

II. XML-basierte elektronische Publikationen in Universitätsbibliotheken - eine Frage von Dokumenttypdefinitionen?

Projekte, Ansätze und Ergebnisse eines NDLTD-Workshops zu Dokumenttypdefinitionen für Hochschulschriften

Susanne Dobratz

1. Einleitung

Die Frage, wie XML in Universitätsbibliotheken angewendet werden kann, um die elektronischen Hochschulpublikationen in hoher Qualität zur Verfügung zu stellen, zu archivieren und recherchierbar zu machen, war Thema eines NDLTD-Expertenworkshops,⁷ der vom 10. bis zum 12. Mai 2000 am Rechenzentrum der Humboldt-Universität zu Berlin stattfand. Entstanden ist die Idee eines Arbeitstreffens aus den Präsentationen verschiedenartiger XML-Projekte auf dem 3. NDLTD-Symposium im März in St. Petersburg/ Florida.⁸

Die Erkenntnis, dass weltweit eine Reihe verschiedenartiger Ansätze zur Nutzung SGML/XML-basierter Publikationskonzepte existiert und die Notwendigkeit, im Sinne einer international übergreifenden Recherche interoperabel zu sein, ließ Experten aus Finnland, Norwegen, Schweden, den USA, Frankreich, Portugal, Großbritannien und Deutschland nach Berlin kommen.

Das Ziel des Workshops war es, Gemeinsamkeiten und Unterschiede zwischen den einzelnen Dokumenttypdefinitionen zu identifizieren, ein Modell für die Nutzung von Standard-DTDs zu entwerfen und zu diskutieren, Erfahrungen in der Dokumentenkonvertierung auszutauschen, sich gegenseitig die entwickelten Werkzeuge zur Verfügung zu stellen und sich auf eine einheitliche Interpretation und Benutzung des „Dublin Core Metadatensatzes für Dissertationen“ zu einigen.

7 Siehe: <<http://dochost.rz.hu-berlin.de/epdiss/dtd-workshop/index.html>>

8 Siehe hierzu und zu NDLTD allgemein den vorangegangenen Bericht in diesem Heft.

2. Was ist XML?

Die *eXtensible Markup Language* (XML) ist eine erweiterte Auszeichnungs- oder Metasprache, die Regeln für die strukturelle Auszeichnung von Dokumenten unabhängig vom Ausgabemedium definiert. XML stellt eine reduzierte Version der *Structured Generalized Markup Language* (SGML) dar, die seit 1986 ein ISO-Standard ist. SGML hat im Bereich des Internetpublishing durch seine hohe Komplexität und durch die sehr teuren Bearbeitungswerkzeuge keine Breitenwirkung erzielen können und sich nur in begrenzten Bereichen, wie z. B. bei der Erstellung technischer Dokumentationen durchgesetzt. Die Grundphilosophie von SGML und XML besteht darin, eine strikte Trennung von Inhalt, Struktur und Layout eines Dokumentes vorzunehmen.

Die bestehenden Projekte stützen sich entweder auf den SGML-Standard (ISO 8879 mit Korregendum K vom 4.12.1997) oder auf die Definition des World Wide Web Consortiums (W3C) XML 1.0 vom 10.2.1998. Den Dreh- und Angelpunkt der vorgestellten Projekte bilden jedoch die Dokumenttypdefinitionen (DTDs). Eine Dokumenttypdefinition im Sinne von XML definiert Regeln oder Vorlagen, nach denen gleichartige Dokumente zu gestalten sind. Eine DTD beschreibt dabei das Inhaltsmodell einer Klasse von Dokumenten. Sie besteht aus folgenden Komponenten:

- einer Notationsdeklaration, die Notationen für Fremdformate, z. B. für Graphik (gif, jpeg), Mathematik (Tex, LaTeX), 3D-Objekte (VRML) und ähnlichem definiert, für die keine eigene XML-Dokumenttypdefinition existiert,
- einer Entitydeklaration, in der Zeichenersetzungen von Zeichen definiert werden, die nicht standardmäßig im benutzten Zeichensatz (z. B. ASCII) enthalten sind, z. B. wird Ü als Ü definiert,
- einer Elementtypdeklaration, in der die Hierarchie der Elementtypen aufgebaut wird, wobei es immer ein eindeutiges oberstes root-Element geben muss, dem die anderen Elementtypen untergeordnet werden. Elemente aus externen DTDs werden einfach in die Hierarchie unter Angabe der DTD und ggf. des zugehörigen Namespaces eingebunden. Namespaces sind ein Instrument, um gleichlautende Namen von Elementtypen aus verschiedenen DTDs eindeutig zuzuordnen,
- und einer Attributlistendeklaration, in der die den Elementtypen zugehörigen Attribute und deren zulässige Wertebereiche definiert werden.

