

## **3 Erschließung**

### **3.1 Überblick**

Der wichtigste Teil jeder Bibliothek und ebenso jeder elektronischen Bibliothek ist die Erschließung der Dokumente. Ohne eine gute Erschließung werden auch wissenschaftlich sehr nützliche und interessante Publikationen nutzlos "verstauben". Deswegen sollte man sich von Anfang an Gedanken um eine effektive Erschließung der elektronischen Dissertationen machen.

In den folgenden Kapiteln wird zunächst die Funktion von Metadaten und von RDF erläutert, die im Bereich der Erschließung elektronischer Dokumente u.a. die Funktion der bibliographischen Beschreibung übernehmen können und für verschiedene Wege der Erschließung wichtig sind. Danach wird die Erschließungsfunktion des lokalen Katalogs erläutert, der jedoch im Rahmen der Globalisierung der Wissenschaften und zunehmender internationaler Zusammenarbeit von Wissenschaftlern weltweit immer mehr in den Hintergrund tritt. Den wichtigeren Erschließungsmöglichkeiten durch Suchmaschinen, Projekten zur Erschließung von Ressourcen eines bestimmten Fachgebietes und anderen nationalen und internationalen Erschließungsmöglichkeiten sind weitere Kapitel gewidmet.

## 3.2 Metadaten

Der Begriff Metadaten besagt generell, daß es sich um Daten über Daten handelt.<sup>45</sup> Die wohl bekannteste Form von Metadaten ist der Kartenkatalog einer Bibliothek. Die Informationen auf einer Katalogkarte sind Metadaten über ein Buch.<sup>46</sup> Eigentlich entsprechen Metadaten somit einer bibliographischen Beschreibung, da aber Metadaten nicht nur Bücher, sondern jedes vorstellbare Objekt beschreiben können sollen, ist der Begriff "bibliographische" Beschreibung hier nicht mehr passend. Außerdem müssen bei der Beschreibung von beispielsweise elektronischen Dokumenten teilweise neue und andere Daten erhoben werden, als bei der Beschreibung von Büchern oder Musikalien.

Metadaten werden also gebildet, um eine elektronische Erschließung von elektronischen Dokumenten und anderen Objekten zu ermöglichen. Sie können entweder Teil des zu beschreibenden Mediums sein oder aber getrennt vorliegen.

### 3.2.1 Dublin Core Metadaten

Ursprünglicher Sinn der Definition der Dublin Core Metadatenpezifikation war es, einen Minimalsatz von Erschließungselementen zu definieren, die zu verbesserter Präzision und Retrievalfähigkeit von elektronischen Dokumenten führen sollte.<sup>47</sup> Denn wie jedes bibliothekarische Regelwerk setzt auch der effektive Einsatz von Metadaten einen gewissen Standardisierungsgrad voraus.<sup>48</sup> Zur Entwicklung solcher Standards hat es in den vergangenen Jahren verschiedene Ansätze geben. Überwiegend durchgesetzt hat sich das von der Dublin Core Metadata Initiative auf verschiedenen Workshops entwickelte und inzwischen 15 Elemente umfassende *Dublin Core Metadata Element Set*<sup>49,50</sup>.

---

<sup>45</sup> vgl. Henze, Volker ; Michael Schefczik: Metadaten : Beziehungen zwischen Dublin Core Set, Warwick Framework und Datenformaten. // In: Bibliotheksdienst. - 31 (1997), S. 413

<sup>46</sup> vgl. Miller, Paul: Metadata for the masses. - URL: <http://www.ariadne.ac.uk/issue5/metadata-massees/> [Stand 9.7.1999]

<sup>47</sup> vgl. Rusch-Feja, Diann: Dublin Core Metadata : auf dem Weg zur Entwicklung eines Internet-Standards - 4. Dublin Core Metadata Workshop in Canberra. // In: Bibliotheksdienst. - 31 (1997), S. 622

<sup>48</sup> Schimmer, Ralf: Einführung in Metadaten. - URL: <http://www2.sub.uni-goettingen.de/metall.html> [Stand 9.7.1999]

<sup>49</sup> vgl. Dublin Core Metadata for Resource Discovery / Stuart L. Weibel ... - URL: <http://www.faqs.org/rfcs/rfc2413.html> [Stand 8.7.1999]

<sup>50</sup> vgl. Dublin Core Metadata Element Set: Reference Description / Dublin Core Metadata Initiative. - URL: [http://purl.org/dc/about/element\\_set.htm](http://purl.org/dc/about/element_set.htm) [Stand 8.7.1999]

Die Dublin Core Metadata Initiative beschreibt die Zielsetzung des Dublin Core Element Reference Sets folgendermaßen:

"The Dublin Core is a metadata element set intended to facilitate discovery of electronic resources. Originally conceived for author-generated description of Web resources, it has attracted the attention of formal resource description communities such as museums, libraries, government agencies, and commercial organizations."<sup>51</sup>

**Die Spezifikation des Dublin Core Element Reference Sets umfaßt folgende Metadaten:<sup>52</sup>**

1. *Title*

Den durch den Autor oder Herausgeber an die Ressource vergebenen Titel

2. *Creator*

Die primär für den intellektuellen Inhalt der Ressource verantwortliche Person oder Organisation (z.B. der Autor bei Textdokumenten, der Komponist / Interpret bei Tonaufnahmen oder der Künstler oder Fotograf bei Bildern / Fotografien)

3. *Subject*

Thema der Ressource, beschrieben durch Stich- bzw. Schlagwörter. Die Benutzung von genormtem Wortschatz (Klassifikationen / Thesauri) wird empfohlen

4. *Description*

Eine textliche Beschreibung des Inhalts der Ressource (z.B. Abstract)

5. *Publisher*

Die für die Veröffentlichung und Verfügbarmachung der Ressource in der vorliegenden Form verantwortliche Organisation (z.B. Verlag, Universität, Firma)

6. *Contributor*

Im Feld Creator nicht genannte Personen oder Körperschaften, die einen wesentlichen aber sekundären Beitrag zur vorliegenden Ressource geleistet haben (z.B. Herausgeber, Übersetzer, Illustratoren)

7. *Date*

Das Datum der Verfügbarmachung der Ressource in der vorliegenden Form. Empfohlenes Format: YYYY-MM-DD (wie beschrieben in <http://www.w3.org/TR/NOTE-datetime>, einem ISO 8601-Profil)

---

<sup>51</sup> The Dublin Core: A Simple Content Description Model for Electronic Resources. - URL: <http://purl.org/dc/> [Stand 9.7.1999]

<sup>52</sup> Stand 1. Juni 1999, vgl. RFC 2413 bzw. "Dublin Core Metadata Element Set: Reference Description" ([http://purl.org/DC/about/element\\_set.htm](http://purl.org/DC/about/element_set.htm))

#### 8. *Type*

Die Kategorie der Ressource, wie z.B. Homepage, Gedicht, Novelle, Technischer Report etc. Die Auswahl sollte aus einer normierten Liste erfolgen, die sich zur Zeit in der Entwicklung befindet. Näheres unter <http://sunsite.berkeley.edu/Metadata/types.html>

#### 9. *Format*

Das Datenformat der Ressource. Es sollte Aufschluß geben über zum Anschauen oder zur Benutzung der Ressource notwendiger Hard- und Software. Die Auswahl sollte aus einer normierten Liste erfolgen, die sich zur Zeit in der Entwicklung befindet

