Berlin) stellte zu Beginn fest, dass die vorgestellten Beiträge sehr unterschiedliche Ebenen ansprechen. Die zentralen drei Problembereiche waren nach Frau Dobratz: a) allgemeine Anforderungen an das Dokument; b) das Verhältnis zwischen dem Dokumentenmodell und den Digitalisaten und c) spezifische Anforderungen der Dokumentenmodelle.

Tobias Steinke (DNB) fragte nach dem eigentlichen Normierungsbedarf für Dateiformate. Aus der Sicht der Langzeitarchivierung sollte man möglichst wenige Formate verwenden. Die komplexen Objektformate könnten nach Ansicht von Markus Enders (SUB-Göttingen) als Container dienen, der bald vorliegen sollte. Angelika Menne-Haritz (Bundesarchiv) empfahl, nicht alle Formate zu standardisieren, um nicht durch solche Eingriffe den technischen Fortschritt zu bremsen. Hinsichtlich der Formate könne man sich am Beispiel von Baden-Württemberg orientieren. Die EAD biete eine präzise Darstellung der Elemente, wobei Konvertierungsformate und Präsentationsformate identisch seien. Darauf hin erwiderte Jens Ludwig (SUB-Göttingen), dass man sich keiner Handlungsmöglichkeiten berauben soll und den Markt nicht durch zu viele Standards eingrenzen solle. Susanne Dobratz (Humboldt-Universität Berlin) erachtete eine Normierung einzelner Formate ebenfalls als nicht dringend notwendig. Eine Anforderungsanalyse der Dateiformate könne im Rahmen eines DIN-Fachberichts erstellt werden.

Die Moderatorin Susanne Dobratz erkannte im Ergebnis der Diskussion noch keine Notwendigkeit eines Normierungsverfahrens beim DIN. Es sei vielmehr ein DIN-Fachbericht anzustreben. Da das METS-Editorial-Board in eine ISO-Zertifizierung führen könnte, sollte Deutschland in naher Zukunft die genauen Anforderungen in einem Round-Table von nestor ausarbeiten.

3.6. Metadaten zur Digitalisierung

Einführung:

In der abschließenden Sektion des Frankfurter Workshops standen die Metadaten zur Digitalisierung im Mittelpunkt der Diskussion. Die Teilnehmer machten darauf aufmerksam, dass Metadaten im digitalen Informationszeitalter neben Volltext-Indices eine entscheidende Rolle spielen. Im nationalen Portal vascoda, das die heterogen verteilten digitalen Metadaten-Sammlungen in Deutschland (wie z. B. OPACs, OLC, Informationsführer etc.) über einen Sucheinstieg nachweist, ist das vascoda Application Profile als nationaler Standard bereits entwickelt worden. Zurzeit finden Überlegungen statt, als einen Schritt in Richtung qualitätsgeprüfte und standardisierte Lieferungen von Metadaten, ein Zertifizierungsverfahren einzuführen. Auch im Rahmen des Kompetenzzentrums Interoperable Metadaten (KIM), das als DCMI Affiliate etabliert werden soll, werden Zertifizierungsverfahren für die Entwicklung von standardisierten, interoperablen Metadatenformaten entwickelt.

Kernsätze der Vorträge:

Den Auftakt bildete der Vortrag von Heike Neuroth (SUB-Göttingen), die über die Einführung von KIM berichtete. In ihrem Kurzvortrag empfahl die Referentin die Etablierung einer KIM „Kernarbeitsgruppe“, die u. a. für die Formulierung von Empfehlungen für interoperable Metadatenmodelle verantwortlich sein sollte. Des Weiteren regte sie die DIN-Zertifizierung von DCMES und eine deutschsprachige Übersetzung des DCMI Abstract Models an. Weitere vorgestellte Maßnahmen bezogen sich auf die DIN-Zertifizierung von nationalen Aplication Profiles (z. B. vascoda, ELAN) sowie auf das W3C Semantic Web Development unter Berücksichtigung deutscher Entwicklungen und Nachnutzungen von Standards in Deutschland.

Im Anschluss daran berichtete Christine Frodl (DNB) über die Standardisierungspraxis an der Deutschen Nationalbibliothek.[14] Seit 1998 wird das

[14] http://www.d-nb.de/standardisierung/afs/afs.htm.

im Rahmen eines DFG-Projektes entwickelte Metadatenformat für Online-Hochschulschriften, MetaDiss, für den Austausch von Metadaten und dazugehörigen Volltexten in Deutschland verwendet. Das Potenzial des neuen, XML-basierten-Formats XMetaDiss besteht nach den Angaben von Frau Frodl unter anderem in der Nutzung von hierarchischen Strukturen und der damit verbundenen Vermeidung von Zuordnungsfehlern, in der Schaffung eines Formates für Online-Hochschulschriften, das per OAI-Protokoll ausgetauscht werden kann und in der Schaffung eines zu internationalen Standards kompatiblen Formates, damit deutsche Online-Hochschulschriften in internationale Metadatensuchmaschinen eingebunden werden können. Die Referentin kommt zu dem Ergebnis, dass man mittelfristig ein gültiges, standardisiertes, international anwendbares Format zur Beschreibung und zum Austausch von Metadaten und Volltexten deutscher und internationaler Hochschulschriften in Zusammenarbeit mit dem Standards Committee der Networked Digital Library of Theses and Dissertations entwickeln solle. Diese Entwicklung sollte weltweit etablierte Standards wie z. B. Dublin Core und den ETD-MS-Standard der NDLTD, aber auch national in der Anwendung befindliche Formate wie XMetaDiss, TEF (Théses électroniques francaises) etc. berücksichtigen.

Zum Schluss präsentierte Stefan Wolf (Bibliotheksservice-Zentrum Baden-Württemberg) den Gebrauch der Metadaten im Bibliotheksservice-Zentrum in Konstanz. Demnach entscheidet sich die Leistungsfähigkeit einer Installation oder Institution bei der Standardisierung von Metadaten daran, mit welcher Qualität sie ohne Verlust von Differenzierungsgraden oder standardisierten Vorleistungen (Normdateien) unterschiedlich strukturierte oder „dialektal“ gefärbte Varianten eines Metadatenformates in ihren für langfristige Pflege, Verwaltung und Suche vorgesehenen Datenbestand eingliedern könne. Herr Wolf betonte, dass die Zukunft aller Metadatendarstellung ausschließlich in XML-Darstellungen liege, die durch XML-Schema definiert seien.

Diskussion:

In der Diskussion fragte Günter Mühlberger (Universität Innsbruck), wie groß die Verbreitung von XMetaDiss in Deutschland tatsächlich sei. Stefan Wolff (Bibliotheksservice-Zentrum Baden-Württemberg) und Nikola Korb (DNB) bestätigten die große Durchdringung von XMetaDiss für den Hochschulbereich in Deutschland und im deutschsprachigen Raum. Die Zahl der Stan-

dards müsse man nach Ansicht Wolfs aber verringern. Darüber hinaus lenkte Stefan Wolf den Blick auf die Verwendung des Dokumentenservers OPUS. Nach der Meinung von Nikola Korb soll man zwar den Blick nicht nur auf OPUS legen, jedoch sei der Server im Moment der potenteste Vertreter auf dem Markt. Zudem berichtete Korb, dass in der Deutschen Nationalbibliothek für die wissenschaftlichen Publikationen XMetaDiss und für die Verlagspublikationen ONIX verwendet wird, sodass eine doppelte Strategie in der DNB vorliege. Heike Neuroth (SUB-Göttingen) wies in ihrem Redebeitrag darauf hin, dass man in diesem Bereich nicht zu sehr auf Standardisierungen setzen sollte. Im Ergebnis werde sich auf dem Markt ein tragfähiges Verfahren durchsetzen. Da seit kurzem ein internationaler Standard für E-Prints bereits vorliege, empfahl sie, XMetaDiss auf ein internationales Niveau zu heben. Hierfür müsse man alle entscheidenden Gremien mit einbeziehen. Stefan Wolf (Bibliotheksservice-Zentrum Baden-Württemberg) bestätigte, dass man mit dem arbeiten müsse, was der Markt hergebe. Tobias Steinke (DNB) betonte die Notwendigkeit von inhaltlichen Metadaten bei der Digitalisierung (Strukturdaten). Eine zukünftige Formatentwicklung werde sich sehr stark an XML orientieren. Daraufhin schlug Heike Neuroth (SUB-Göttingen) vor, ein Gespräch zwischen den vorhandenen Verantwortlichen und Entwicklern von XMetaDiss und den Vertretern der DINI – Deutsches Institut für Netzwerkinformation e. V. – anzuregen. DINI sei die entscheidende Instanz, die Frage nach dem geeigneten Umfeld der Daten zu beantworten. Es müsse ein Container geschaffen werden, an den man von außen „andocken“ könne. Zwar sieht Angelika Menne-Haritz (Bundesarchiv) in METS eine gute Voraussetzung auf dem Weg zu einem Standard, jedoch mahnte Reinhard Altenhöner (DNB) an, darüber hinaus ein konkretes Konzept zu entwickeln. Hierfür solle man aber noch eine breite Grundlage für einen Kompromiss finden und Gütekriterien für die Qualität der Metadaten formulieren. Schließlich müsse man sich nach Ansicht Altenhöners beim Mapping auf einer jeweils gleichen Ebene bewegen. Heike Neuroth (SUB-Göttingen) resümierte, dass man bei den eventuellen Normverfahren noch ganz am Anfang stünde und man sich noch auf der Ebene der Definition von Kriterien befinde. Die auf dem Workshop angestoßenen interoperablen Konzepte müssten in weiteren Schritten erst entwickelt werden.

In der Zusammenfassung der Diskussion von Reinhard Altenhöner (DNB) könne man in Zukunft auf der Grundlage von XMetaDiss und unter

Berücksichtung internationaler Entwicklungen einen nationalen Standard anstreben. Dabei warf er die Frage auf, wie man internationale Partner konkret mit einbeziehen könne. Einen konkreten Bedarf eines Normierungsverfahrens beim DIN erkannte Altenhöner aber nicht. Der Problembereich der Interoperabilität sei noch ungelöst und außerdem müssen noch Prüfkriterien für die Formate entwickelt und aufeinander bezogen werden. Dublin Core könne als Grundmuster auf abstrakter Ebene verwendet werden, wie Reinhard Altenhöner weiter ausführte. Der Ansatz von Heike Neuroth (SUB-Göttingen) gehe hierbei in die richtige Richtung: Man müsse im Sinne einer Vorarbeit für DIN die vorliegenden Projekte weiterentwickeln und Gütekriterien definieren. Der geeignete organisatorische Arbeitsbereich hierfür sei KIM. Zudem müsse DINI aktiv miteinbezogen werden.