Zur Formulierung einer DTD wird eine spezielle Syntax (nicht XML <...>) benutzt.

Das Publizieren von Dokumenten kann in XML auf zweierlei Art geschehen:

1. durch sogenannte „wohlgeformte“ Dokumente, für die keine DTD benötigt wird, die jedoch streng hierarchisch mit einem root-Element aufgebaut und vollständig ausgezeichnet (mit sogenannten „Tags“ versehen, also getaggt) sind;
2. kann eine Publikation mit sogenannten „validen“ oder „gültigen“ XML-Dokumenten erfolgen. Diese Dokumente benötigen eine DTD.

Zur Zeit befindet sich innerhalb des W3C eine Alternative zur Verwendung von DTDs in der Diskussion. XML-Schemas definieren ähnlich wie DTDs Grundstrukturen und Inhaltsmodelle für eine Klasse von Dokumenten. Mit dem Hintergrund der Speicherung in Datenbanken bieten Schemas zusätzlich die Möglichkeit, vordefinierte Datentypen aus der Programmiersprachen (JAVA)- und Datenbankwelt zu benutzen, z. B. Real, String, Byte, und auf vordefinierte Wertebereiche und deren Grenzen, z. B. minLength, MaxExclusive, scale, encoding, etc. zuzugreifen. Ein Vorteil von XML-Schemas ist der, dass zur Definition von Dokumentklassen im Gegensatz zu DTDs XML als Syntax benutzt wird.

3. XML-Projekte und Dokumenttypdefinitionen

Die Tatsache, dass die auf dem Markt befindlichen XML-Autorensysteme sich aus den verschiedenen Gründen noch nicht durchgesetzt haben, führte dazu, dass an jeder Universitätsbibliothek eine andere Strategie verfolgt wurde. Es wurden verschiedene DTDs entwickelt und unterschiedliche Werkzeuge erstellt und genutzt. Damit haben sich sehr unterschiedliche Strategien z. B. für die Konvertierung von Dokumenten aus den gängigen Textverarbeitungssystemen nach XML etabliert.

Die meisten dieser Projekte begannen 1995 - 1997, in einer Zeit, in der von XML zwar schon die Rede war, zu der es jedoch kaum Werkzeuge oder standardisierte XML-DTDs gab. Aus heutiger Sicht ist es daher dringend erforderlich, dass man diese heterogenen Eigenentwicklungen neu überdenkt und diese zugunsten bereits etablierter Standards anpasst. Dabei bildet die Überführung der strukturiert vorliegenden Daten in neue Strukturen das geringste Problem.

4. Der Workshop

Als Gastgeber eröffnete *Dr. Peter Schirmbacher*, Direktor des Rechenzentrums der Humboldt-Universität zu Berlin, die Veranstaltung mit einer Einord-

nung dieses Workshops in die Aktivitäten innerhalb der NDLT-Initiative und den Zusammenhang zur ETD-Initiative der UNESCO.

UNESCO-Clearinghouse⁹

Prof. Peter Diepold präsentierte eine Datenbank, die als Ergebnis einer Diskussion auf dem UNESCO-Workshop vom September 1999 in Paris¹⁰ beschlossen wurde. Diese Datenbank soll weltweit alle Experten, Projekte und Institutionen erfassen, die auf dem Gebiet des elektronischen Publizierens aktiv sind. Weiterhin sollen über diese Datenbank, die an die Struktur des deutschen Bildungsservers angelehnt ist, von verschiedenen Projekten und Personen entwickelte Materialien zur Nachnutzung erfasst und bereitgestellt werden.