#### 10. *Identifizier*

Zeichen- oder Zahlenfolge zur eindeutigen Identifikation der Ressource, z.B. URL, URN oder aber auch andere weltweit eindeutigen Bezeichner wie ISBN oder ISSN

#### 11. *Source*

Zeichen- oder Zahlenfolge zur eindeutigen Identifikation derjenigen Ressource, aus der die vorliegende Ressource erstellt bzw. abgeleitet wurde. Handelt es sich z.B. bei der vorliegenden Ressource um die PDF-Version eines Buches, könnte im Feld Source die ISBN-Nummer des Buches angegeben werden

#### 12. *Language*

Die Sprache in der die Ressource verfaßt wurde. Die Angabeform sollte möglichst den RFC 1766 folgen

#### 13. *Relation*

Beziehungen der vorliegenden Ressource zu anderen Ressourcen. Hier könnten z.B. die Beziehung zwischen jeweils als eigenes Dokument vorliegenden Kapiteln eines Buches dargestellt werden. Die Auswahl sollte aus einer normierten Liste erfolgen, die sich zur Zeit in der Entwicklung befindet

#### 14. *Coverage*

Räumliche und zeitliche Charakterisierung des intellektuellen Inhalts der Ressource. Eine räumliche Charakterisierung könnte z.B. durch Längen- und Breitengrade gegeben werden. Eine zeitliche Charakterisierung wäre z.B. durch Angabe des Zeitalters oder auch eines Datums möglich

#### 15. *Rights*

Copyright-Vermerk oder Link zu einem Copyright-Vermerk. Außerdem könnten hier Angaben gemacht werden zu Nutzungsbedingungen für die vorliegende Ressource

Neben diesen grundlegenden Elementen wurde festgelegt, daß es möglich sein soll eine Elementverfeinerung durch Subelemente vorzunehmen. Außerdem ist die Benutzung von sogenannten Qualifiern definiert worden, die zur Präzisierung von Inhalten der Metadaten-elemente beitragen sollen.

So könnte z.B. ein Subelement von DC.Title mit dem Namen DC.Title.Translated definiert werden und das Metadatum DC.Title.Translated mit dem Qualifier LANG="eng" präzisiert werden. Auf Benutzung von Subelementen und Qualifiern beruht auch die Definition des Dissertationsmetadatenatzes des DFG-Projekts "Dissertationen Online".

### 3.2.2 Der Dissertationsmetadatenatz von "Dissertationen Online"

Seit Jahresbeginn 1998 hat sich das Teilprojekt Metadaten des DFG-Projekts "Dissertationen Online" in Zusammenarbeit mit den Fachgesellschaften und der Deutschen Bibliothek mit der Entwicklung von für elektronischen Dissertationen geeigneten Metadaten auf der Grundlage des Dublin Core Metadata Element Sets beschäftigt. Das Ergebnis ist ein Satz von Metadaten, der einerseits wichtige Daten zur sinnvollen fachlichen Erschließung von Dissertationen enthält, andererseits sowohl für die Aufnahme von Dissertation an der Deutschen Bibliothek benötigte als auch für den Geschäftsgang der Universitäten relevante Angaben berücksichtigt.<sup>53</sup>

Um die Übernahme dieses Dissertationsmetadatenatzes von möglichst vielen deutschen Universitätsbibliotheken zu fördern bzw. zu unterstützen, wurde gleichzeitig ein bereits vorhandenes Softwaretool, der *My Meta Maker for Theses* (MMMfT), weiterentwickelt und der Öffentlichkeit zur Verfügung gestellt.<sup>54</sup> Die Version 1.2 des MMMfT unterstützt den seit 1.1.1999 in seiner endgültigen Form vorliegenden Metadatenatz des Teilprojektes Metadaten. Mit Hilfe des MMMfTs kann der Doktorand die zu seiner Dissertation gehörigen Metadaten eingeben und lokal auf seinem Rechner abspeichern. Wie die Übergabe der Metadaten an die jeweilige Universitätsbibliothek erfolgt wurde offen gelassen. Das zum MMMfT gehörige Administrationswerkzeuge *Edit* erlaubt nach Abgabe der vom MMMfT erzeugten Datei in der Universitätsbibliothek das Editieren und das Hinzufügen von administrativen Metadaten. Schließlich kann die mit allen Metadaten versehene Datei in ein Distributionsverzeichnis gestellt werden.

---

<sup>53</sup> vgl. Metadaten im Umfeld von Dissertationen. - URL: <http://www.ub.uni-duisburg.de/dissonline/comment2metatags.html> [Stand 2.6.1999]

<sup>54</sup> vgl. Projektziele des Teilprojekts Metadaten des DFG-Projekts "Dissertationen Online". - URL: <http://www.ub.uni-duisburg.de/dissonline/ziele.html> [Stand 8.7.1999]

Die Spezifikation des Dissertationsmetadatenatzes des Teilprojekts Metadaten des DFG-Projekts "Online-Dissertationen" sieht folgende Metadaten vor (Stand 01.01.1999).<sup>55</sup>

<b>a) Dissertationsspezifische Daten</b>		
<b>Bibliographisch</b>		
<i>Kategorie zug. Element des DC</i>	<i>Dublin Core Subklassifikationen</i>	<i>Beschreibung des Feldinhaltes</i>
Titel DC.Title	DC.Title DC.Title.Alternative DC.Title.Translated DC.Title.Alternative.Translated	Titel der Dissertation Untertitel der Dissertation Titel der Dissertation in alternativer Sprache Untertitel der Dissertation in alternativer Sprache
Autor DC.Creator	DC.Creator.PersonalName DC.Creator.PersonalName.Address DC.Creator.PersonalName.DateOfBirth DC.Creator.PersonalName.PlaceOfBirth	Name, Vorname des/der Promovierenden Adresse des/der Promovierenden Geburtsdatum des/der Promovierenden Geburtsort des/der Promovierenden
Universitätsbibliothek DC.Publisher	DC.Publisher.CorporateName DC.Publisher.CorporateName.Address	Name der Universitätsbibliothek Vollständige Postanschrift der UB
Identifizierungskennung DC.Identifier	DC.Identifier	Eindeutige Identifizierungskennung, z.B. URL, URN
<b>Fachspezifisch</b>		
<i>Kategorie zug. Element des DC</i>	<i>Dublin Core Subklassifikationen</i>	<i>Beschreibung des Feldinhaltes</i>
Thema DC.Subject	DC.Subject DC.Subject DC.Subject	Themenspezifische Metadaten Klassifizierungen gemäß der Fachthesauri (vgl. Verknüpfungen / Schemes und Scheme Qualifier, z.B. MSC) Stichworte
Beschreibung DC.Description	DC.Description DC.Description.Note	Inhaltliche Beschreibung der Dissertation, Zusammenfassung oder Link auf eine solche Annahmehinweise des Fachbereichs
Relationen DC.Relation	DC.Relation	Mit der Dissertation verbundene Objekte, z.B. zugrundeliegende Veröffentlichungen
<b>Administrativ</b>		
<i>Kategorie zug. Element des DC</i>	<i>Dublin Core Subklassifikationen</i>	<i>Beschreibung des Feldinhaltes</i>
Herausgeber DC.Contributor	DC.Contributor.Referee DC.Contributor.Advisor DC.Contributor.CorporateName DC.Contributor.CorporateName.Address	Gutachter Betreuer / Doktorvater Universität, Fachbereich
Daten DC.Date	DC.Date.Submitted DC.Date.Created DC.Date.Accepted	Datum der Antragstellung zur Promotion Datum der rechtswirksamen Abgabe der Pflichtexemplare bzw. des Pflichtfiles Datum der mündlichen Prüfung / Promotionsdatum