4. Workshop-Ergebnisse und Ausblick

In dem letzten Abschnitt werden die erzielten Workshop-Ergebnisse für die behandelten Themenbereiche der Langzeitarchivierung digitaler Objekte und der Digitalisierung in knapper Form zusammengefasst und ein Ausblick für die weitere Arbeit eines vorgesehenen Normierungsprozesses vorgestellt.

In Frankfurt am Main wurde am 2. und 3. November 2006 im Rahmen eines Workshops erstmals auf nationaler Ebene der Standardisierungsbedarf für zwei zentrale Bereiche der deutschen Gedächtnisorganisationen, nämlich die Langzeitarchivierung digitaler Objekte und der Digitalisierung, ermittelt. Eine der Zielvorgaben der Veranstalter war es, eine breit angelegte Expertengruppe aus den unterschiedlichen Gedächtnisorganisationen und Forschungseinrichtungen an einen Tisch zu führen. Die hohe Beteiligung aus Deutschland, Österreich und der Schweiz zeigte, dass die einzelnen Institutionen die Notwendigkeit von normierten bzw. zertifizierten Standards in diesen Bereichen als sehr wichtig erachten. An dem Workshop nahmen insgesamt 50 Experten aus 29 Institutionen aus dem deutschsprachigen Raum an beiden Veranstaltungstagen teil. Dabei beteiligten sich am ersten Tag 43 und am zweiten Tag 39 Experten. Insgesamt waren 32 Teilnehmer an beiden Veranstaltungstagen anwesend.[15] Darüber hinaus belegte der Workshop, dass die INS-Initiative den geeigneten Rahmen für die angestrebten Standardisierungsaktivitäten bietet.[16] Die Initiative bietet die geeignete Plattform, die deutschen Gedächtnisorganisationen noch stärker in den europäischen Kontext einzubinden und die nationalen Interessen konkurrenzfähiger zu machen.

Der Frankfurter Workshop führte zu unterschiedlichen Ergebnissen. Für alle vorgestellten Themenblöcke konnte zwar grundsätzlich ein Normierungsbedarf in Deutschland herausgearbeitet werden, jedoch ist der genaue Nor-

[15] Insgesamt waren zehn Teilnehmer aus sieben beteiligten Universitäten anwesend. Weitere sieben Teilnehmer kamen aus den Staatsarchiven bzw. der Archivschule. Die zahlenmäßig am stärksten vertretene Gruppe bildeten die 21 Teilnehmer aus den fünf eingeladenen Bibliotheken. Der Mitveranstalter DIN - Deutsches Institut für Normung e. V. stellte zwei Vertreter. Museen, Verlage, Rechenzentren, Digitalisierungsfirmen waren mit jeweils einem Teilnehmer vertreten und die sonstigen sechs anwesenden Forschungseinrichtungen stellten auf dem Workshop in Frankfurt am Main ebenfalls jeweils einen Teilnehmer. Siehe: Teilnehmerliste im Anhang.

[16] Das DIN wird in seinem voraussichtlichen Fördervolumen finanzielle Mittel mit einplanen, die für eine internationale Gremienarbeit (beispielsweise für Treffen von PREMIS in den USA) anfallen können. Welche nationalen bzw. internationalen Gremien für einen Normierungsprozess im Detail zu beteiligen sind, bedarf mitunter noch weiterer systematischerer Klärung.

mierungsgrad in vielen Bereichen noch nicht so konkretisiert bzw. auf Erfahrungswissen fundiert, um daraus einen unmittelbaren und breit angelegten initialen Normierungsprozess abzuleiten. Eine Reihe von Themen bedarf noch einer intensiveren Diskussion, und die Experten werden in naher Zukunft weitere Verhandlungen darüber durchführen. Trotzdem ermittelten die Teilnehmer für einzelne Bereiche konkrete Ansatzpunkte.

Die folgenden Empfehlungen und Maßnahmenvorschläge beziehen sich im Kern auf die abschließende mündliche Zusammenfassung des Workshops:

Persistent Identifier

Ergebnisse: Die Teilnehmer heben die „N2T-Initiative“ als lohnenswertes Projekt für eine Standardisierungsaktivität hervor, wobei noch die genaue Normrelevanz zu prüfen ist. Für den Bereich „DOI“ wird im Rahmen der TIB und nestor die weitere Entwicklung des angestrebten ISO-Standards untersucht. Für die URN wird auf den RfC bei der IETF hingewiesen, der für einen Normierungsprozess zugrunde gelegt werden sollte. Ob daraus auch ein Verfahren für einen ISO-Standard wird und welche Anforderungen daraus resultieren, müssen die einzelnen Communities prüfen. Die weitere Koordination dieser Aufgaben wird nestor übernehmen.

Fazit: Die Workshop-Teilnehmer beschließen, dass für den Bereich Persistent Identifier der Start in eine Normierungsaktivität beim DIN sinnvoll ist. Außerdem wird die Bildung eines Konsortiums für die Aufnahme eines Normprozesses empfohlen. Weitere Aktivitäten sollten den Normierungsprozess begleiten.

Vertrauenswürdige Archive

Ergebnisse: Im Falle des Kriterienkatalogs für vertrauenswürdige Archive halten die Workshop-Teilnehmer fest, dass die aufgestellten Anforderungen im Rahmen eines Normprozesses auf den Kern reduziert werden sollten. Die Teilnehmer sind sich über den Bedarf eines Normierungsverfahrens einig,

weil „Vertrauenswürdigkeit“ per se eine hohe Affinität zu Standards und Normen besitzt. Die Möglichkeiten für eine nationale Initiative werden geprüft. Im weiteren Vorgehen wird zunächst das existierende Normenwerk zugrunde gelegt, insbesondere die Bereiche IT-Sicherheit sowie die bestehenden internationalen Normen für Qualitätsmanagement, Schriftgutverwaltung und digitales Archiv (ISO 9000ff., 15489 und 14721). In nestor sind im Dezember 2007 und im Dezember 2008 insgesamt zwei Workshops vorgesehen, auf denen die Arbeitsergebnisse präsentiert werden.[17]

Fazit: Die Workshop-Teilnehmer beschließen, dass nach Abschluss der notwendigen Vorarbeit in nestor ein Normierungsprozess beim DIN eingeleitet wird.

Metadaten zur Langzeitarchivierung und Objektformat

Ergebnisse: In dem Themenfeld Metadaten zur Langzeitarchivierung und Objektformat werden unterschiedliche Ansätze und Umsetzungsverfahren formuliert. Die nächsten Handlungsschritte dienen deshalb dazu, den Bedarf für eine einheitliche Konzeption unter besonderer Berücksichtigung der prozesshaften Metadaten weiter zu ermitteln. Des Weiteren beschließen die Teilnehmer, dass aufgrund der unterschiedlichen Ansätze die (Weiter-) Entwicklung eines adäquaten Austauschformats vorangetrieben wird. Die weitere Arbeit soll nestor übernehmen.

Fazit: Für eine Normierung beim DIN ist es noch zu früh. Es müssen noch notwendige Vorarbeiten für ein Normierungsverfahren geleistet werden. Am 7.12.2006 trifft sich an der SUB eine nestor-Arbeitsgruppe, um im Rahmen der Meilensteinplanung die angesprochenen Themenbereiche weiter zu planen.[18]

---

[17] Hierbei wird auch unter Berücksichtigung der OCLC/RLG und DCC/DPE/Planets auf eine internationale Einigung von "common principles" abgezielt.

[18] Im Rahmen der nestor-Arbeitsgruppe soll bis April 2007 das Mapping zwischen LMER u. PREMIS und bis Mai 2007 die Entwicklung eines ""Universal Object Formats“ abgeschlossen werden.

Qualität digitaler Kulturgüter

Ergebnisse: Die Workshop-Teilnehmer halten für den Bereich der Digitalisierung fest, dass weitere Maßnahmen für einheitliche Verfahren zu führen sind, um über den Entwicklungsstand von 1997 hinauszukommen. Die Teilnehmer empfehlen, die Normbereiche a) Vorlagemessung, b) der Festlegung von MTF und c) ICC-Profile weiterzuentwickeln. Des Weiteren werden die Dokumentation des Scan-Produktionsprozesses und die Funktion des Qualitätsmanagements für die messbaren Prozesse hervorgehoben. Am Ende des Diskussionsprozesses soll ein DIN-Fachbericht stehen. Als notwendiger Schritt wird die Einsetzung eines runden Tisches in nestor angeregt.

Fazit: Die Teilnehmer erkennen derzeit noch keine stabile Grundlage für einen Normierungsprozess.

Dateiformate und Datenmodelle von Digitalisaten

Ergebnisse: In der Sektion Dateiformate und Datenmodelle von Digitalisaten müssen detaillierte Anforderungen an die Dokumentenspeicherung und Dokumentation noch klarer formuliert werden. Die Teilnehmer halten fest, dass man die Zahl der Formate zwar begrenzen, jedoch eine gewisse Vielfalt wegen des technischen Fortschritts zulässig sei. Für die weitere Arbeit wird die Einrichtung einer Arbeitsgruppe empfohlen.

Fazit: Für eine Normierung beim DIN ist es noch zu früh. Zunächst müssten weitere Fragen formuliert und präzisiert werden, bevor auch konkrete Lösungsansätze entwickelt werden könnten. Das Ergebnis der Diskussion sollte ein DIN-Fachbericht sein.

Metadatenformate

Ergebnisse: Die Teilnehmer beschließen für den Bereich der Metadatenformate, dass unter Berücksichtigung der internationalen Entwicklung Kriterien

für die Interoperablität von Metadaten-Formaten auf der Basis von XML entwickelt werden sollen. Bei der Einsetzung von geplanten Arbeitsgruppen sollte DINI – die Deutsche Initiative für Netzwerkinformation - einbezogen werden. Die Arbeitsergebnisse sollen verbreitet und die Anforderungen eindeutig benannt werden. Als Plattform der weiteren Arbeit wird KIM, ein kooperatives DFG-Projekt der SUB und der DNB zum Aufbau eines Kompetenzzentrums Metadaten, empfohlen.

Fazit: Für eine Normierung beim DIN ist es derzeit noch zu früh. Zunächst müssten auch hier weitere Fragen formuliert und präzisiert werden, bevor auch konkrete Lösungsansätze entwickelt werden könnten.