DTDs für elektronische Dissertationen

Nach der Vorstellung aller DTDs¹¹ stellte sich heraus, dass sich alle DTDs der Workshopteilnehmer nach ähnlichen Prinzipien aufbauen. So besteht die klassische Dissertation (die in Form einer Monographie erscheint) aus 3 Hauptstrukturen: dem erweiterten *Titelblatt* mit Abstracts, Erklärungen etc., dem eigentlichen *Dissertationskorpus*, der Text, Bild, Ton, Video, Tabellen u.ä. einhalten kann, sowie den *Anhängen*, die die fachlichen Anhänge, die Bibliographie, sowie Danksagungen u.ä. enthalten.

Es wurden präsentiert:

- die *ETD-ML.DTD* der Virginia Polytechnic Institute and State University (Virginia Tech) von *Anthony Atkins*,
- die *DiML.DTD*, die innerhalb des deutschen Dissertationen Online-Projektes entwickelt wurde, von *Matthias Schulz*,
- die *TDM.DTD* der University of Iowa von *Phil Potter*,
- die *HutPubl.DTD* der Technischen Universität Helsinki von *Tuija Sonkkila*,
- die Nutzung der *TEI-Light.DTD* in Ann Arbor und Lyon von *Paul Schaffner* von der University of Michigan at Ann Arbor und *Viviane Bouletreau* von der Universität Lyon 2,
- die Nutzung der *ISOBook.DTD* an der Universität Oslo von *Havard Fosseng*,

9 <<http://dbs.schule.de/unesco/>> Bei Interesse tragen Sie sich bitte in diese Datenbank ein.

10 <<http://www.unesco.org/webworld/etd/>>

11 Materialien und DTDs findet man unter: <<http://dochoost.rz.hu-berlin.de/epdiss/dtd-workshop/cdrom/index.htm>>

- ein Konzept, welches die Nutzung einer *TEI-basierten DTD* mit Erweiterungen für die Naturwissenschaften vorsieht, stellte *Per Åkerlund* von der Schwedischen Universität für Landwirtschaften in Uppsala vor.

All diese Dokumenttypdefinitionen werden als sogenannte Autoren-DTDs angesehen. Dies bedeutet, dass es sich hier um DTDs handelt, die dafür entworfen wurden, dass Autoren sie direkt im Schreibprozess benutzen. Das Problem der Verfügbarkeit und der Leistungsfähigkeit (z. B. bei der Darstellung der getaggtten Dokumente) von XML-Autorensystemen, sowie die weite Ausbreitung von Textverarbeitungssystemen wie Microsoft Word verhindert jedoch bislang die Verbreitung eines echten XML-Authoring.

Eine weitere Ursache dafür, dass sich an den verschiedenen Universitäten jeweils andere DTDs durchgesetzt haben, ist in der fachlichen Ausrichtung dieser Universitäten zu suchen. So sind die Universitäten Lyon und Michigan, die die *TEI-Light.DTD* benutzen, vor allem geisteswissenschaftlich ausgerichtet. Gleiches gilt für die Universität Oslo. Probleme bei der Nutzung der *TEI.DTD* oder *DocBook.DTD* treten vor allem an den Universitäten auf, die eine starke naturwissenschaftliche Ausrichtung bedienen müssen, so z. B. die Humboldt-Universität Berlin, die Technische Universität Helsinki oder die Schwedische Universität für Landwirtschaften in Uppsala. Die Vertreterinnen aus Lyon und Helsinki erwähnten, dass eine Dissertation an ihren Standorten oft nur eine Zusammenstellung verschiedener Zeitschriftenartikel darstellt, die durch ein Deckblatt miteinander verbunden sind.

Weitere Ansätze zur Informationsstrukturierung

Martin Hess von der Universität Frankfurt/Main präsentierte einen völlig anderen Ansatz der Strukturierung. Ausgehend von den in Postscript vorliegenden Dokumenten werden über eine automatische Strukturerkennung die Überschriften mit den Seitenzahlen extrahiert und in einer SGML-Struktur gespeichert. Alle anderen Teile des Dokuments einschließlich des Volltexts werden ignoriert. Die Postscript-Seiten werden dann mit Ghostscript (bzw. für ältere, digitalisierte Dissertationen mit einer OCR-Software) erneut generiert und die so neu entstandenen Textseiten werden mit dem extrahierten Inhaltsverzeichnis verknüpft.