<sup>55</sup> Metadaten im Umfeld von Dissertationen : Dissertationsmetadatenatz. - URL: [http://www.ub.uni-  
duisburg.de/dissonline/metatags.html](http://www.ub.uni-duisburg.de/dissonline/metatags.html) [Stand 2.6.1999]

Urheberrechte DC.Rights	DC.Rights DC.Rights	Copyright Hinweise des Promovenden / Einverständniserklärungen Link auf die Ausführungsbestimmungen des Fachbereiches
<b>b) Formaldaten</b>		
<i>Kategorie zug. Element des DC</i>	<i>Dublin Core Subklassifikationen</i>	<i>Beschreibung des Feldinhaltes</i>
Sprache DC.Language	DC.Language	Abkürzungen der Sprache in der die Dissertation verfaßt ist nach ISO 639-2 (z.B. ger für deutsch)
Format DC.Format	DC.Format	MIME Bezeichner der/des Präsentationsformate/s
Typ DC.Type	DC.Type	Dokumentenart DC.Type="Text.PhDThesis"
Quelle DC.Source	DC.Source	Bezeichner, welcher die Quelle der Onlineversion eindeutig identifiziert, z.B. Tex-Datei oder Word- Dokument
<b>Links</b>		
<i>Kategorie zug. Element des DC</i>	<i>Dublin Core Subklassifikationen</i>	<i>Beschreibung des Feldinhaltes</i>
Schema SCHEME	SCHEME.DC SCHEME.PACS1998 SCHEME.CR1998	Link zum jeweiligen Klassifikationsschema

### 3.3 RDF

"The Web is quickly becoming the world's fastest growing repository of data. [...] RDF provides the necessary foundation and infrastructure to support the description and management of this data. RDF can transform the Web into a more useful and powerful information resource."<sup>56</sup>

**Tim Berners-Lee**, Direktor des World Wide Web Consortiums (W3C)

Die große Vielfalt von Metadaten im World Wide Web macht eine Infrastruktur notwendig, die die Koexistenz unterschiedlicher, unabhängig voneinander entwickelter sowie neu zu entwickelnder Metadaten-Standards (z.B. des Dublin Core Metadata Element Sets und des Australian Government Locator Service (AGLS) Metadata Standards<sup>57</sup>) unterstützt.

Denn die Metadaten des Dublin Core Sets sind tatsächlich nur ein kleiner Teil der Metadaten, die auf dem World Wide Web vorstellbar wären. Ebenso wie Autor, Titel und Verlag eines Buches wären z.B. Maler, Titel und besitzendes Museum eines Bildes vorstellbar. Oder Produktname, Firma und verkaufender Großhändler eines Scanners. Diese Liste ließe sich, wie man bald merkt, endlos fortsetzen und eine Entwicklung von Standards auch für solche Metadaten ist absehbar bzw. schon vollzogen. Es wird jedoch nicht möglich sein, die Entwicklung solcher Standards zentral zu steuern oder zu dokumentieren.

Das Resource Description Framework (RDF) - eine XML<sup>58</sup>-Applikation entwickelt vom World Wide Web Consortium (W3C) 1999<sup>59</sup> - stellt die Grundlage bereit für die Interoperabilität zwischen verschiedenen Metadaten-Standards. Es versucht die Gemeinsamkeiten von existierenden und vorstellbaren Metadatenstandards zu identifizieren und eine Möglichkeit für WWW-Autoren bereitzustellen, verschiedene Metadatenstandards zu integrieren und den Austausch von Metadaten zu vereinfachen.

RDF kann in einer Vielzahl von Anwendungsgebieten benutzt werden, z.B. im *Information Retrieval* durch Verbesserung von Suchmaschinen, in der *Katalogisierung* und im Aufbau von *digitalen Bibliotheken* zur besseren Beschreibung von Webseiten, durch *intelligente*

---

<sup>56</sup> W3C Issues Recommendation for Resource Description Framework (RDF). - URL: <http://www.w3.org/Press/1999/RDF-REC> [Stand 12.7.1999]

<sup>57</sup> Australian Government Locator Service (AGLS). - URL: <http://naa.gov.au/GOVSERV/AGLS/default.htm> [Stand 12.7.1999]

<sup>58</sup> Extensible Markup Language (XML) 1.0. - URL: <http://www.w3.org/TR/REC-xml> [Stand 14.7.1999]

*Softwareagenten* zur Vereinfachung der gemeinsamen Nutzung und des Austauschs von Wissen, in *der Bewertung des Inhalts* von Webseiten oder zur Beschreibung *geistigen Eigentums*.<sup>60</sup>

### 3.3.1 Das grundlegende RDF-Modell

Die Grundlage von RDF ist ein Modell zur Repräsentation benannter Eigenschaften und Werte dieser Eigenschaften. Damit übernimmt RDF die bekannte und bewährte Philosophie von Attribut-Werte-Paaren.

Das RDF-Modell besteht aus 3 Objekt-Typen:

**Ressourcen:** Alles, was von RDF beschrieben wird, nennt man Ressourcen (engl. "Resources"). Dabei können Ressourcen z.B. eine Webpage, Teile einer Webpage, eine Sammlung von Webpages, aber auch Objekte sein, auf die gar nicht über das Web zugegriffen werden kann, z.B. Bücher, CDs, Geräte und allgemein jede vorstellbare Einheit, der man einen URI (Uniform Resource Identifier) zuordnen kann.

**Eigenschaften:** Eine Eigenschaft (engl. "Property") ist, wie der Name schon sagt, eine spezifische Charakteristik oder ein spezifisches Attribut, das der Beschreibung einer Ressource dient. Jede Eigenschaft definiert erlaubte Werte und Beziehungen mit anderen Eigenschaften.

**Aussagen:** Eine Aussage (engl. "Statement") ist eine spezielle Ressource zusammen mit einer Eigenschaft dieser Ressource und dem Wert dieser Eigenschaft. Diese 3 Teile einer Aussage nennt man auch Subjekt, Prädikat und Objekt. Das Objekt (also der Wert einer Eigenschaft) kann entweder eine weitere Ressource oder ein Literal sein.<sup>61</sup>

Bsp.:

Folgende Tatsache soll beschrieben werden:

---

<sup>59</sup> Resource Description Framework (RDF) Model and Syntax Specification. - URL: <http://www.w3.org/TR/REC-rdf-syntax/> [Stand 12.7.1999]

<sup>60</sup> vgl. Resource Description Framework (RDF) Model and Syntax Specification ...