5. Anhang: Teilnehmer, Programm und eingereichte Thesenpapiere

Teilnehmer:

Teilnah-me2.11	Teilnahme 3.11	Name	Institution
+	+	Altenhöner, Reinhard	Deutsche Nationalbibliothek
+	-	Bischoff, Dr. Frank M.	Archivschule Marburg
+	+	Brandt, Olaf	SUB Göttingen
+	+	Brase, Jan	Technische Informationsbibl. UNI Hannover
+	+	Brocks, Holger	FernUniversität Hagen
+	+	Brodersen, Maren	Deutsche Nationalbibliothek
+	+	Dobratz, Susanne	Humboldt Universität zu Berlin
+	+	Enders, Markus	SUB Göttingen
+	+	Ermert, Axel	Institut für Museumskunde
+	+	Faigle, Maja	Georg Thieme Verlag KG
-	+	Flores-Herr, Dr. Nicolas	Deutsche Nationalbibliothek
+	+	Frodl, Christine	Deutsche Nationalbibliothek
+	+	Grahle, Roman	DIN-Deutsches Institut für Normung e.V.
+	+	Gubler, David	ABTM Uni Basel
+	-	Hackel, Dr. Siegfried	Phys.-Techn. Bundesanstalt, Braun-schweig
+	+	Hemmje, Prof. Dr. Matthias	FernUniversität in Hagen
+	+	Höhler, Philipp	Deutsche Nationalbibliothek
+	-	Huth, Karsten	Bundesarchiv
+	+	Jäschke, Gerald	FernUniversität in Hagen
-	+	Jannidis, Prof. Dr. Fotis	TU Darmstadt
+	+	Jehn, Dr. Mathias	Deutsche Nationalbibliothek
+	-	Kahnert, Dr. Wolfgang	Landesarchiv NRW – Technisches Zen-trum
+	-	Kappelhoff, Dr. Bernd	Landesarchiv Niedersachsen
+	-	Klump, Dr. Jens	GeoForschungsZentrum Potsdam
+	+	Korb, Dr. Nikola	Deutsche Nationalbibliothek
+	+	Ludwig, Jens	SUB Göttingen
-	+	Lüdeling, Anke	Humboldt Universität zu Berlin
-	+	Maier, Dr. Gerald	Landesarchiv Baden-Württemberg
-	+	Menne-Haritz, Prof. Dr. An-gelika	Bundesarchiv
+	+	Mühlberger, Dr. Günter	Universität Innsbruck
+	+	Müller, Florian	Imaging and Media Lab, Basel
+	+	Neuroth, Dr. Heike	SUB Göttingen
+	-	Rödig, Peter	Universität der Bundeswehr München
+	+	Roth, Andreas	Deutsche Nationalbibliothek
+	+	Sahle, Patrick MA	SUB Göttingen
+	+	Sauter, Prof. Dipl.-Ing. Dietrich	Institut für Rundfunktechnik GmbH, München
+	+	Schäfer, Dr. Udo	Kulturbehörde - Hamburg Staatsarchiv

Teilnah-me2.11	Teilnahme 3.11	Name	Institution
+	+	Schaßan, Torsten	Foto Marburg
+	+	Schmidt, Dr. Frieder	Deutsche Nationalbibliothek
+	-	Stäcker, Dr. Thomas	Herzog August Bibliothek Wolfenbüttel
+	+	Steinke, Tobias	Deutsche Nationalbibliothek
-	+	Steinmetz, Prof. Dipl.-Ing. Arnd	Hochschule Darmstadt
+	+	Stockmann, Ralf	SUB Göttingen
+	+	Strathmann, Stefan	SUB Göttingen
+	-	Suchodoletz, Dirk v.	Rechenzentrum Universität Freiburg
-	+	Thaller, Prof. Dr. Manfred	Universität zu Köln
+	-	Wende, Dr. Ingo	DIN Deutsches Institut für Normung e.V.
+	+	Wolf, Stefan	Bibliotheksservice-Zentrum Ba-Wü, Konstanz
+	+	Wolf-Klostermann, Dr. Thomas	BSB München
+	-	Wollschläger, Dr. Thomas	Deutsche Nationalbibliothek

Programm:

Persistent Identifier **10:00 – 12:30 Uhr : Moderation Herr Altenhöner (Deutsche Nationalbibliothek)**		
1.1	Einführung zu organisatorischen Rahmenbedingungen	Referent: Herr Enders (SUB-Göttingen)
1.2	Einführung zum Stand URN-Einsatz	Referentin: Frau Dr. Korb (DNB)
1.3	Einführung zum Stand DOI in Deutschland	Referent: Herr Dr. Brase (TIB-Hannover)
Vertrauenswürdige Archive **12:30 – 13:30 Uhr : Moderation Herr Dr. Bischoff (Archivschule Marburg)**		
2.1	Einführung: Sachstandsbericht Kriterienkatalog von nestor	Referentin: Frau Dobratz (HUB)
Metadaten zur Langzeitarchivierung und Objektformate **14:00 – 16:30 Uhr : Moderation Frau Dr. Neuroth (Staats- und Universitätsbibliothek Göttingen)**		
3.1	UOF und LMER (kopal)	Referent: Herr Steinke (DNB)
3.2	METS und PREMIS (MathArc)	Referent: Herr Brandt (SUB-Göttingen)
3.3	Beispiele für andere Metadatenformate	Referent: Herr Dr. Wolf-Klostermann (BSB)

Qualität digitalisierter Kulturgüter **09:00 – 10:00 Uhr : Moderation Herr Dr. Mühlberger (Department for Digitisation and Digital Preservation University Innsbruck Library)**

1.1	Einführung SUB-Sachstandsbericht	Referent: Herr Stockmann (SUB-Göttingen)
1.2	Einführung BSB-Sachstandsbericht	Referent: Herr Dr. Wolf-Klostermann (BSB)
1.3	Einführung Maschinell erhobene Qualitätsstandards	Referent: Her Dr. Flores-Herr (DNB)
Dateiformate und Dokumentenmodelle von Digitalisaten **10:00 – 11:00 Uhr : Moderation Frau Dobratz (Humboldt-Universität Berlin)**		
2.1	Praxis GDZ	Referent: Herr Enders (SUB-Göttingen)
2.2	Praxis BSB	Referent: Herr Dr. Wolf-Klostermann (BSB)
2.3	Praxis „daofind“	Referentin: Frau Prof. Dr. Menne-Haritz (Bundesarchiv)
Metadaten **11:30 – 12:30 Uhr : Moderation Herr Altenhöner (Deutsche Nationalbibliothek)**		
3.1	Einführung SUB	Referentin: Frau Dr. Neuroth (SUB-Göttingen)
3.2	Bericht über die Praxis an der DNB	Referentin: Frau Frodl (DNB)
3.3	Über den Gebrauch von Metadaten an der BSZ in Baden-Württemberg	Referent: Herr Wolf (BSZ-BaWü)

Thesenpapiere:

Olaf Brandt

METS und PREMIS (MathArc)

Kurze Angaben zu Organisation

Die Niedersächsische Staats- und Universitätsbibliothek Göttingen ist als innovativer Anbieter von integrierten (digitalen) Informationsdienstleistungen in vielen Bereichen der Standardisierung tätig. Zu nennen sind hier exemplarisch die Tätigkeiten in der Dublin Core Metadata Initiative, dem METS Editorial Board, dem W3C und dem PREMIS Editorial-Committee.

Problemaufriss

Um die langfristige Verfügbarkeit von digitalen Objekten zu gewährleisten sind Informationen über diese Objekte - Langzeitarchivierungsmetadaten - essentiell notwendig. Diese Metadaten müssen die wahrscheinlich benötigten Informationen für die Benutzbarkeit und damit für ihre Verarbeitung in die Zukunft tragen. In diesem jungen Bereich hat sich noch kein gemeinsamer Standard herausgebildet. Es existieren zwei relevante Entwürfe in Deutschland: LMER und PREMIS.

Mit PREMIS können digitale Objekte, Ereignisse, Rechte, „Agenten" und ihre Verknüpfungen beschrieben werden. Somit ist es auch möglich die für die langfristige Verfügbarkeit wichtigen Kontextinformationen, die mit digitalen Objekten verbunden sind, zu beschreiben. Hier schlägt sich nieder, dass PREMIS von unterschiedlichsten Institutionen (Bibliotheken, Museen, Archiven, privaten Firmen) entwickelt wurde und nicht nur Bibliotheken anspricht.

PREMIS ist zugleich eine Initiative, die die Entwicklung und die Pflege der gleichnamigen Metadatenspezifikation und der dazu entwickelten XML-Schemas übernimmt. Die überinstitutionelle und internationale Zusammenarbeit hat PREMIS regen Zuspruch und Anerkennung gebracht, darunter zwei Auszeichnungen (Digital Preservation Award der DPC und den Preservation Publication Award der SAA). Verständnisfragen und Fragen über Implementierungen werden in der regen Mailingliste der PREMIS implementors group diskutiert.

Ein international besetztes Editorial Committee (in dem die SUB vertreten ist) steuert die Pflege des Standards und der XML-Schemas und erstellt Vorschläge für best practices.

Im MathARC-Projekt, welches den Austausch von Daten zwischen zwei Langzeitarchiven zum Ziel hat, haben sich die Implementierungen von uneinheitlichen Metadatensets (LMER und PREMIS) als schwierig zu handhaben herausgestellt.

Der beste Weg für den Austausch ist daher sicherlich ein einheitliches Datenformat. Hier bietet sich die Chance, ein gemeinsames Datenformat auf der Basis von PREMIS zu entwickeln und zu standardisieren. Es könnte entweder ein standardisiertes Datenformat von Metadaten und Contentdaten angestrebt werden oder ein einfacheres Austauschformat – mit einem gemeinsamen Minimalset an Metadaten. Für beides bietet sich PREMIS an.

Man hat in Deutschland und im deutschsprachigen Raum schon erste praktische Erfahrungen mit PREMIS sammeln können (z. B. im Bundesarchiv, im Schweizer Bundesarchiv, an der SUB Göttingen). Weitere Verbreitung wird PREMIS durch die deutsche Übersetzung des „Data Dictionary" erfahren.

Im Rahmen von kopal wird gerade ein Vorschlag für die Implementierung von PREMIS in das bestehende System, welches u. a. auf LMER aufbaut, erarbeitet.

Dr. Jan Brase, Technische Informationsbibliothek der Universität Hannover

Einführung zum Stand DOI in Deutschland

Projekt/zur Institution

Im DFG-Projekt "Publikation und Zitierfähigkeit wissenschaftlicher Primärdaten" wurde eine Infrastruktur zur Registrierung von DOIs für Wissenschaftliche Datensätze geschaffen und erfolgreich getestet. Mit dem System wurden an der Technischen Informationsbibliothek (TIB) bereits über 450.000 Primärdatensätze aus dem Bereich der Geowissenschaften mit persistenten digitalen Identifikatoren versehen. Auf alle Datensätze kann offen über die Kataloge der beteiligten Datenarchive als sogenannte Publikationsagenten zugegriffen werden. Ein Teil der Datensätze ist auch über den Online-Katalog der TIB verfügbar und in Kooperation mit der Deutschen Bibliothek mit URNs versehen.
Für die Kataloge wurde ein ISO690-und Dublin-Core-kompatibles Metadatenschema entwickelt, mit Zusätzen für die Registrierung und Besonderheiten der Datenpublikation.