Fach-DTDs und Spezial-DTDs für Multimedia

Neben den Haupt-DTDs, die für größere Dokumente genutzt werden sollen, kam die Diskussion über die Nutzung bestimmter Spezial-DTDs auf. So verwenden einige der DTDs Standards wie das *CALS-Tabellenmodell* (z. B. *DiML*) oder eine dem *BiBTeX*-Standard nachempfundene Struktur (z. B. *ETD*-

ML, DiML) für die Auszeichnung des Literaturverzeichnisses. Hier besteht jedoch noch Einigungsbedarf. Der Vorteil der Verwendung von Standardmodulen in DTDs ist vor allem der, dass gängige Konvertierungs- und Bearbeitungsmechanismen sowie Werkzeuge oder Standardsoftware bereits zur Verfügung stehen, die sonst speziell programmiert und an die Eigenentwicklungen angepasst werden müssen.

Christof Steinbeck vom Max-Planck-Institut für chemische Ökologie in Jena gab eine Einschätzung des aktuellen Entwicklungsstandes und Verbreitungsgrades der Chemical Markup Language (CML). Die Version 1.0 der Markup-Sprache von *Peter Murray-Rust* und *Henry Rzepa* wurde 1999 formal veröffentlicht¹² und dient der Auszeichnung von 2- und 3-dimensionalen chemischen Strukturen. Defizite existieren noch bei der Beschreibung von Reaktionsgleichungen, kristallographischen und spektroskopischen Daten. Für die Auszeichnung makromolekularer Objekte wird von den Chemikern z. B. die *BioML.DTD* verwendet. Um jedoch die hohe Akzeptanz in der relativ kleinen Gemeinde der Markup-Chemiker nicht zu gefährden, wird CML einem kontrollierten Versionierungsprozess unterzogen. Viele freie Werkzeuge unterstützen CML bereits als Import- und Exportformat, kommerzielle Anbieter halten sich jedoch bisher bei einer Implementation zurück. Eine gleichartige Darstellung des Standes der MathML.DTD gab es leider nicht.

Susanne Dobratz gab einen Überblick über weitere DTD-Entwicklungen im Multimediabereich. Als nutzbare und bereits weit unterstützte Alternative zu Bitmaps im WWW hat sich die Scalable Vector Graphics (SVG) des W3C bereits partiell etabliert. Exportfilter von Corel Draw und Adobe Illustrator, sowie eine große Menge freier Tools existieren bereits und sind über das W3C zugänglich. Die Firma Adobe stellt ein freies Browser-Plugin zur Verfügung. In den Diskussionen wurde jedoch deutlich, dass SVG nicht als proprietäres Auszeichnungsformat für 2-dimensionale Graphikobjekte anzusehen ist, sondern eine gute Möglichkeit ist, um aus fachlich strukturierten Daten, die in CML, MathML o. ä. vorliegen, mittels XSLT¹³-Technik graphische Darstellungen für die Netzpräsentation in einem Vektorformat zu generieren. X3D (eXtensible 3 D) soll ein komponentenbasierter Nachfolger des VRML97- Standards werden und 2002 von der ISO standardisiert werden. Entwickelt wird diese Sprache von einem Industriekonsortium.

12 Journal Chem. Inf. Comput. Sci. 39 (1999) 928.

13 eXtensible Stylesheet Language Transformations: Mit Hilfe von XSLT kann man ein XML-Dokument, welches einer bestimmten DTD folgt, in ein oder mehrere neue XML-Dokumente umwandeln, die völlig anderen DTDs entsprechen.

Weiterhin wurden die Entwicklung und die Möglichkeiten der Synchronized Multimedia Integration Language des W3C (*SMIL*) als Synchronisationsinstrument für Medienpräsentationen vorgestellt, welches von vielen Medienplayern bereits unterstützt wird, so von Realplayer, Apple Quicktime, Grins, Soja und anderen. SMIL könnte sich zur Beschreibung von Inhalten von Videosequenzen in digitalen Publikationen einsetzen lassen. Die *VoiceML* des W3C befindet sich im Moment noch in einem sehr frühen Entwicklungsstadium und ist bisher noch nicht einsetzbar. Interessant wird *VoiceML* für die Anwendung in den Bereichen Sprachsynthese, Sprachgrammatiken und zur Repräsentation natürlicher Sprache sein. Für die Darstellung von Noten wurden 2 konkurrierende DTDs vorgestellt, *MML* (Universität Pretoria) und *MusicalML* (Connection Factory: eine niederländische Firma).