<sup>61</sup> vgl. Lamprecht, Stephan: RDF. // In: Internet Professional. - 3 (1999), 1, S. 94

*Der Autor von "Bereitstellung und Erschließung von elektronischen Dissertationen" ist Tobias Fritz*

Hier wäre

das Subjekt (die Ressource)	<i>"Bereitstellung und Erschließung von elektronischen Dissertationen"</i>
das Prädikat (die Eigenschaft)	Autor
das Objekt (Literal)	Tobias Fritz

Ein Beispiel aus dem WWW wäre z.B. der Autor einer Webpage:

*Der Autor der Seite <http://www.vetmed.fu-berlin.de/~fritz/index.html> ist Tobias Fritz*

Hier wäre

das Subjekt (die Ressource)	<i><a href="http://www.vetmed.fu-berlin.de/~fritz/index.html">http://www.vetmed.fu-berlin.de/~fritz/index.html</a></i>
das Prädikat (die Eigenschaft)	Autor
das Objekt (Literal)	Tobias Fritz

### 3.3.2 Die grundlegende RDF Syntax

Die Repräsentation obiger Aussage in RDF bzw. als XML-Dokument und unter Benutzung des RDF-Namespaces und des DC-Namespaces sieht folgendermaßen aus:

```
<?xml version="1.0"?>
<RDF
  xmlns = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:DC = "http://purl.org/DC#">
  <Description about =
    "http://www.vetmed.fu-berlin.de/~fritz/index.html">
    <DC:Creator>Tobias Fritz</DC:Creator>
  </Description>
</RDF>
```

Die Angabe der beiden Namespaces bedeutet, daß alle Eigenschaften der Beschreibung aus einer der beiden Namespaces stammen und der RDF-Namespace den Default-Wert darstellt.<sup>62</sup>

An dieser Stelle erkennt man den am Anfang des Kapitels beschriebenen großen Vorteil von RDF: die Integration verschiedener Vokabulare bzw. Metadaten-Standards.

---

<sup>62</sup> vgl. Ianella, Renato: An idiot's guide to the Resource Description Framework. - URL: <http://www.dstc.edu.au/RDU/reports/RDF-Idiot/> [Stand 15.7.1999]

Da eine weitere Erläuterung des RDF-Modells und der RDF-Syntax über den Rahmen dieser Arbeit hinausgehen würde, verweise ich für eine genauere Beschreibung auf die "Resource Description Framework (RDF) Model and Syntax Specification"<sup>63</sup> des W3C.

### **3.3.3 RDF und elektronische Dissertationen**

Der Einsatz von RDF für die Beschreibung elektronischer Dissertationen ist wiederum hauptsächlich wegen der Integrationsmöglichkeit verschiedener Metadatenstandards empfehlenswert. Insbesondere zur Beschreibung multimedialer Inhalte von Dissertationen werden neben Dublin Core weitere Metadaten-Standards herangezogen werden müssen. Daneben wird der Austausch von Metadaten sowie die Suche nach Metadaten durch RDF wesentlich vereinfacht und verbessert, so daß auch aus diesem Grund RDF für elektronische Dissertationen benutzt werden sollte.

Ein erster Modellentwurf eines dissertationsspezifischen DC-Metadatenatzes in RDF wurde im Frühjahr 1999 vom Teilprojekt Metadaten des DFG-Projekts "Dissertationen Online" entwickelt (siehe Anhang 2: Modell eines RDF-Datensatzes für Dissertationen).

---

<sup>63</sup> Resource Description Framework (RDF) Model and Syntax Specification ...

### 3.4 Lokale Kataloge

Neben einer Erschließung über das Internet und in speziellen Katalogen bzw. Referateorganen bleibt natürlich auch der Nachweis elektronischer Dissertationen im lokalen Katalog wichtig. Trotz oder gerade wegen ihres informationswissenschaftlichen Mehrwerts sollten elektronisch Dokumente nicht grundsätzlich an anderer Stelle als gedruckte Publikationen erschlossen werden. Andernfalls würde wieder eine Situation mit unübersichtlichen *Partikularnachweisen* entstehen, die gerade im Zeitalter der Online-Bibliothekskataloge als historisch zu gelten hat.<sup>64</sup>

Um unnötigen Arbeitsaufwand zu vermeiden, sollten die schon vorhandenen Metadaten zur automatischen Erstellung der RAK-Titelaufnahme benutzt werden. Zu diesem Zweck wurde z.B. an der Universität Stuttgart in Zusammenarbeit mit dem Bibliotheksservicezentrum Baden-Württemberg (BSZ) eine Dublin-Core-Konvention festgelegt und ein Konverter entwickelt, der die Metadaten nach dieser Konvention auf das Format des Südwestdeutschen Bibliotheksverbundes (SWB) abbildet und in dessen Verbunddatenbank einspielt.<sup>65</sup> Die Entwicklung eines Dublin-Core-Metadatensatzes für Online-Dissertationen sollte auch in dieser Hinsicht bearbeitet werden und sämtliche für RAK notwendigen Daten enthalten, um eine Nachbearbeitung der automatisch generierten Datensätze zu vermeiden. Außerdem sollte die Titelaufnahme neben den Elementen nach RAK in der Fußnote die URL der elektronischen Dissertationen enthalten. Damit werden auch Benutzer des lokalen Kataloges auf die elektronische Version dieser Arbeit hingewiesen. Weiterhin sollte der lokale Katalog möglichst über das WWW abfragbar sein, und die URL einen direkten Link zur Web-Seite der Dissertation darstellen.

---

<sup>64</sup> vgl. Stephan, Werner ; Frank Scholze: Online Publikationsverbund : Erfassung und Organisation elektronischer Hochschulschriften. // In: Bibliotheksdienst. - 33 (1999), S. 97

<sup>65</sup> vgl. Stephan, Werner ; Frank Scholze: Online Publikationsverbund ... S. 97

### **3.5 Suchmaschinen**

Natürlich sollten die elektronischen Dissertationen oder zumindest die Titelseiten der elektronischen Dissertationen auch von ganz gewöhnlichen Suchmaschinen wie Altavista, Fireball, Hotbot u.a. gefunden werden, da diese sehr häufig benutzt werden und oft die schnellste Möglichkeit darstellen, um Informationen zu einem bestimmten Thema im WWW zu finden.

Zum Anmelden neuer Seiten bzw. neuer Dissertationen bei den verschiedenen Suchmaschinen können entweder die Schnittstellen der einzelnen Suchmaschinen oder kommerzielle oder frei im Web verfügbare Dienste benutzt werden, die die gleichzeitige Anmeldung einer neuen Seite bei vielen verschiedenen Suchmaschinen erlauben. Wird keine Anmeldung von neuen Dissertationen vorgenommen, bleibt es mehr oder weniger dem Zufall überlassen, ob diese innerhalb des nächsten Monats oder innerhalb der nächsten Jahre von einer oder mehreren Suchmaschinen indexiert werden.

Es ist deshalb zu empfehlen, zu Beginn eines Projektes zur Sammlung elektronischer Dissertationen eine Auswahl von Suchmaschinen zu treffen und sich näher mit den Möglichkeiten der Eintragung neuer Seiten in die Datenbanken dieser Suchmaschinen zu beschäftigen. Bei der Auswahl der Suchmaschinen sollte auch darauf geachtet werden, ob diese die Suche nach (Dublin Core) Metadaten erlauben, da nur so die in die Titelseiten integrierten Metadaten bei der Suche nützlich sind.

Danach könnte eine automatische Anmeldung bei den ausgewählten Suchmaschinen in den Workflow der Einarbeitung der elektronischen Dissertationen integriert werden. Als Alternative zu der für diesen Mechanismus notwendigen Programmierarbeit stehen kommerzielle oder frei verfügbare Dienstleister zur Verfügung, die dem Nutzer je nach Umfang des Services und des Preises die Anmeldung bei mehreren hundert Suchmaschinen teilweise oder komplett abnehmen.