Problemaufriss

Der Nutzen eines persistenten Identifikators zum Nachweis wissenschaftlicher Inhalte und Wahrung von Urherberrechten ist unbestritten. Der DOI wird in Deutschland allerdings meistens mit kommerziellen Anbietern wie Crossref oder Medra gleichgesetzt.
Die Aktivitäten der TIB als nicht-kommerzieller Registrierungsagentur sollen den weltweit anerkannten Digital Object Identifier (DOI) der wissenschaftlichen Community in Deutschland verfügbar machen. Die Kosten für die Registrierung werden dabei so gering wie möglich gehalten.
Standardisierungsrelevante Aktivitäten
Seit August 2006 ist eine SC9 Arbeitsgruppe eingerichtet, die als ISO Projekt 26324 einen Standard für den DOI entwerfen soll.

Handlungsvorschlag

Die TIB strebt in Kooperation mit den Deutschen Forschungsförderern und weiteren Fachinformationseinrichtungen an, den DOI in der wissenschaftlichen Community in Deutschland weiter kostenneutral zur Verfügung zu stellen.

Susanne Dobratz, Humboldt Universität zu Berlin

Organisatorische Angaben

Die Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung wurde im Rahmen des nestor-Projektes (2003-2006) am 12.10.2004 etabliert und wird im Rahmen von nestor II weiter geführt. Sie besteht aus Vertretern aus den Bereichen Bibliotheken, Verlagswesen, Rechenzentren, Medienzentren, Archivwesen, Museumswesen, Großforschungseinrichtungen, Zertifizierungsexperten. (16 Mitglieder)
AG Leitung:
Susanne Dobratz (Humboldt Universität zu Berlin) und
Dr. Astrid Schoger (Bayerische Staatsbibliothek München)

Problemaufriss

Die zunehmende Produktion digitaler Daten in allen Bereichen des Lebens führt dazu, dass Bibliotheken, Archive, Museen, aber auch Forschungseinrichtungen und Medienzentren digitale Archive aufbauen. Dabei mangelt es an Standards für die technische und organisatorische Realisierung, die zum Aufbau derartiger Archive herangezogen werden können und die Produzenten und Lieferanten digitaler Objekte als auch den Nutzern dieser Archive die Vertrauenswürdigkeit des Archivs in der Form garantieren können, dass die Verfügbarkeit, Lesbarkeit und besonders die Authentizität und Integrität sowie Identität der digital archivierten Objekte erhalten bleibt. Die Interessenten dieses Kriterienkataloges sind:

- Endnutzer, Kooperationspartner, Öffentlichkeit
- Produzenten / Lieferanten digitaler Ressourcen
- Betreiber, Management, Auftraggeber, Politik,

die in Erfüllung eines gesetzlichen Auftrags oder zum Bestehen auf dem Markt allgemein anerkannte Maßstäbe der Bewertung ihrer digitalen Archive benötigen.

Standardisierungsrelevante Aktivitäten

Daher hat nestor auf der Basis umfassender Umfragen und Diskussionen im Juni 2006 einen Kriterienkatalog [1] als Entwurf zur öffentlichen Kommentierung erstellt, der die Funktion hat, Richtlinien zum Aufbau eines vertrauenswürdigen Archivs zu definieren, die über die Spezifik der einzelnen Archive hinweg Gültigkeit besitzen. Dabei gehen die nestor-Aktivitäten mit internationalen Entwicklungen einher, die insbesondere durch die Research Libraries Group in den USA [2] und dem Digital Curation Centre in Großbritannien [3] geschehen und die zu ähnlichen Kriterienkatalogen geführt haben. Die RLG/NARA-Checkliste wird zur Durchführung von Test-Audits eingesetzt und zur Entwicklung von Evaluierungs- und Zertifizierungsverfahren.
Die RLG beabsichtigt, die RLG/NARA-Entwicklungen über die (ISO-)Familie der OAIS-Standards [4] zu standardisieren. Daher ergibt sich die dringende Notwendigkeit über eine DIN und ISO Standardisierung die internationalen Aktivitäten zu koppeln

und international gültige Richtlinien für vertrauenswürdige digitale Langzeitarchive zu definieren.

[1] Draft „Kriterienkatalog vertrauenswürdiger digitale Langzeitarchive – Version 1 (Entwurf zur öffentlichen Kommentierung)"; http://www.nbn-resolving.de?urn:nbn:de:0008-2006060710 (Draft Version in Englisch liegt vor, ist noch nicht veröffentlicht)

[2] RLG/NARA: Audit Checklist for the Certification of Trusted Digital Repositories, 31.08.2005; http://www.rlg.org/en/page.php?Page_ID=20769

[3] Keine Veröffentlichung eines Katalogs, http://www.dcc.ac.uk

Handlungsvorschlag

Mittelfristige Ziele (2007):

1. Forcierung einer internationalen Einigung auf „Common Principles", als Desiderat aus den Entwicklungen in nestor, OCLC/RLG (Robin Dale, David Giaretta) und DCC/DPE (Seamus Ross) sowie dem Planets-Project (Max Kaiser)

Längerfristige Ziele (2007 – 2009):

2. Überarbeitung der Kriterien des Katalogs für eine Standardisierung
3. Entwicklung eines Metrik für die Kriterien des Katalogs

Kontakt:

Dr. Astrid Schoger, Bayerische Staatsbibliothek München, Digitale Bibliothek,
astrid.schoger@bsb-muenchen.de ,
Tel. 089/28638 2600
Susanne Dobratz, Humboldt-Universität zu Berlin, Universitätsbibliothek
10099 Berlin dobratz@cms.hu-berlin.de,
Tel. 030/2093 7070

Markus Enders, Staats- und Universitätsbibliothek Göttingen

Persistent Identifier

Persistent Identifier (PI) dienen zunächst nur einmal dazu, Objekte eindeutig über einen längeren Zeitraum zu identifizieren. In der Praxis werden sie daher sehr eng zusammen mit einem Objekt gespeichert (bspw. in Dokumentenmanagementsystemen, Mapping Tabellen etc.). Um diese PIs auch für den Zugriff auf die Dokumente nutzen zu können, wurden mittlerweile eine Vielzahl unterschiedlicher Resolver-Systeme aufgebaut. Allen Systemen ist gemeinsam, dass sie neben mehr oder weniger proprietären Resolver-Mechanismen auch ein Interface zum Ansprechen via http-URLs besitzen. So lassen sich bspw. Handles auch ohne entsprechendes Browser-Plugin auflösen.
Um das Resolving für die PIs auch dauerhaft zu gewährleisten, müssen u. a. folgende Szenarien bedacht werden:

- Repositorysysteme können sich ändern und (kleinere) Organisationen mit der Pflege der PI-Datenbank überfordert sein.
- Organisationen und Organisationsstrukturen können sich ändern. Die Verwendung von Organisationsbezeichnungen, oder -abkürzungen in Identifiern ist aus diesen Gründen nicht sinnvoll.
- Namensräume können sich während der Zeit auf mehrere Organisationen aufteilen (Namespace-splitting).
- Hersteller von Resolver-Systemen können über längeren Zeitraum die Pflege der Software einstellen. Die Nutzung von proprietären Software ist daher problematisch.

Ein langfristig persistentes Resolving kann daher nur auf einer übergeordneten Ebene erfolgen. Eine solche Initiative ist mit dem Name-to-Thing (N2T) Konsortium geplant, welches den N2T-Resolver betreibt. Dieser Resolver soll mit allen bereits vorhandenen Identifier-Systemen zusammenarbeiten, in dem der jeweilige Identifiertyp als „Namespace“ genutzt wird.

http://n2t.info/**NAA**/... (**NAA** = N2T NAA Number, eg, **12345**)
http://n2t.info/ark:/**NAA**/... (**NAA** = ARK NAA Number, eg, **12345**)
http://n2t.info/urn:**NAA**:... (**NAA** = URN Naming Authority, eg, **nbn**
http://n2t.info/hdl:**NAA**/... (**NAA** = Handle Naming Authority, eg, **12345**)
http://n2t.info/doi:**NAA**/... (**NAA** = DOI Naming Authority, eg, **10.12345**)
http://n2t.info/purl:/**NAA**/... (**NAA** = PURL Resolver Hostport, eg, **purl.org**)

Mittels Rewrite-Rules leitet der N2T-Resolver den Browser des Benutzers auf einen entsprechenden Resolver oder auf ein Dokument weiter.
Der N2T-Resolver soll in mehrfacher Version durch freiwillige Mitglieder des Konsortiums betrieben werden, um so die Zuverlässigkeit zu erhöhen und Ausfallwahrscheinlichkeiten zu minimieren. Da es sich hierbei um einen einfachen Webdienst handelt, der zunächst nur für einen kleinen Teil von Dokumenten überhaupt genutzt wird, sind die Kosten zum Betrieb des Systems relativ gering. Es ist aus diesem Grund nicht geplant, Gebühren zu erheben oder weitere Anforderungen an eine Mitgliedschaft in dem geplanten Konsortium zu stellen.

Bei ersten Analysen fiel besonders die unterschiedliche Struktur von PREMIS und LMER auf. Es bestehen z. B. Unterschiede in den Grundentitäten und in den Hierarchien. LMER besitzt eine flache Struktur, während PREMIS eine hierarchische Struktur mit semantischen Einheiten besitzt und mächtiger ist. Es erweist sich als nicht trivial, eine Rückumwandlung von einmal konvertierten Metadaten durchzuführen.

Es ist aber notwendig, schon bestehende Systeme in die Betrachtungen mit einzubeziehen, um vorhandene Datenbestände migrieren zu können. Deswegen sollten Mappings entwickelt werden, die den Weg von anderen Langzeitarchivierungsmetadaten zu PREMIS ebnen.

Handlungsvorschläge

- Entwicklung eines standardisierten Objektformates auf der Basis von PREMIS und METS, in dem sowohl Metadaten als auch (optional) Contentdaten abgelegt und ausgetauscht werden können.
- Entwicklung und Standardisierung eines gemeinsamen Minimalsets an PREMIS-Metadaten, die den Austausch von Langzeitarchivobjekten erleichtern.
- Entwicklung von Mappings zwischen PREMIS und LMER
- Entwicklung von Handreichungen und Best Practices für die Umsetzung und Implementatierung von entwickelten Standards.
- Durchführung von Schulungen und Workshops zur Verbreitung des entwickelten Wissens.