Konvertierungsstrategien

Die Frage nach der Konvertierung von Dokumenten aus unterschiedlichen Originalformaten nach SGML/XML wurde sehr ausgiebig besprochen.

Da an allen Universitäten Microsoft Word das meistgenutzte Textverarbeitungssystem war, konzentrierten sich die Ansätze auf die Konvertierung von Word .doc oder .rtf -Dateien. Folgende Strategien wurden dabei identifiziert:

1. Mit Hilfe des *MS SGML-Authors for Word*, eines Zusatzprogrammes zu Word, welches Microsoft anbietet, werden an der Virginia Tech und der Humboldt-Universität über Formatvorlagen für Word die Konvertierungen vorgenommen. Mit dieser Methode wurden an der Virginia Tech 2 Dissertationen konvertiert, an der Humboldt-Universität wurde bereits eine kritische Masse von 100 Dokumenten übertragen. An der Humboldt-Universität ist die Bibliothek für die Konvertierung zuständig, an der Virginia Tech University liegt sie in der Hand des Autors. Da die Humboldt-Universität hier die ausgefeilteste Methode besitzt, wurden von *Uwe Müller* und *Burghard Güther* die Word-Formatvorlage und die Konvertierung an einem Beispiel ausführlich vorgeführt.
2. Die freie Software *Majix* wird an der University of Iowa genutzt. Hier müssen Autoren, die an dem Pilotprojekt teilnehmen, ihre Dissertation selbst nach XML konvertieren. *Majix* 1.1 ist ein kostenloses Werkzeug auf JAVA-Basis, welches ausschließlich mit Microsoft erstellte RTF-Dateien nach XML umwandelt.
3. *FrameMaker+SGML* wird von der Technischen Universität Helsinki benutzt. Mit dieser Methode wurden allerdings 1997/1998 nur einige Dokumente mit Hilfe einer *FrameMaker*-Dokumentvorlage (einer EDD) konvertiert, da es sowohl bei den Autoren als auch bei der Bibliothek, die für die

Umwandlung verantwortlich ist, Akzeptanzprobleme bezüglich der DTD gab.

4. Ein auf dem *Omnimark*-System basiertes *rtf2sgml*-script findet an der University of Michigan Anwendung. Hier wurden bereits 20 Dissertationen von der Bibliothek nach SGML konvertiert. Die Universität Michigan hat bereits in vielen Digitalisierungsprojekten Erfahrungen mit Omnimark gesammelt. Omnimark ist ein programmiersprachenartiges System, welches speziell auf die Konvertierung von SGML-Dokumenten aus und in andere Dokumentformate sowie andere DTDs zugeschnitten ist. Da es seit einiger Zeit für den Bildungsbereich frei verfügbar ist, hat sich weltweit eine breite Nutzergemeinschaft etabliert, die über spezielle WWW-Seiten Scripte wie *rtf2sgml* austauscht. Auch die Universität Lyon konvertiert ihre Dokumente aus Word, Staroffice, WordPerfect über den RTF-Umweg mittels Omnimark-script nach XML.
5. Die Universität Oslo nutzt für die Konvertierung aus Word *Balise*, ein Werkzeug, welches ähnlich wie Omnimark arbeitet. Da das Projekt dort erst am Beginn steht, ist noch keine repräsentative Menge von SGML-Dokumenten entstanden.