Wenn auch dem Fachwissenschaftler andere Wege zur Verfügung stehen, um aktuelle Informationen zu einem bestimmten Thema zu finden, so können Suchmaschinen trotzdem oft der schnellste und einfachste Zugang zu neuen elektronischen Dissertationen sein. Aus diesem Grund und um möglichst vielen Personen möglichst schnell den Zugriff auf neue Dissertationen zu gewähren, ist eine Auseinandersetzung mit dem Thema "Suchmaschinen" unter dem Gesichtspunkt der Anmeldung neuer Web-Inhalte unumgänglich.

### 3.6 Zusammenarbeit mit anderen Institutionen

In Zeiten, die einerseits durch Informationsüberfluß geprägt sind, in denen Wissenschaftler andererseits hunderte verschiedene Suchmaschinen, Datenbanken, Referateorganen u.ä für eine Recherche hinzuziehen können, ist eine Zusammenarbeit von Institutionen gleicher Fachrichtung von enormer Wichtigkeit. Schon der Benutzer einer normalen Bibliothek steht vor dem Problem, seine Suche nach Literatur nicht auf ein System beschränken zu können, sondern neben dem lokalen EDV-Katalog zunächst den alten Zettelkatalog, dann verschiedene CD-ROM-Datenbanken (wenn er Glück hat mit derselben Oberfläche) und verschiedene Ausgaben und Reihen der Current Contents durchsuchen zu müssen. Möchte der Benutzer zusätzlich das Internet in seine Suche mit einbeziehen, steht er vor zahlreichen Suchmaschinen, den Homepages der Fachbereiche und den Internetangeboten der seinem Fachgebiet zugeordneten nationalen und internationalen Institutionen. Bei diesem Informationsüberfluß die richtigen und wichtigsten Informationsangebote zu finden ist beinahe unmöglich.

Deswegen ist eine Zusammenarbeit von fachlich zusammengehörenden Institutionen früher oder später zwingend. Als Beispiel möchte ich hier auf das PhysNet<sup>66</sup> eingehen, das mit PhysDis<sup>67</sup>, PhysDoc<sup>68, 69</sup> und MetaPhys<sup>70</sup> drei sowohl nationale als auch internationale Quellen einschließende Suchmöglichkeiten für Physiker geschaffen hat.

#### 3.6.1 PhysDoc / PhysDis

PhysDoc ermöglicht eine Suche verteilter Dokumente auf Servern von Physik-Fachbereichen in Europa bzw. den USA. Welche Dokumente damit erschlossen werden, liegt an den einzelnen Fachbereichen. Die Einbindung der einzelnen Fachbereiche erfolgt manuell nach Anmeldung bei PhysDoc. Die ideale Anbindung eines Fachbereichs an PhysDoc erfolgt durch Installation und Betrieb eines Harvest Gatherers an dem jeweiligen Fachbereich.

---

<sup>66</sup> PhysNet : Physics Departments Worldwide. - URL: <http://www.physik.uni-oldenburg.de/EPS/PhysNet/> [Stand 7.5.1999]

<sup>67</sup> PhysDis : Physics Theses in Europe. - URL: [http://www.physik.uni-oldenburg.de/EPS/PhysNet/link-lists/noframe\\_dis\\_europe.html](http://www.physik.uni-oldenburg.de/EPS/PhysNet/link-lists/noframe_dis_europe.html) [Stand 19.5.1999]

<sup>68</sup> PhysDoc : Physics Documents in Europe. - [http://www.physik.uni-oldenburg.de/EPS/PhysNet/link-lists/noframe\\_doc\\_europe.html](http://www.physik.uni-oldenburg.de/EPS/PhysNet/link-lists/noframe_doc_europe.html) [Stand 19.5.1999]

<sup>69</sup> PhysDoc : Physics Documents in the USA. - [http://www.physik.uni-oldenburg.de/EPS/PhysNet/link-lists/noframe\\_doc\\_us.html](http://www.physik.uni-oldenburg.de/EPS/PhysNet/link-lists/noframe_doc_us.html) [Stand 19.5.1999]

<sup>70</sup> MetaPhys : The Physics Document Meta Search Engine. - URL: <http://www.physik.uni-oldenburg.de/MetaPhys/> [Stand 7.5.1999]

Dieser Harvest Gatherer sammelt und formatiert an dem jeweiligen Fachbereich vorhandene Informationen und stellt sie dem bei PhysDoc betriebenen Harvest Broker zur Verfügung. Die durch den Benutzer bei PhysDoc eingegebene Suche erfolgt dann über den Harvest Broker. Damit wird ein Teil der Arbeitslast auf die einzelnen Fachbereiche ausgelagert, gleichzeitig können so die Fachbereiche selber eine Auswahl treffen, welche Informationen sie dem Harvest Broker zur Verfügung stellen wollen. Zur Verbesserung der Suchergebnisse empfiehlt PhysDoc die Integration von Dublin-Core-Metadaten und stellt dazu den "MyMetaMaker to Markup Resources in Physics" zur Verfügung.

Die Einrichtung eines solchen verteilten Systems auf der Basis von Harvest Gatherern und einem Harvest Broker halte ich für die im Moment beste Möglichkeit der Erschließung von Internet-Ressourcen eines Fachgebietes. Auch einer Eingrenzung dieses Systems z.B. ausschließlich auf elektronische Dissertationen steht nichts im Wege und wurde z.B. für den Fachbereich Physik mit PhysDis realisiert. Damit könnte die zeitraubende Suche nach elektronischen Dokumenten/Dissertationen eines bestimmten Faches auf eine einzige Suche an einer einzigen Stelle eingeschränkt werden. Nähere Erläuterungen zum Harvest System befinden sich im übernächsten Kapitel.

Möchte man neben elektronischen Dokumenten auch verschiedene Datenbanken gleichzeitig durchsuchen, ist die Einrichtung eines Services wie MetaPhys zu empfehlen.

### 3.6.2 MetaPhys

"**MetaPhys** is a Meta Search Engine which serves as a unified query interface to the databases of a various publishers with relevance in physics."<sup>71</sup>

Dieses Zitat beschreibt treffend die Funktion von MetaPhys. Mit Hilfe eines Java-Programms wird hier nicht nur PhysDoc durchsucht, sondern auch verschiedene kommerzielle Datenbanken als auch das Los-Alamos E-Print Archive sowie der E-Print-Server der American Physical Society. Die Erstellung eines solchen Services ist meiner Meinung nach für jedes Fachgebiet sehr hilfreich und erstrebenswert, allerdings mit umfangreichem Arbeits- und Programmieraufwand verbunden, da es keinen durchgehend genutzten Standard für Online-Datenbanken gibt. Z39.50 ist zwar das populärste und weit verbreitetste Protokoll zur

---

<sup>71</sup> MetaPhys : The Physics Document Meta Search Engine. - URL: <http://www.physik.uni-oldenburg.de/MetaPhys/about.html> [Stand 8.7.1999]

Abfrage von Datenbanken, ist jedoch trotzdem nur bei einem Bruchteil der über das Web abfragbaren Datenbanken verfügbar, da es ursprünglich für hochstrukturierte (z.B. bibliographische Daten) entwickelt wurde, die auf dem WWW leider noch eher selten anzutreffen sind.<sup>72</sup>