Markus Enders, Staats- und Universitätsbibliothek Göttingen

Dokumentenmodelle:

Ein großer Teil von Informationen aus dem „Buchzeitalter“ ist lediglich in Form von längeren Texten zugänglich. Das Digitalisieren dieser Texte bedeutet weitaus mehr als lediglich die Seiten digital abzulichten und für den Zugriff bereitzuhalten. Vielmehr ermöglicht erst die Beschreibung eines Dokumentes das Auffinden und Nutzen eines Dokuments. Die Nutzungsmöglichkeiten sind dabei stark abhängig von dem Dokumentenmodell. Ein Dokumentenmodell sieht ein Dokument nicht nur als Ganzes, sondern teilt – unabhängig von der verwendeten Technologie zum Speichern dieser Metainformationen – das Dokument in kleinere Einheiten mit ihren Relationen untereinander. Für die Digitalisierung lassen sich sehr unterschiedliche Dokumentenmodelle finden:

- globale Sicht auf das Dokument: Das Dokument besteht lediglich aus einer Datei; Metadaten sind nur auf bibliographischer Ebene vorhanden; es lassen sich keine internen Teile des Dokuments ansprechen. Dies ist das typische Dokumentenmodell von DMS-Systemen aus dem Office-Bereich.
- einfach strukturierte Sicht auf das Dokument: Neben bibliographischen Metadaten existiert eine Struktur, welche die inhaltlich, logische Gliederung (logische Struktur) eines Dokumentes abbildet. Der eigentliche Inhalt (Content-Files) sind direkt mit zugehörigen
- zweifach strukturierte Sicht auf das Dokument: Neben der logischen Struktur enthält das Dokument eine weitere Struktur, die die physischen Einheiten des Dokumentes abbildet: die gebundene Einheit als auch einzelne Seiten. Da es sich ebenfalls um eine hierarchisch organisierte Struktur handelt, lassen sich mit diesem Modell auch Einheiten innerhalb der Seite (also bspw. Spalten, Kästen etc.). Elemente aus beiden (logischer und physischer) Strukturen verknüpfen. Der eigentliche Inhalt (Content-Files) werde lediglich aus der physischen Struktur verlinkt.

Inhalte können also je nach Bedarf / Anwendungsszenario auf Basis unterschiedlicher Dokumentenmodelle erfasst werden. Da die Entwicklung eines Dokumentemodells – sowie die Ausarbeitung eines entsprechenden Speicherformats (in bspw. METS od. MPEG21/DIDL) - komplex sein kann, ist es sinnvoll, ein einmal entwickeltes Dokumentenmodell nachzunutzen. Voraussetzung dafür ist die Dokumentation dieser Modelle unter den Gesichtspunkten der Granularität (welche Einheiten innerhalb des Dokuments gibt es), der Relationen (wie hängen diese Einheiten zusammen) und der verwendeten Metadatenstandards (für deskriptive, technische und Langzeitarchivierungsmetadaten).
Für METS-Dateien stellen die „METS-Profiles“ eine Möglichkeit der Dokumentation über die Nutzung und Bedeutung der einzelnen Elemente da. Durch eine zentrale Registry wird diese öffentlich verfügbar und ist somit nachnutzbar.
Eine Dokumentation ist gerade auch für den Bereich der Langzeitarchivierung wichtig, da ggfs. die Dokumentation auf Basis von eher technisch orientierten Methoden

(wie bspw. XML-Schema) nicht ausreichend ist. Die Bedeutung und Nutzung von bspw. bestimmten XML-Elementen kann Variationen unterworfen sein.

Autor: Dr. Nicolas Flores-Herr

Güte der Digitalisierung von Bildern und Drucken

Konzept der Deutschen Nationalbibliothek

a. Optimale Abtastung der zu digitalisierenden Vorlagen. Speicherintensive Scanparameter (hohe Scanauflösung, hohe Farbtiefe) sind zunächst kein Qualitätskriterium, sondern bei unvorteilhafter Wahl ein großer Kostenfaktor bei der Archivierung der Digitalisate. Von daher müssen die Scanparameter derart gewählt werden, dass die digitale Rekonstruktion des Originals ohne Informationsverlust möglich ist. Es gilt die im Bild der Vorlage enthaltene Information derart zu digitalisieren, dass insbesondere sämtliche Details und der gesamte Farbumfang erfasst werden.

b. Vorlagenabhängige Scanparameter. Abschätzung und im Idealfall Messung des Farbumfangs und der maximalen Ortsfrequenz für bestimmte Vorlagentypen (z. B. Offsetdrucke). Die Scanparameter sind derart zu wählen, dass bei einer Erhöhung keine Qualitätsverbesserung messbar ist.

Druckvorlage	Rasterweite (f_{max})	Minimale Abtastrate(f_{abtast})
Zeitung	≤ 100 lpi	220 ppi
Zeitschrift	120 – 150 lpi	330 ppi
Buch	≤ 200 lpi	440 ppi

Auflösung nach dem Nyquist-Shannon Abtasttheorem $ppi_{Scanner} > 2 * lpi_{Vorlage}$

c. Berücksichtigung der Übertragungsfunktion (MTF) des Scanners statt des dpi-Wertes. Die Übertragungsfunktion (MTF) von Scannern (linke Abbildung) beschreibt deren Güte bei der Digitalisierung von Bildkontrasten (Ordinate linke Abbildung) in Abhängigkeit der Ortsfrequenz der Vorlage (Abszisse linke Abbildung). Durch die MTF wird die Erhaltung von Bildinformation der analogen Vorlage bei der Digitalisierung bestimmt. Im Bereich der Maximalauflösung sinkt typischerweise der Bildkontrast (Tiefpaßfilter). Durch Messung der MTF wird quantifizierbar, ab welcher Abtastrate (dpi) keine weiteren Bilddetails der Vorlage erfasst werden, sondern nur der Speicherverbrauch des Digitalisats steigt. Auf der rechten Seite sind drei Scans derselben alphanumerischen Scanvorlage (Target) abgebildet, die mit drei unterschiedlichen Scannern erzeugt wurden. Die jeweiligen MTFs der Scanner sind in das linke Diagramm eingezeichnet. Aus der rechten Abbildung wird deutlich, dass die Qualität der Digitalbilder von Scannern mit besserer MTF über den gesamten Ortsfrequenzbreich auch für das menschliche Auge klarer und schärfer wirken.

Im linken Diagramm sind die Übertragungsfunktionen (MTFs) dreier Scanner in Abhängigkeit der Ortsfrequenz aufgetragen. In der rechten Abbildung sind die mit den respektiven Scannern erzeugten Digitalbilder derselben alphanumerischen Testvorlage (Target) zu sehen. Die MTF beschreibt den

Kontrast (Ordinate) in Abhängigkeit des Detailgrades der Bilddetails (Abszisse; Ortsfrequenz). Je höher der Kontrastwert bei einer gegebenen Ortsfrequenz ist, desto höher ist die Güte der Übertragung von Details aus der Vorlage in den digitalen Bereich.

d. Keine höhere Auflösung durch Interpolation. Viele Scanner, auch von professionellen Herstellern, besitzen eine optische Auflösung (in dpi), die nicht alle Details der Vorlage erfassen kann. Um die geringe Auflösung der Optik zu kompensieren (und aus marketingstrategischen Gesichtspunkten), werden durch Interpolation zusätzliche Pixel berechnet, meist gefolgt von einer Kontrastanhebung und Schärfung. Durch die „Erfindung" von Pixeln, die der Scanner niemals „gesehen" hat, werden Ausgabeformate mit höheren dpi-Werten erzeugt. Interpolationsverfahren bringen keinerlei Vorteile, da es physikalisch unmöglich ist, Bilddetails aus der Vorlage herauszuarbeiten, die nicht von den Sensoren des Scanners registriert wurden. Im Gegenteil, der Speicherverbrauch des Digitalisats steigt quadratisch mit der Abtastrate bzw. mit dem am Scanner eingestellten dpi-Wert. Weiterhin werden die für die Interpolation gebräuchlichen Bildbearbeitungsschritte meistens nicht vom Hersteller dokumentiert; die Veränderung des Originalbildes ist nicht nachvollziehbar.

Diskussions- / Handlungsvorschläge

a. Erstellung einer Tabelle für die maximale Ortsfrequenz der Bilder der gängigsten Scanvorlagen und, daraus abgeleitet, minimale Abtastrate des Scanners.

b. Messung der Übertragungsfunktion (Modulation Transfer Function) des eingesetzten Scanners. Durch Messung der Übertragungsfunktion (MTF; nach ISO Norm) soll überprüft werden, ob der Scanner ausreichend Kontrast im Bereich maximaler Auflösung besitzt. Falls der Kontrast bei maximaler Auflösung zu niedrig ist, muss entweder ein anderer Scanner beschafft werden oder mit einer entsprechend niedrigeren Auflösung gearbeitet werden, da ansonsten Speicherplatz verschwendet wird. Im Zuge dieser Messung sollte ebenfalls überprüft werden, ob durch Änderung des dpi-Wertes eine Verbesserung des MTF errreicht werden kann. Falls durch Erhöhung des dpi-Wertes keine Verbesserung der MTF erzielt wird ist, wäre es für eine optimale Ausnutzung des Speicherplatzes sinnvoll, mit der niedrigsten Abtastrate, die gleichzeitig die bestmögliche MTF besitzt, zu digitalisieren.

Übertragungsfunktionen (MTFs) eines Scanners bei unterschiedlichen dpi-Werten. Durch die Erhöhung des dpi-Wertes wird keine Verbesserung der Auflösung erzielt, weil die MTFs nahezu identisch sind. Für Scans mit einer Abtastrate größer als 250 dpi wird nur der Speicherbedarf des Digitalisats erhöht. Ein Scan mit 500 dpi benötigt gegenüber einem Scan mit 250 dpi das Vierfache an Speicherplatz ohne dass mehr Informationen bzw. Bilddetails aus der analogen Vorlage digitalisiert werden.

c. Messung und Speicherung der MTFs für die verwendeten Scanner als technische Metadaten.

d. Mehr Qualität für Speicherplatz. Langzeitverfügbarer Archivspeicher für Digitalisate ist ein wichtiger Kostenfaktor bei der Bewahrung digitaler Kulturgüter. Von daher ist es angeraten, Digitalisate mit Scannern oder Digitalkameras zu erzeugen, die eine dokumentierte MTF besitzen. Die angegebene MTF des Herstellers ist selbstverständlich zu überprüfen. Sollte es Vorlagen geben (Handschriften oder Kunstdrucke) die aufgrund einer unzureichenden MTF nicht durch einen vorhandenen Scanner digitalisiert werden können, sollte überprüft werden, inwieweit digitale Spiegelreflexkameras bessere Übertragungsfunktionen besitzen oder ein anderer Scanner verwendet werden kann. Durch die genaue Kenntnis des analogen Ausgangsmaterials und der MTF des Scanners können die Digitalisierungsparameter derart gewählt werden, dass die im Bild der Vorlage enthaltene Information ohne Qualitätsverlust mit minimalem Speicherverbrauch digitalisiert werden kann. Auf diese Weise wird die Handhabe der Digitalisate komfortabler (z. B. CPU-Last bei Bildbearbeitung, Netzwerkauslastung oder Bereitstellung im Internet) und die Kosten für Langzeitarchivierung werden drastisch gesenkt.