Einen Eindruck der Problematik, wie man *LaTeX*-Dokumente nach XML, besonders jedoch nach MathML umwandeln kann, gab *Sebastian Rahtz* vom Rechenzentrum der University of Oxford. Dieses Problem wird für einige Wissenschaftsbereiche als sehr essentiell angesehen. Laut Sebastian Rahtz stellt eine derartige Konvertierung zwar ein großes Problem dar, das jedoch nicht als unlösbar anzusehen ist. Er stellte das TeX4ht-System vor und zeigte an Beispielen dessen Anwendung bei der Konvertierung von LaTeX nach MathML. Sein Fazit war, dass eine derartig komplizierte Überführung mathematischer Strukturen von keinem anderen als dem Autor des Dokuments selbst zu leisten ist, da die Anwendung von LaTeX oder TeX sehr unterschiedlich und oft nicht korrekt geschieht, so dass selbst bei dem Publizieren in LaTeX stilistische Hilfskonstrukte benutzt werden, um die Darstellung mathematischer Formeln entsprechend zu forcieren, anstatt die bereits vorgefertigten, semantisch orientierten LaTeX-Auszeichnungen zu benutzen. Sebastian Rahtz stellte ebenfalls klar, dass die von ihm vorgeschlagenen Wege nicht für eine Massenproduktion geeignet sind, sondern für jedes Dokument erneut angepasst werden müssen.

Da XML-Dokumente allein im WWW von der breiten Masse der Benutzer bisher noch nicht visualisiert werden können, ist es notwendig, eine Konvertierung in niedriger strukturierte Dateiformate wie HTML oder PDF vorzunehmen. An der Humboldt-Universität und an der Virginia Tech University benutzt man

hierzu in perl geschriebene Programme (etd2html: Virginia Tech, diml2html: Humboldt-Universität), um eine HTML-Ausgabe der Dokumente zu erzeugen. In Helsinki wurde zu diesem Zweck ein DSSSL-Stylesheet geschrieben. Die Document Style Semantics and Specification Language ist ein ISO Standard und gilt als die Stilkomponente zu SGML. Sie ist sehr komplex, aber auch sehr leistungsfähig, vor allem in der Erstellung von Printausgaben für SGML-Dokumente, ist aber auch zur Generierung von HTML-Output geeignet. Die University of Iowa arbeitet im Moment testweise mit einem ersten CSS (Cascading Style Sheet), wogegen die Universität Lyon mit einem XSL- (eXtensible Style Language) Stylesheet experimentiert, um eine Visualisierung direkt aus dem XML-Dokument heraus ohne statische Konvertierung nach HTML vorzunehmen. Das Problem bei den beiden letzten Ansätzen besteht im Moment noch darin, dass allein der Internet Explorer 5.0 und testweise Netscape/Mozilla6.0 als Standardbrowser in der Lage sind, XML mit CSS oder XSL-Stylesheets zu interpretieren und darzustellen. Daher arbeitet Lyon zusätzlich noch mit einem Script, um eine statische Konvertierung vorzunehmen.

Retrieval

Kerstin Zimmermann, Carl von Ossietzky-Universität Oldenburg, stellte einen Ansatz zur verteilten Suche in Dokumenten dar. Dabei wurde in die Details des *Harvest-Systems* eingeführt und die Fähigkeit des Systems zur Realisierung einer verteilten Suche in stark strukturierten Dokumenten dargestellt. Frau Zimmermann ging sehr ausführlich auf die im Fach Physik in Deutschland, Europa und der Welt etablierte Suche in dokumentformat-heterogenen Hochschulschriften in der Physik ein. Sie schloss ihren Vortrag mit einem Ausblick auf eine Möglichkeit, die Dokumentenserver der Teilnehmer über eine minimale Strukturierung in Form der Dublin Core Metadaten der Dokumente zusammenzuschließen. Dies könnte über eine Suchmaske wie die im Dissertationen Online-Projekt entwickelte Suchmaschine *THEO*¹⁴ geschehen.

Das Metadatenproblem

Eine intensive Diskussion wurde zum Thema der Speicherung und des Formates für dokumentbeschreibende Metadaten geführt. Zu Beginn der Diskussion stellte *Thorsten Bahne*, Universität Duisburg, den in Deutschland abgestimmten *Dublin Core Metadatensatz* des Dissertationen Online-Projektes vor. Hierbei ging es eher um die Diskussion der Semantik und Interpretation der einzelnen Felder als um eine Syntax der Darstellung. Eine Auseinanderset-