### 3.6.3 Das Harvest-System

Harvest ist ein 1994 unter der GNU Public License (GPL) erschienenes und somit kostenloses Indiziersystem für Unix, dessen Entwicklung 1994 von der Internet Research Task Force Group on Resource Discovery (IRTF-RD) begonnen wurde. Da dieses Projekt seit 1996 nicht mehr finanziert wird, wird Harvest heute vom Tardis Project an der University of Edinburgh weiterentwickelt und dokumentiert.<sup>73</sup>

Harvest ist eine dezentrale Suchmaschine, d.h. es bietet eine verteilte Lösung für das Problem der Indexierung im Internet verfügbarer Daten. Durch das System eines *zentralen Brokers* der seine Daten von den *lokal auf jedem Webserver installierten Gatherern* erhält werden viele Probleme herkömmlicher Suchmaschinen und Webcrawler vermieden:

1. Die Belastung des Internet wird verringert, da die Daten lokal gesammelt werden
2. Es können problemlos Broker für bestimmte Fächer betrieben werden; damit werden die üblichen hohen Prozentsätze von irrelevanten Treffern bei konventionellen (interdisziplinären) Suchmaschinen vermieden
3. Die Aktualität der Daten ist im Vergleich zu konventionellen Suchmaschinen exzellent, denn der Zeitabstand der Aktualisierungen kann je nach Wunsch eingestellt werden; damit werden auch erst kürzlich hinzugekommene Daten berücksichtigt und die nach einigen Untersuchungen in konventionellen Suchmaschinen enthaltenen toten Links in einer Höhe von 15%<sup>74</sup> werden schnell aus der Datenbank entfernt

---

<sup>72</sup> vgl. Powell, James ; Edward A. Fox : Multilingual federated searching across heterogeneous collections. // In: D-Lib Magazine. - 4 (1998), 8, URL: <http://www.dlib.org/dlib/september98/powell/09powell.html> [Stand 20.05.1999]

<sup>73</sup> Harvest Indexer. - URL: <http://www.tardis.ed.ac.uk/harvest/> [Stand 19.5.1999]

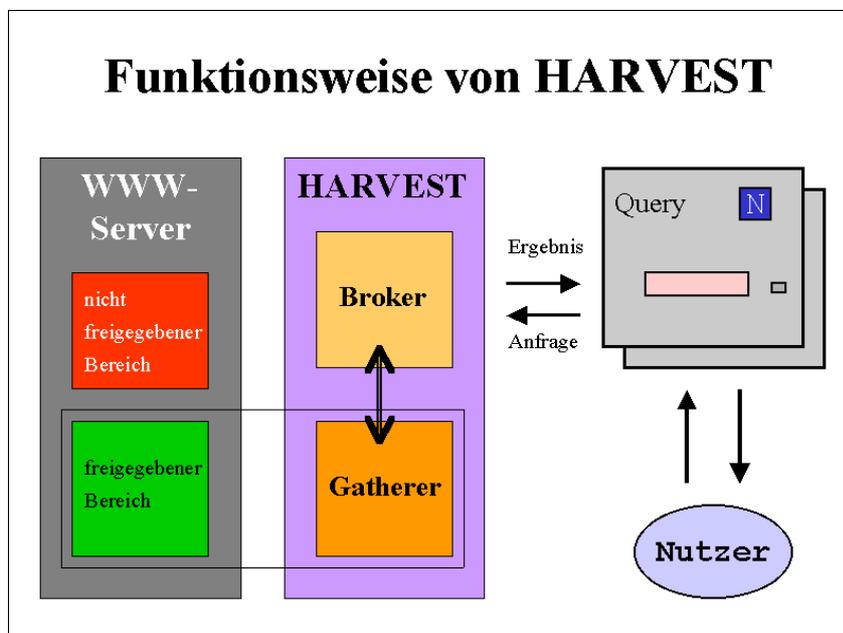
<sup>74</sup> vgl. Umlauf, Konrad: Suchmaschinen im Internet. - Berlin : Institut für Bibliothekswissenschaft der Humboldt-Universität zu Berlin, 1999. - (Berliner Handreichungen zur Bibliothekswissenschaft und Bibliothekerausbildung ; 58) (Materialien zur Fortbildung ; 6). - URL: <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h58/> [Stand 19.5.1999]

### 3.6.3.1 Funktionsweise von Harvest<sup>75</sup>

Das Harvest System besteht aus 2 Teilen: dem sogenannten *Gatherer* und dem sogenannten *Broker*.

Der *Gatherer* sammelt die Daten, arbeitet sie auf und stellt sie für den Abruf durch eine oder mehrere Broker bereit. Der *Broker* ruft die Daten von einem oder mehreren Gatherern ab, archiviert sie und stellt ein Interface in Form einer HTML-Seite für Benutzeranfragen zur Verfügung.<sup>76</sup>

Folgende Grafik stellt die grundlegende Funktionsweise von Harvest dar:<sup>77</sup>



#### 3.6.3.1.1 Der Harvest Gatherer

Die Aufgabe des Gatherers besteht, wie schon erwähnt, im Sammeln, Aufarbeiten und Bereitstellen von Daten. Er besteht aus 4 Komponenten: dem *Enumerator*, dem Programmteil *Essence*, einer *Indexdatenbank* und dem *Gatherer-Daemon*.

<sup>75</sup> vgl. Borggraefe, Stefan: Reiche Ernte : hierarchische Suchindizes mit Harvest. // In: iX. - (1999), 2, S. 126 - 129

<sup>76</sup> vgl. Was ist Harvest? - URL: <http://www.physik.uni-oldenburg.de/~eprint/schul/einl.html> [Stand 20.5.1999]

<sup>77</sup> Funktionsweise von Harvest. - URL: <http://www.physik.uni-oldenburg.de/~eprint/schul/harvfunk.html> [Stand 20.5.1999]

Dem **Enumerator** muß zunächst mitgeteilt werden, wo er seine Suche nach Daten starten soll. Diese erste Seite indiziert er, folgt in dieser Seite gefundenen Links, indiziert wiederum die neu gefundenen Seiten etc. Um nicht zu einem gewöhnlichen Gatherer einer konventionellen Suchmaschine zu werden, müssen ihm bestimmte Limits vorgegeben werden. Ein Limit könnte z.B. die Einschränkung auf einen bestimmten Server sein. Außerdem läßt sich spezifizieren, wie der Enumerator seine Daten sammelt. Hier steht das HTTP-Protokoll, das FTP-Protokoll, das NNTP-Protokoll, Gopher sowie das lokale Dateisystem zur Verfügung. Insbesondere letzteres unterscheidet Harvest von Suchrobotern anderer Suchmaschinen, denn diese können nur die vier vorher genannten Protokolle benutzen. Durch die Nutzung des lokalen Dateisystems läßt sich z.B. auch das Sammeln auf bestimmte Verzeichnisse reduzieren, unabhängig von den in den Dateien vorhandenen Links.