Christine Frodl, Deutsche Nationalbibliothek
Digitalisierung und digitale Sammlungen

Thesenpapier zu Punkt 3.2 Metadatenformate - Praxisbericht DNB zu Online-Hochschulschriften – MetaDiss und XMetaDiss

1. Kurze organisatorische Angaben zum Projekt / zur Institution

Christine Frodl
Deutsche Nationalbibliothek
Arbeitsstelle für Standardisierung
Tel.: 069 – 1525 -1404
E-Mail: c.frodl@d-nb.de

Die Deutsche Nationalbibliothek[i] ist die zentrale Archivbibliothek und das nationalbibliografische Zentrum der Bundesrepublik Deutschland. Die Arbeitsstelle für Standardisierung[ii] leistet die hauptamtliche Facharbeit im Bereich der bibliothekarischen Standardisierung für Deutschland, Österreich und die Schweiz. Sie vertritt die Interessen in Gremien auf nationaler und internationaler Ebene und ist als Organisationseinheit der Deutschen Nationalbibliothek für die Leitung und Koordination der Arbeit in den Teilbereichen Regelwerke, Normdateien, Datenformate und Kommunikationsschnittstellen sowie für die Geschäftsführung des Standardisierungsausschusses verantwortlich. Online-Hochschulschriften[iii]: Seit Juli 1998 sammelt die Deutsche Nationalbibliothek Online-Dissertationen und -Habilitationen, und betreut im Rahmen der Aktivitäten der Koordinierungsstelle DissOnline die Metadatenformate MetaDiss und XMetaDiss[iv] sowie die Transferschnittstelle für Online-Hochschulschriften.

2. Problemaufriss

Seit 1998 wird das im Rahmen eines DFG-Projekts entwickelte Metadatenformat für Online-Hochschulschriften MetaDiss für den Austausch von Metadaten und dazugehörigen Volltexten in Deutschland verwendet. Dieses Format basiert auf dem Dublin-Core-Metadaten-Element-Set von 1998 und auf HTML 4. Da die Einbindung in HTML nicht mehr dem Stand der Technik entsprach, wurde im August 2004 das neue, mit den Hochschulbibliotheken abgestimmte Format XMetaDiss entwickelt. Es basiert auf XML, UTF-8 und dem aktuellen Dublin-Core-Standard[v]. Das Potenzial von XMetaDiss besteht unter anderem in der Nutzung von hierarchischen Strukturen und der damit verbundenen Vermeidung von Zuordnungsfehlern, in der Schaffung eines Formates für Online-Hochschulschriften, das per OAI-Protokoll ausgetauscht werden kann und in der Schaffung eines zu internationalen Standards kompatiblen Formates, damit deutsche Online-Hochschulschriften in internationale Metadatensuchmaschinen eingebunden werden können.

3. Standardisierungsrelevante Aktivitäten

Die Deutsche Nationalbibliothek hat auf Basis des Formats MetaDiss und gemeinsamer Absprachen zwischen den deutschen Hochschulbibliotheken im Jahre 2004 das Metadatenformat XMetaDiss vorgelegt. Es liegt als Referenzbeschreibung in der Version 1.3 in deutscher[vi] und englischer Sprache vor und basiert auf dem weltweit verbreiteten Dublin-Core-Standard. Des Weiteren wurde ein Mapping von XMetaDiss zum ETD-MS-Standard[vii] der Networked Digital Library of Theses and Dissertations[viii] (NDLTD) angefertigt, das es ermöglicht, deutsche Hochschulschriften im NDLTD-Gesamtkatalog zu verzeichnen. Im März 2006 ging ein Vorschlag an das NDLTD Standards Committee bezüglich der Standardisierung der Roles von Beteiligten an Online-Hochschulschriften wie z. B. Committee member, Chair, Co-Chair, Referee etc., der positiv aufgenommen wurde.

Weiterhin ist auf eine neue Entwicklung, das „ePrints Application Profile"[ix] hinzuweisen, das auf einer Kombination der FRBR[x] und dem Dublin Core Abstract Model[xi] basiert und wissenschaftliche Netzpublikationen, die in institutionellen Repositories Großbritanniens vorgehalten werden, also auch Online-Hochschulschriften, beschreiben soll.

4. Handlungsvorschlag

Mittelfristig Entwicklung eines gültigen, standardisierten, international anwendbaren Formats zur Beschreibung und zum Austausch von Metadaten und Volltexten deutscher und internationaler Hochschulschriften in Zusammenarbeit mit dem Standards Committee der Networked Digital Library of Theses and Dissertations.

Die Entwicklung sollte weltweit etablierte Standards wie z. B. Dublin Core und den ETD-MS-Standard der NDLTD, aber auch national in der Anwendung befindliche Formate wie XMetaDiss, TEF[xii] (Théses électroniques francaises) etc. berücksichtigen. Dabei sollten auch die gegenwärtigen Arbeiten am ePrints Application Profile Berücksichtigung finden.

[i] http://www.d-nb.de
[ii] http://www.d-nb.de/standardisierung/afs/afs.htm
[iii] http://www.d-nb.de/wir/projekte/koordinierungsstelle_dissonline.htm
[iv] http://www.d-nb.de/standards/xmetadiss/xmetadiss.htm
[v] http://www.dublincore.org/documents/dcmi-terms/
[vi] http://www.d-nb.de/standards/pdf/ref_xmetadiss_v1-3.pdf
[vii] http://www.ndltd.org/standards/metadata/current.html
[viii] http://www.ndltd.org/info/description.en.html
[ix] http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile
[x] http://www.ifla.org/VII/s13/frbr/frbr.pdf
[xi] http://www.dublincore.org/documents/abstract-model/
[xii] http://www.abes.fr/abes/documents/tef/recommandation/tef.pdf

Dr. Nikola Korb, Deutsche Nationalbibliothek

Uniform Resource Name (URN)
-Struktur und Anwendung-

1. Organisatorische Angaben zum Projekt/zur Institution

Dr. Nikola Korb
Deutsche Nationalbibliothek
Abteilung Informationstechnik
Tel: 069/1525-1703
E-Mail: n.korb@d-nb.de

2. Problemaufriss

URNs sind Persistent Identifier und können die URLs in Nachweissystemen ersetzen und als stabile Referenzen in den Objekten selbst angewendet werden. Der Pflegeaufwand für die Aktualisierung der Verweise reduziert sich. Digitale Publikationen haben einen weltweit eindeutigen Identifier und sind damit zuverlässig zitierbar und über ihre gesamte Lebensdauer zu identifizieren. Auch im Rahmen der Langzeitarchivierung lässt sich durch die Verwendung von URNs eindeutig ein bestimmtes Objekt identifizieren. Eine URN besteht aus mehreren hierarchisch aufgebauten Teilbereichen. Dazu zählen der Namensraum (Namespace, NID), der sich aus mehreren untergeordneten Unternamensräumen (Subnamespaces, SNID) und einer eindeutigen Nummer (Namespace Specific String, NISS) zusammensetzen kann wie z. B. "URN:NID:SNID-NISS". URNs, die von der Deutschen Nationalbibliothek administriert werden, haben die folgende allgemeine Struktur: urn:nbn:de:[Verbundabk.]:[Sigelnummer]-[Nummer][Prüfziffer] bzw. urn:nbn:de:[Verlagsnummer]-[Nummer][Prüfziffer]

urn demonstriert die Auflösungsroutine, "nbn" bezeichnet den internationalen Namensraum "National Bibliography Number" (NBN), ein Namensraum, der von den Nationalbibliotheken kooperativ verwaltet wird. "de" steht für den Unternamensraum "de", für den die Deutsche Nationalbibliothek verantwortlich ist. [Verbundabk.] ist das Buchstabenkürzel des jeweiligen Bibliotheksverbundes, [Sigelnummer] das Bibliothekssigel der Bibliothek bzw. ein vierstelliger Zahlencode für Institutionen außerhalb der Bibliotheksverbünde. [Nummer] muss eine eindeutige (Produktions-) Nummer sein. Die [Prüfziffer] dient der Validierung des gesamten URN-Strings.

3. Standardisierungsrelevante Aktivitäten

- Strategien zur URN-Vergabe; vgl. dazu http://nbn-resolving.de/urn:nbn:de:1111-200606299 sowie http://www.persistent-identifier.de/?link=860
- Internationale Entwicklungen http://www.persistent-identifier.de/?link=510

Ziel: Infrastruktur zur „Browser-unterstützten“ URN-Auflösung aufbauen (vgl. RFC2276, 3401-3404)

4. Handlungsvorschlag

DIN für RFC 2141 - URN Syntax (Category „Standards Track“);
DIN für RFC 3406 - Namespace Definition Mechanism (Category “Best Current Practice”)

Prof. Dr. Angelika Menne-Haritz, Bundesarchiv

Die Präsentation von Digitalisaten aus Archivgut: Das Projekt <daofind>

Das Projekt daofind ist ein Projekt des Bundesarchivs. Es wurde von November 2004 bis Oktober 2006 mit Unterstützung durch die Mellon-Foundation, New York, durchgeführt. Sein Ziel war es, in einer Pilotanwendung die kombinierte Nutzung der drei internationalen Standards

* EAD (Encoded Archival Description)
* EAC (Encoded Archival Context)
* METS (Metadata Encoding and Transmission Standard)

für einen verbesserten Zugang zu Archivgut aus deutschen Archiven über das Internet zu testen.

Die Ergebnisse, die auf der Website des Projektes vorgestellt werden, sind die prototypische Beständeübersicht in EAD und EAC mit Links zu den 3 in EAD kodierten Findbüchern. Die Findbücher führen zu den digitalisierten Beständen mit 30.000 Images. Die Präsentation der Digitalisate wird über METS gesteuert.

EAD wird für die Herstellung und Präsentation von strukturierten Online-Findbüchern verwendet, in denen durch die Gliederung wie in einem Inhaltsverzeichnis navigiert, nach Begriffen gesucht, Indexbegriffe genutzt und der Text durchblättert werden kann. Die Elemente und Attribute der EAD-DTD bezeichnen typische Teile von archivischen Findbüchern. Mit XSLT-Stylesheets können sie in beliebiger Form präsentiert werden.

EAC hilft bei der Nutzung verschiedener Namen und Bezeichnungen von Personen, Organisationen oder Orten, indem verschiedene Formen mit Erklärungen ihrer Verwendung in den Beständen für die Recherche angeboten werden können. EAC beschreibt die typische Struktur der expliziten Kontextinformationen in archivischen Erschließungsangaben auf der Ebene der Beständeübersicht oder der Findbücher.