14 <<http://www.iuk-initiative.org/iwi/TheO/>>

zung zu der Thematik, ob man die zum Teil in den hochstrukturierten Dokumenten enthaltenen Metadaten wie Autor, Titel, etc. eher dokumentintern oder als zusätzliches Dokument neben der eigentlichen Publikationen verwalten und archivieren sollte, war bereits im Vorfeld in Gang gekommen. Obwohl die TEI-Light.DTD, die HUT-Publ.DTD und die ETD-ML.DTD sowie die DiML.DTD die Dublin Core Metadaten zum Teil direkt als solche kennzeichnen oder - mit anderen Bezeichnungen versehen - einschließen, einigten sich die Teilnehmer darauf, dass man zusätzlich dazu im Sinne der Interoperabilität einen weltweit abgestimmten externen Metadatensatz benutzen sollte.

In der Kürze der Zeit war trotz der sehr guten Vorbereitung des Diskussionsleiters *Nuno Freire* von der portugiesischen Nationalbibliothek kaum eine Einigung zu Detailfragen zu erzielen. Als nächste Aufgabe wurde von den Teilnehmern erkannt, dass zunächst im eigenen Land eine Diskussion zur Interpretation und Anwendung des Dublin Core-Datensatzes für elektronische Publikationen in Universitätsbibliotheken geführt werden muss, bevor man zu einer globaleren Absprache kommen kann. Man verpflichtete sich, die von Nuno Freire gewartete NDLTD-Mailingliste zu Metadaten zu nutzen, um so innerhalb der nächsten Monate zu einem gemeinsamen Vorschlag zu gelangen.

Ergebnisse und Ausblick

Im Verlaufe des Workshops wurden weitere Fragen andiskutiert. So zum Beispiel das Problem der Zitierung und der Bewahrung der Seitenzahlen zu diesem Zweck. Keiner der Workshopteilnehmer hatte bisher Erfahrungen in der Anwendung von XPath, XLink oder ähnlicher Konzepte vorzuweisen. Dies ist also ein Gebiet, das von allen in Zukunft noch untersucht und auf seine Anwendbarkeit im Bereich von Universitätsbibliotheken noch validiert werden muss.

Eine weitere sehr interessante Diskussion entzündete sich an der Frage, wer wann und vor allem was mit XML auszeichnet. Dabei wurde festgestellt, dass in der momentanen Situation 2 Strategien vonnöten sind: eine Langzeitstrategie und eine mittelfristige Lösung.

Die Langzeitstrategie sollte vorsehen, dass die Autoren direkt in XML-Auto-rensystemen ihre Dokumente unter Nutzung der genormten DTDs oder XML-Schemata schreiben. Die mittelfristige Lösung ist natürlich die der Konvertierung aus gängigen Textsystemen nach XML. Allen Beteiligten erschien es dabei wichtig, dass den Autoren entsprechende Dokumentvorlagen angeboten werden. Zusätzlich würde es die korrekte Anwendung dieser Vorlagen erleichtern, wenn entsprechende validierende Komponenten, z. B. in Word, zur

Verfügung stehen würden, die nur DTD-konforme Auszeichnungen zulassen. Hier sollen verstärkt Methoden gesucht werden.

Als wichtiges Ergebnis des Workshops ist auch die Entscheidung zu werten, sich nicht direkt mit den Details einzelner DTDs zu befassen, sondern sich eher über globalere Dokumentstrukturen abseits von DTDs zu verständigen, so etwa über die Nutzung bestimmter, fachlich ausgerichteter Teil-DTDs oder Schemata.

Zur Frage, wie man die Autoren am besten für das elektronische Publizieren gewinnen kann, ohne Zwang auszuüben und ohne auf Argumente wie das der geringeren Kosten zu verweisen, wurde von allen Teilnehmern einhellig festgestellt, dass gute Beispiel-Dokumente der beste Weg dafür sind. Ziel sollte es in der nächsten Zeit sein, einige überzeugende Beispiele zu erarbeiten, die die Leistungsfähigkeit von XML demonstrieren. Dabei ist es wichtig, auch Autoren aus der LaTeX-Gemeinde davon zu überzeugen, dass die Qualität von XML für das elektronische Publizieren der Qualität des Textsatzsystems TeX bei der Erzeugung von Printausgaben mindestens ebenbürtig ist.