In einem zweiten Schritt übernimmt **Essence** die Dateityperkennung und gleichzeitig den (vorher konfigurierten) Ausschluß bestimmter Dateien von der Indexierung. Denn mit Harvest besteht nicht nur die Möglichkeit normalen ASCII-Text zu indizieren, sondern auch mit Hilfe spezieller *Summarizer* z.B. PDF-, PostScript-, RTF-, WordPerfect- und TeX-Dokumente. Außerdem trägt die Dateityperkennung zur *sinnvollen* Indizierung der Dateien bei, da es zur Unterscheidung zwischen Füllwörtern und relevanten Schlüsselwörtern wichtig ist, daß der Gatherer die Sprache verarbeiten kann, in der die Informationen eingebettet sind.<sup>78</sup> Ohne eine Dateityperkennung für in HTML verfaßte Dateien wäre es z.B. nicht möglich festzustellen, daß es sich bei "<body>" um ein irrelevantes und damit nicht zu indizierendes Wort zur Formatierung handelt, während die Wörter aus dem Satz "The doctor was examining his body" natürlich in den Index aufgenommen werden müssen.

Neben den im Summarizer vorgegebenen Algorithmen zur Entscheidungsfindung über die Relevanz von Begriffen ist es möglich, eigene Konfigurationen vorzunehmen. An dieser Stelle könnte dem Summarizer für HTML z.B. vorgeschrieben werden nur die Daten aus den Metatags der HTML-Dateien zu indizieren und so eine Datenbank über Metadaten aufzubauen.

Zum Abschluß werden die von den Summarizern nicht verworfenen Daten als Attribut-Wert-Paare im Summary Object Interchange Format (SOIF) in der **Indexdatenbank** abgelegt und durch den **Gatherer-Daemon** (*gatherd*) den Brokern zur Verfügung gestellt.

---

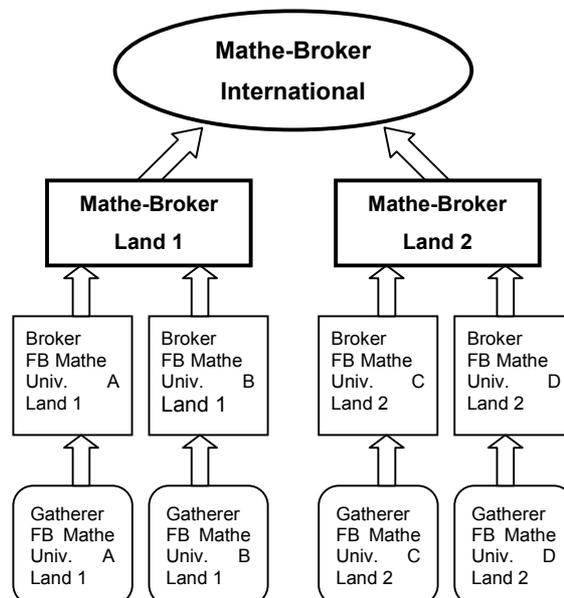
<sup>78</sup> vgl. Borggraefe, Stefan ; Oliver Schade: Weniger weit weg : Programmpakete zur Indizierung der eigenen Website. // In: iX. - (1999), 2, S. 89

### 3.6.3.1.2 Der Harvest Broker

Der grundlegende Aufbau des Brokers ist folgender: er besteht aus einem *Collector*, einer *Indexdatenbank im SOIF-Format*, einer *Glimpse-Indexdatenbank* und einem *Query-Manager*.

In einem ersten Schritt bezieht der Broker durch den **Collector** Daten von einem oder mehreren Gatherern sowie von einem oder mehreren anderen Brokern. Im zweiten Schritt baut der Broker aus diesen Daten wiederum eine **Indexdatenbank im SOIF-Format** auf. Diese Daten stehen dann (wie die SOIF-Daten des Gatherers) wiederum anderen Brokern zur Verfügung. Im dritten Schritt erzeugt bzw. aktualisiert das Indizier- und Suchprogramm Glimpse aus den SOIF-Daten eine **Glimpse-Indexdatenbank**. Und viertens beantwortet ein **Query-Manager** mit Hilfe eines Query-Interfaces in Form einer HTML-Seite Benutzeranfragen an die Datenbank. Bei diesen Benutzeranfragen sind sämtliche Optionen des Programms Glimpse erlaubt, z.B. unscharfe Suche und die Benutzung boolescher Operatoren.

Folgende Grafik stellt noch einmal einen möglichen Aufbau eines fachspezifischen Harvest-Systems dar:



### 3.7 Fächerübergreifende nationale und internationale Erschließung

#### 3.7.1 Nationale Erschließung

Neben dem fachlichen Zugriff auf elektronische Dissertationen sollte es auch eine fächerübergreifende nationale Verzeichnung elektronischer Dissertationen geben. Die zuständige Stelle ist die jeweilige Nationalbibliothek.

"Es ist das erklärte Ziel Der Deutschen Bibliothek, in der Zukunft ihren Benutzern elektronische Publikationen anbieten zu können."<sup>79</sup>

In Deutschland hat *Die Deutsche Bibliothek* den gesetzlichen Auftrag zur vollständigen Sammlung, Erschließung und dauerhaften Sicherung deutschsprachiger und in Deutschland erschienener Publikationen einschließlich physisch (z.B. auf CD-ROM) verbreiteter digitaler Publikationen. Die Ablieferung von Netzpublikationen ist jedoch bisher im Gesetz über Die Deutsche Bibliothek nicht vorgesehen.

Trotzdem werden auf der Basis der freiwilligen Meldung und Abgabe seit dem 1. Juli 1998 auch digitale Netzpublikationen gesammelt. Die "Eröffnung" der *nationalen digitalen Bibliothek* erfolgte am 1. Oktober 1999. In einem ersten Schritt zum Aufbau dieser nationalen digitalen Bibliothek werden **Online-Dissertationen und -Habilitationen** gesammelt und auf einem Dokumentserver der Deutschen Bibliothek archiviert.

Die Meldung des Vorliegens von Online-Dissertationen und -Habilitationen erfolgt entweder per Email oder per HTML-Formular an Die Deutsche Bibliothek, die sie vom Server der meldenden Institution auf den Dokumentserver DEPOSIT.DDB.DE übernimmt. Da zur Strukturierung der zu meldenden Daten Metadaten auf Grundlage des Dublin Core Metadata Element Sets verwendet werden sollen, hat Die Deutsche Bibliothek in Zusammenarbeit mit dem DFG-Projekt "Dissertationen Online" METADISS entwickelt, einen speziellen Metadatensatz für Online-Hochschulschriften.<sup>80</sup>

---

<sup>79</sup> Stephan, Werner: Dissertationen online. // In: Dialog mit Bibliotheken. - 9 (1997), 2, S. 22-25

<sup>80</sup> Die Deutsche Bibliothek: METADISS - Format des Metadatensatzes für Online-Hochschulschriften Der Deutschen Bibliothek und des Projekts "Dissertationen Online". - URL: <http://deposit.ddb.de/metadiss.htm> [Stand 26.5.1999]

Mit dieser Vorgehensweise als Grundlage ist es den beteiligten Universitätsbibliotheken ohne weiteres möglich, eine Meldung an Die Deutsche Bibliothek vollkommen automatisiert ablaufen zu lassen - unter Voraussetzung der Verwendung von METADISS. Auch bei der Deutschen Bibliothek können damit die über die Metadatenchnittstelle empfangenen Daten ausschließlich maschinell weiterverarbeitet werden. Stellt Die Deutsche Bibliothek strukturelle Inkompatibilität zu HTML 4.0 fest oder das Fehlen als obligatorisch definierter Elementbezeichner, wird der jeweilige Metadatensatz in einem automatisierten Verfahren an den Absender zurückgeliefert.