METS erfasst und strukturiert alle erforderlichen Angaben für digitalisiertes Archivgut und unterstützt die Verwaltung digitaler Objekte, die aus einer oder mehreren Dateien oder aus zusammengehörigen Gruppen von digitalen Reproduktionen bestehen können und verknüpft sie mit den zugehörigen Online-Findbüchern.

Die gleichen Findbücher und die vollständige Beständeübersicht sind in der Suchmaschine des Bundesarchivs MidosaSEARCH zusammen mit anderen Online-Findbüchern übergreifend recherchierbar. Die Images werden in der nächsten Version ebenfalls durchsuchbar integriert. Die Suchmaschine erlaubt gleichzeitig und in ständigem Wechsel eine strukturierte Navigation, eine Volltextsuche, die Nutzung von bei der Erschließung ausgewählten Indexbegriffen und ein Blättern wie in einem E-book. Fundstellen der Volltextsuche werden im Kontext aufgeblättert.

Ergebnis des Projektes ist neben der Pilotanwendung für die Präsentation von Images in einer speziell für Schriftgut entwickelten Form der Prototyp eines Werkzeuges für die Erstellungen der kombinierten Präsentation unter Nutzung der drei internationalen Standards am archivischen Arbeitsplatz. Basierend auf Eclipse bietet es unter Nutzung von Java und XML sowie XSLT eine fachsprachliche Umgebung mit allen erforderlichen Funktionen einschließlich der HTML-Generierung an.

Die Projektseite www.daofind.de stellt neben Berichten aus dem Projekt Informationen und Übersetzungen zu den drei Standards bereit.

Heike Neuroth (SUB-Göttingen): Metadatenformate

Angaben zur Institution:

Die SUB Göttingen ist seit ca. 10 Jahren mit der Entwicklung von Metadatenformaten beschäftigt und ist im Rahmen des vascoda-Projekts für das Kompetenzzentrum Metadaten, Standards und Heterogenitätsbehandlung verantwortlich.
Im Laufe der letzten Jahre sind unter der SUB Federführung eine Reihe von sog. Application Profiles entwickelt worden:

- VLib Application Profile
- ELAN Application Profile
- Vascoda Application Profile, vMIF (zurzeit in der Version 2.0)
- Vascoda Collection Level Application Profile

In diesen Entwicklungs- und Forschungsschwerpunkt ist auch das DFG Projekt KIM („Kompetenzzentrum Interoperable Metadaten" – Aufbau eines deutschsprachigen DCMI Affiliates) anzusiedeln, dass seit Oktober 2006 von der SUB geleitet wird.
Mit dem Projekt „Kompetenzzentrum Interoperable Metadaten" (KIM) – Aufbau eines deutschsprachigen DCMI Affiliates wird der Versuch unternommen, nicht einzelne Metadaten-Elemente zu definieren.

Problemaufriss

Metadaten spielen in der heutigen (digitalen) Wissens- und Informationsgesellschaft eine zunehmend wichtige Rolle. Sie dienen der Beschreibung, Verwaltung, Recherche, Nachweis und dem langfristigen Zugriff auf digitale Objekte bzw. Ressourcen.

In vielen Anwendungen und lokalen Projekten werden domainspezifische Metadatenformate (sog. Application Profiles) entwickelt, die mehr oder weniger nationale oder internationale Metadatenformate wie z. B. das Dublin Core Metadata Element Set zugrunde legen. Allerdings zeigt sich, dass auch bei Berücksichtigung nationaler Standards die Metadatenformate wenig interoperabel und damit die Sammlungsbeschreibungen der (digitalen) Objekte nicht oder nur mit hohem Aufwand austauschbar sind (vgl. auch aktuelle Metadatenlieferungen an vascoda). Ohne ein Mindestmaß an (semantischer) Interoperabilität wird es kaum gelingen, Portale mit nationalen, europäischen oder internationalen Informationen aufzubauen oder auch nur in Richtung Semantic Web zu gelangen. Diese Interoperabilität kann durch ein interoperables Metadatenmodell, wie es das DCMI Abstract Model vorschlägt, erreicht werden. Auch aktuelle Entwicklungen im Rahmen der internationalen Diskussionen um AACR bzw. RDA gehen in diese Richtung.

Standardisierungsrelevante Aktivitäten & Empfehlungen

- Etablierung einer KIM „Kernarbeitsgruppe", verantwortlich z. B. für die Formulierung von Empfehlungen für interoperable Metadatenmodelle, Workshops etc.
- DIN Zertifizierung von DCMES
- Deutschsprachige Übersetzung des DCMI Abstract Models
- DIN Zertifizierung von nationalen Aplication Profiles (z. B. vascoda, XMetaDiss, ELAN)

- W3C Semantic Web Developments: Berücksichtigung deutscher Entwicklungen, Nachnutzung von Standards in Deutschland …

Tobias Steinke, Deutsche Nationalbibliothek

UOF und LMER

1. Projektbeschreibung

Im BMBF-geförderten Projekt kopal wird kooperativ ein Langzeitarchiv digitaler Informationen aufgebaut. Unter der Leitung der Deutschen Nationalbibliothek und mit den Partnern SUB Göttingen, GWDG und IBM soll in der Projektlaufzeit von 2004 bis 2007 eine nachnutzbare Lösung entstehen, die für alle Arten von digitalen Objekte geeignet ist und keine Einschränkungen bei Dateiformaten vorsieht. Kern ist das System DIAS von IBM, welches vom Dienstleister GWDG in Göttingen betreut wird. Das Generieren der Archivobjekte, das Einspielen und der Zugriff erfolgt über die von DNB und SUB entwickelte Open-Source-Software koLibRI.
Weitere Informationen: http://kopal.langzeitarchivierung.de

2. Problemaufriss

Eine wichtige Schlüsselkomponente für die spätere Durchführung von Formatmigrationen zur Langzeitverfügbarkeit ist die Gewinnung von Metadaten zur Langzeitarchivierung. Dazu gehören technische Informationen zu den einzelnen Dateien, aber auch eine Historie der erfolgten Dateiformatänderungen. Ein Archivobjekt sollte zum Zeitpunkt der Archivierung diesen „Beipackzettel" haben und auch beim Austausch mit anderen Langzeitarchiven mitführen. Ähnlich Dublin Core bei den beschreibenden Metadaten wird ein einheitlicher Kern von Langzeitarchivierungsmetadaten benötigt. Zudem soll-te eine möglichst universelle Objektstruktur für die Archivierung und den Austausch vereinbart werden, die einerseits eine verlässliche automatische Verarbeitung inklusive der Metadaten zur Langzeitarchivierung ermöglicht, andererseits flexibel genug ist, um auch zukünftig weitere Informationen aller Art aufzunehmen.

3. Standardisierungsrelevante Aktivitäten

Die Deutsche Nationalbibliothek hat 2003 auf Basis eines Datenmodells der Nationalbibliothek von Neuseeland die Spezifikationen der Langzeitarchivierungsmetadaten für elektronische Ressourcen (LMER) vorgelegt. In Zusammenarbeit mit Regionalbibliotheken und Bibliothekszentren wurde die erste Version weiterentwickelt und liegt derzeit in der Version 1.2 vor. Im Projekt kopal wurde eine generische Struktur für Archivobjekte erstellt, die auf LMER und dem Standard METS der Library of Congress (LoC) basiert. Diese Struktur wurde unter dem Namen Universelles Objektformat (UOF) 2005 veröffentlicht. 2006 wurde für die Verwendung von METS im UOF ein eigenes METS-Profile bei der LoC registriert. Mit koLibRI liegt eine freie Software zur Erzeugung und Verarbeitung von Objekten im UOF vor.
LMER: http://www.d-nb.de/standards/lmer/lmer.htm
UOF: http://kopal.langzeitarchivierung.de/index_objektspezifikation.php.de

4. Handlungsvorschlag

- LMER 1.2 als genormter Kern von Metadaten zur Langzeitarchivierung
- UOF als Basis für den genormten Objektaustausch zwischen Langzeitarchiven

Ralf Stockmann (SUB-Göttingen): Qualität digitalisierter Kulturgüter

(GDZ-Erfahrungen)

Angaben zur Institution:

Das GDZ wurde 1997 mit Unterstützung der DFG als eigenständige Abteilung an der Niedersächsischen Staats- und Universitätsbibliothek in Göttingen eingerichtet. Dabei handelt es sich um eines der beiden Kompetenzzentren für Bibliotheken und wissenschaftliche Einrichtungen, die konzeptionelle und technische Beratung auf dem Gebiet der Digitalisierung von Datenbeständen anbieten. Die Mitarbeiter des GDZ bieten folgende Leistungen an: Beratung bei Digitalisierungsprojekten / Konzeption von Digitalisierungsvorhaben (Work-flow/Arbeitsschritte) / Digitalisierung, Bildbearbeitung und Qualitätssicherung / Strukturelle Erschließung von Digitalisaten / Integration der Digitalisate in das eigene digitale Archiv inkl. Web-Zugang / Softwareentwicklung.
Seit Bestehen wurden im GDZ rund 5 Mio. Seiten digitalisiert, ca. 90 % davon bitonal.

Problemaufriss

Entscheidend für die bei der Digitalisierung anzulegende Qualität ist die Frage nach den späteren Nutzungsarten:

- Inhalt vollständig lesen/verstehen können
    - o Am Bildschirm (Auflösungen steigen!)
    - o Im Ausdruck (Laserdrucker / Tintenstrahl)
- „Anmutung“ reproduzieren
    - o Welche Anmutung ist gemeint?
    - o „Original“ = bitonal?!
- Volltexterfassung
    - o Double key (unkritisch wenn lesbar)
    - o OCR (Abhängigkeit von Auflösung, Farbtiefe, Verschmutzungsgrad)
- Reprint / Faksimile

In einigen Bereichen (Farbdigitalisierung!) besteht ein (derzeit) nicht aufhebbarer Widerspruch zwischen Qualitätsanspruch und Massendigitalisierung.

Standardisierungsrelevante Aktivitäten

- Einsatz nur hochwertiger (Buch)Scanner (Zeutschel, I2S)
- Schlechte Erfahrungen mit Outsourcing (hier könnte eine Standardisierung besonders helfen!)

- Wenn es die Vorlage zulässt: 600dpi bitonal
- Aufwändige Nachbearbeitung: (De)Speckling, Rotation, Ausrichtung, Seitengrößenanpassung
    - Bitonal: im Badgebetriete, Farbe: händisch
- Qualitätskontrolle aller Seiten (häufige Fehlerbilder: fehlende/doppelte/verschobene Seiten, Belichtung, Farbwahl)
- Software für verteilten Workflow (Goobi)

Empfehlungen

- DIN-Zertifizierung sehr wünschenswert, um vergleichbare Qualitätsstandards herzustellen
- Mehrere Qualitätsstufen notwendig, abgestimmt auf Nutzungsszenarien
- Ausgesprochen hilfreich um Vergleiche mit externen Dienstleistern herzustellen (in Kombination mit Dokumenten- und Metadatenmodellen)

Wolf-Klostermann

Metadaten zur Langzeitarchivierung

Beispiel eines Objektmodells

Hintergrund: Projekt und Institution

DFG-Projekt der Bayerischen Staatsbibliothek in Kooperation mit dem Leibniz-Rechenzentrum München (2005-2007). Aufbau einer organisatorischen und technischen Infrastruktur für die Langzeitarchivierung von Netzpublikationen aus dem breiten Spektrum einer Universalbibliothek - Bibliothekarisches Archivierungs- und Bereitstellungssystem (BABS) auf der technischen Basis von DigiTool (ExLibris) und TSM (IBM). Maßnahmen zu Langzeitarchivierung der Dokumente und Entwicklung von Bereitstellungslösungen.

Grundprinzip

Zerlegung komplexer Objekte und Verwaltung auf der Ebene einzelner Dateien (Datei Objekt digitale Entität). Verwaltung von beliebigen „Manifestationen“ eines Objekts. Zusammenlegung von Einzelobjekten zu „komplexen Objekten“. Zuordnung von Metadatensätzen sowohl auf Einzelobjektebene als auch für komplexe Objekte.

Aufbau einer digitalen Entität

Element	**Funktion/Umsetzung**
PID	Identifkator
Control Section	Verwaltungsdaten
Metadata Section	
descritptive	MARC oder Dublin Core
technical: text/image/audio/video	PREMIS, JHOVE
accessrights	
copyrights	METS RightsDeclarationMD
changehistory	PREMIS
preservation	URN mit NBN
filestructure	METS
Relations (part-of, includes, manifestation)	Relation zu anderen Objekten über PID
Stream ref	zugeordnete Daten

Diskussions-/Handlungsvorschlag

a) Der richtige Zeitpunkt für eine Standardisierung: Die verwendeten Objekt- und Metadatenmodelle sind relativ jung und kaum in Gebrauch. Erfahrung.

b) Austauschbarkeit entscheidend: Transformation in andere Objektmodelle muss definiert werden, z. B. UOF. BA/ES - Digitale Bibliothek 2./3.11.2006 Standardisierungsbedarf im Kontext Langzeitarchivierung und Digitalisierung Stand: 31.10.2006 09:15 ● TWK ●

c) Sinn und Zweck bestimmter MD ist zu hinterfragen. Ist z. B. die Extraktion der technischen Metadaten immer sinnvoll?

Stefan Wolf, Bibliotheksservicezentrum Baden-Württemberg in Konstanz (BSZ)

Digitalisierung und Digitale Sammlungen, 3.November 2006:

3. Metadatenformate // 3.3 BSZ Baden-Württemberg: Metadaten

Beschreibende Metadaten werden aus vielfältigen Zusammenhängen an das Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) geliefert:

- Monographienkatalogisierung: 12 Mio Titelaufnahmen sind nach RAK/MAB in der Verbunddatenbank des Südwestdeutschen Bibliotheksverbundes (SWB) unter einem Format vereinigt.
- Daten aus der Museumsdokumentation
- Daten aus den Findbüchern der Archivverwaltung
- ...

Die Differenzierung und Strukturierung der Metadaten erfolgt nach vorhandenen Standards.

Vorhandene Standards werden konsequent genutzt:

- XMetaDissPlus als Format für Beschreibung und Transfer von Metadaten zu Hochschulschriften (im weiten Sinn) stützt sich semantisch und strukturell auf XMetaDiss als dem maßgeblichen Format für Dissertationen und Habilitationen
- XMetaDissPlus-Daten können der weiteren Verarbeitung zugeführt werden.
- Digitalisierte Publikationen und Digitale Sammlungen werden sachgerecht bibliographisch in der Verbunddatenbank erschlossen.

Spezialbedarf wird den vorhandenen Standards eingegliedert.

Die Pflege der Metadaten erfolgt prinzipiell und ausschließlich auf dem fachoriginären Verzeichnungsinstrument.

Die Leistungsfähigkeit einer Installation oder Institution bei der Standardisierung von Metadaten entscheidet sich daran, mit welcher Qualität sie ohne Verlust von Differenzierungsgraden oder standardisierten Vorleistungen (Normdateien) unterschiedlich strukturierte oder „dialektal" gefärbte Varianten eines Metadatenformates in ihren für langfristige Pflege, Verwaltung und Suche vorgesehenen Datenbestand eingliedern kann.

Zu unterscheiden ist immer zwischen

- Datenhaltungsformat
- Datentausch/-lieferformat:
- Verpackungsformat (z. B. METS), das weitere Funktionsbereiche einbezieht.

Standardisierung stößt an Grenzen: diese

- liegen in der Autonomie und dem immer vorhandenen Spezialbedarf der beschreibenden Einrichtungen;
- sind darin begründet, dass der Horizont der interessierenden Objekte in den interessierten Institutionen dynamisch ist und immer dynamischer wird.

Die Zukunft aller Metadatendarstellung wird ausschließlich in XML-Darstellungen liegen, die durch XML-Schema definiert sind.

Wolf-Klostermann

Dateiformate und Dokumentenmodelle
Thesenpapier

Konzept der Bayerischen Staatsbibliothek

a) Nur ganze Werke werden digitalisiert, erschlossen und publiziert, keine fragmentarische Digitalisierung.

b) Unterscheidung zwischen Archivformat und Präsentationsformat: Digitalisierung und Langzeitarchivierung erfolgt in bestmöglicher Qualität. Erzeugung digitaler und analoger Derivate nach Bedarf.

c) Grundsätzlich mit Referenzierung: *Kein Digitalisat ohne Nachweis* im Katalog und in anderen Portalen und Suchsystemen. URN-Vergabe. Ggf. Sperrfrist.

d) Einheitliche Workflows zur Erschließung für alle Digitalisate: Bibliographischer Nachweis aus dem Katalog, eindeutiger Dateiname, Erzeugung einer XML-Strukturdatei und von Präsentationsderivaten. URL/URN-Eintrag in Katalog, Langzeitarchivierung.

e) Unterschiedliche Erschließungstiefen möglich, woraus sich die Komplexität der Dokumentstruktur ergibt: 1) Reine Blätterversion (Minimallösung) 2) Aufnahme von Strukturdaten (Kapitel, Illustrationen, Register) 3) Volltexterschließung. Dabei werden universelle Formate genutzt (METS-Strukturdatei, TEI-XML).

f) Authentische Präsentation wird angestrebt: Nachempfindung der analogen Nutzung („Blättern"), immer mit den Images im Vordergrund. Keine Korrektur von „Original"-Fehlern, keine Kurzfassungen oder sonstige Bearbeitungen.

g) Langzeitarchivierung: Objekt wird grundsätzlich als digitales Master zusammen mit der XML-Strukturdatei archiviert.

Diskussions- / Handlungsvorschläge

a) Dateiformate. Unterscheidung von Archivexemplar und Präsentationsderivaten. Festlegung von Archivformaten vorrangig, Präsentationsformate z. B. JPG, PDF. Beschränkung auf wenige nicht-proprietäre oder lange eingeführte Dateiformate.

b) Archivierung in bestmöglicher Qualität.

c) Erschließungstiefe: Formale Mindestanforderungen für die Erschließung (Navigation, Blätterfunktion) bzw. Typisierung von Erschließungstiefen.

d) Dokumentmodell/Datenstrukturen: Strukturen müssen auch bei Verlust des Referenzsystems leicht rekonstruierbar bleiben. Verwendung selbsterklärender Strukturen für Objektmodell und bibliographische, administrative und technische Metadaten. Gekapselte

Speicherung von Objekt und Metadaten. Formale Mindestanforderungen für die Gestaltung.

e) Verfügbarkeit: Nachweis in lokalen, regionalen, nationalen und internationalen Bibliothekssystemen und Portalen. Persistente Identifikatoren. Unbefristete Verfügbarkeit durch Langzeitarchivierung.

Wolf-Klostermann

Qualität digitalisierter Kulturgüter

Thesenpapier

Qualitätskonzept der Bayerischen Staatsbibliothek

a) Digitalisierungsparameter abhängig von Vorlage und Verwendungszweck:

Je nach Objekt Referenz-Scan mit Referenzmaterial (Farbkeil, Graukeil und Maßstab). Farbige Scans mit Farbmanagement an kalibrierten Geräten (ICC-Profil). Grundsätzlich keine elektronische Nachbearbeitung der Images (z. B. unscharf Maskieren, restaurative Retuschen). Masterformat: TIFF unkomprimiert, mit Metadaten im Header.

Handschriften	600 oder 400ppi, farbig, Farbmanagement,
Alte Drucke	400ppi, farbig, Farbmanagement, wenn ohne Abb. auch in Graustufen möglich
Drucke ab 1850	600ppi s/w (bitonal), wenn mit Abbildungen graustufig oder farbig bei 300ppi
Landkarten	400ppi oder 300ppi, farbig
Fotografien	400ppi, graustufig oder farbig
Autographen, Nachlässe, sonstige Objekte	wie Handschriften

b) Umgebungsbedingung und konservatorische Anforderungen von Bedeutung: Definiertes Vorgehen und Festlegung der erlaubten Hilfsmittel je Medientyp. Beteiligung des Instituts für Buchrestaurierung. Scanner mit neutralen Licht- und Wärmequellen. Raumklimatisierung.

c) Qualitätssicherung durch Automatisierung der wesentlichen Workflows: Voreinge-stellte Scan-Parameter je Objekttyp, skriptgesteuertes Post-Processing, zeitgesteuerte Archivierung mit Übertragungskontrolle.

d) Zwischen- und Endkontrolle: Qualitäts- und Vollständigkeitskontrolle während und nach der Produktion. Wartungs- und Testkonzept für die Produktionsumgebung.

Diskussions- / Handlungsvorschlag

a) Definition von *Mindeststandards* nach heutigem Stand der Technik: Farbmanagement, Dateiformate, Referenzmaterial. Formulierung muss möglichen künftigen technischen Entwicklungen Rechnung tragen.

b) Qualitätsdiskussion darf sich nicht auf Scan-Parameter beschränken. Ausweitung auf die Themen Postprocessing, konservatorische Anforderungen, sowie Bereitstellung und Verfügbarkeit (Siehe auch „Dateiformate und Dokumentmodelle“)

c) Qualität lässt sich nur über den Anwendungszweck definieren. Digitalisierung aus welchem Motiv?