Der Zugriff auf digitale Publikationen geschieht wie für konventionelle Medien über eine Suche im OPAC der Deutschen Bibliothek, der im Fall einer digitalen Publikation einen Link auf den Dokumentserver der Deutschen Bibliothek enthält. Ist das Dokument der Deutschen Bibliothek zwar schon gemeldet, aber noch nicht archiviert worden, wird der Benutzer automatisch auf den jeweiligen Universitätsrechner bzw. Verlagsrechner weitergeleitet.<sup>81</sup>

### 3.7.2 Internationale Erschließung

Als größtes internationales Projekt zur Erschließung elektronischer Dissertationen ist hier die *Networked Digital Library of Theses and Dissertations (NDLTD)*<sup>82</sup> zu nennen.

#### Kurzer geschichtlicher Abriß der NDLTD

Nach einem Treffen von Mitarbeitern der Firmen UMI, ArborText, SoftQuad sowie der University of Michigan und der Virginia Tech im Jahr 1987 und umfangreichen Vorarbeiten, Experimenten und weiteren Treffen erhielt die NDLTD im Frühjahr 1996 eine auf ein Jahr beschränkte \$ 91.117 umfassende Unterstützung durch die Southeastern Universities Research Association, Inc. (SURA) sowie Hardwareausstattung durch die Firma IBM im Wert von \$ 250.000 und nahm damit ihre Arbeit auf, damals noch als "National Digital Library of Theses and Dissertations". Zusätzlich wurde im September 1996 eine dreijährige Finanzierung durch das U.S. Department of Education (FIPSE) bewilligt.

Obwohl von der NDLTD SGML bzw. XML immer als ideale Wahl für die Archivierung von elektronischen Dissertationen angesehen wurde, bestanden und bestehen bis heute ernste

---

<sup>81</sup> vgl. Die Deutsche Bibliothek: Dissertationen online als erster Schritt zur nationalen Bibliothek : Pressemitteilung. - Frankfurt a. M., 1998. - S. 1-2

<sup>82</sup> URL: <http://www.ndltd.org> [Stand 9.7.1999]

technische und ökonomische Probleme, die auch bei der NDLTD dazu geführt haben, daß nur ein kleiner Teil der Dissertationen mit Hilfe einer speziellen DTD archiviert wurde, während der überwiegende Teil der Dissertationen im Format PDF abgegeben und archiviert wird. Dennoch wurde auch von der NDLTD eine eigene SGML-DTD entwickelt, die sogenannte ETD-ML (Electronic Thesis and Dissertation Markup Language). Daneben stand die Entwicklung von Konvertern, die z.B. speziell formatierte Latex oder Word-Dokumente in dieses Format umsetzen können, im Vordergrund.<sup>83, 84</sup>

### Stand der NDLTD

"The Networked Digital Library of Theses and Dissertations (NDLTD) is an *international* effort aimed at ensuring that future scholars are more effectively prepared for the Information Age [...] by understanding [electronic] publishing and digital libraries. The NDLTD also supports present day scholars by providing a forum for more timely, detailed and effective exchange of information."<sup>85</sup>

Obiges Zitat stellt deutlich die wichtigsten Aufgaben bzw. Ziele der NDLTD dar, die den Zielsetzungen vieler Projekte zur Erschließung elektronischer Dissertation entsprechen, zusätzlich jedoch auf eine *internationale* Ebene ausgerichtet sind.

Eine wesentliche Aufgabe sieht die NDLTD inzwischen in der Integration internationaler digitaler Bibliotheken in das Projekt: "NDLTD [...] may be a key player in the move to a world digital library."<sup>86</sup>

Zu diesem Zweck wurde und wird die sogenannte "*Federated Search*" entwickelt, eine Website, über die eine einzige Suchanfrage durch einen auf Java basierenden "Federated Search Server" an möglichst viele an der NDLTD teilnehmende Universitäten weitergeleitet werden soll, also eine Metasuchmaschine exklusiv für elektronische Dissertationen bzw. Diplom- und Magisterarbeiten. Im Moment sind nur 12 Universitäten an die "Federated

---

<sup>83</sup> vgl. National Digital Library of Theses and Dissertations : a Scalable and Sustainable Approach to Unlock University Resources / Edward A. Fox ...

<sup>84</sup> vgl. "Virginia Tech to create the standard for electronic theses, dissertations," Virginia Tech Press Release, Spring, 1996. - URL: <http://www.ndltd.org/news/sura.htm> [Stand 29.7.1999]

<sup>85</sup> NDLTD: Encouraging International Collaboration in the Academy / Edward A. Fox ... // In: DESIDOC Bulletin of Information Technology, September 1997. - URL: <http://www.ndltd.org/pubs/dbit.pdf> [Stand 29.7.1999]

<sup>86</sup> vgl. Fox, Edward A.: Networked Digital Library of Theses and Dissertations : a framework for east-west collaboration. Proc. of East meets West, An Asian Digital Library Workshop, 5-7 August 1998, Hong Kong. - URL: <http://www.ndltd.org/pubs/Asia98.htm> [Stand 29.7.1999]

Search" angeschlossen, aber aufgrund der vielen verschiedenen Datenbanken, die in Zukunft mit der "Federated Search" abgefragt werden sollen (62 Universitäten sind inzwischen offizielles Mitglied der NDLTD), wurde die Searchable Database Markup Language (SearchDB-ML) entwickelt, eine Applikation der eXtensible Markup Language (XML) zur Beschreibung einer Suchmaschine.<sup>87</sup> Mit der SearchDB-ML kann von den Teilnehmern der NDLTD jeweils ein "Simple Search Interface" und ein "Advanced Search Interface" beschrieben werden und damit die eigene Datenbank der elektronischen Dissertationen über die Federated Search suchbar gemacht werden.

Der wichtigste Vorteil einer solchen "Federated Search" ist zunächst die enorme Zeitersparnis auf der Suche nach Informationen: statt eine Suchanfrage in vielen verschiedenen Datenbanken jeweils neu formulieren zu müssen (und sich vorher u.U. mit der Syntax der jeweiligen Suchschnittstelle auseinandersetzen zu müssen) muß sie ein einziges Mal formuliert werden, die Distribution an die verschiedenen Datenbanken übernimmt der "Federated Search Server".

Ein weiterer Wert der NDLTD (sowie jeglicher Erschließung elektronischer Dissertationen) liegt in einem Punkt, den ich an das Ende meiner Ausführungen über die NDLTD stellen möchte und der von Fox/McMillan/Eaton treffend formuliert wurde:

"Digital libraries can help us share knowledge and culture on a world scale. [...] ETDs may be a key driving force for sharing knowledge and culture. If all TDs are captured electronically, and most are freely shared, there will be tens of thousands of new works each year, many of which will deal with topics like history, sociology, linguistics, religion, and architecture. These will directly help people learn about other cultures, while more technical works will help readers learn about methods and approaches adopted by groups in remote locations."<sup>88</sup>

---

<sup>87</sup> vgl. Powell, James ; Edward A. Fox : Multilingual federated searching across heterogeneous collections ...

<sup>88</sup> Fox, Edward A. ; Gail McMillan ; John L. Eaton: The Evolving Genre of Electronic Theses and Dissertations. // submitted for publication in: *Genre in Digital Documents*. - URL: <http://www.ndltd.org/pubs/Genre.htm> [Stand 29.7.1999]