

B-WIN4M Baden-Württemberg

Teilprojekt:

Aufbau eines elektronischen wissenschaftlichen
Publikationsverbunds

3. Meilensteinbericht

**Bibliothekarische Analyse von Publikationen an der
Universität Stuttgart und Verfeinerung des organisato-
rischen und technischen Konzepts**

Annette Maile¹ und Frank Scholze²

13.6.1997

¹ Rechenzentrum Universität Stuttgart, Abt. Informationsdienste

² Universitätsbibliothek Stuttgart

Inhaltsverzeichnis

1 EINLEITUNG	1
2 ART UND AUSWAHL DER DOKUMENTE	1
2.1 Abschätzung der maximal anfallenden Dokumente	2
2.2 Kurzbeschreibung ausgewählter Projekte anderer Institutionen	3
2.2.1 VVV der Universität Karlsruhe	3
2.2.2 MATEO der Universität Mannheim	4
2.2.3 Virtueller Medienserver des Südwestdeutschen Bibliotheks-verbunds (SWB)	4
2.3 Analyse des Ist-Zustands ausgewählter Fachbereiche der Universität Stuttgart	4
2.3.1 Informatik	4
2.3.2 Mathematik	5
2.3.3 Physik	6
2.4 Verfügbarkeit der Dokumente	6
3 VERFEINERUNG DES ORGANISATORISCHEN KONZEPTS	7
3.1 Vorbereitung der Dokumente	7
3.2 Einbringen der Dokumente auf den Dokumenten-Server und Erfassen der Metadaten	8
3.2.1 „Upload“ der Dateien auf den Dokumenten-Server	8
3.2.2 Dublin Core als Erfassungsschema der Metadaten	10
3.2.3 Weitere Arbeitsschritte	13
3.3 Speicherung und Bearbeitung der Metadaten	14
3.3.1 Speicherung	14
3.3.2 Bearbeitung	15
3.4 Retrieval der Dokumente	16
3.5 Aufbereiten und Ausgabe der Dokumente	16
4 FAZIT	17

1 Einleitung

In Forschung und Lehre werden eine Vielzahl von Dokumenten produziert. Diese sind wissenschaftliche Publikationen (Studien-, Diplom-, Doktorarbeiten), Forschungsberichte, Konferenzproceedings, Vorlesungsmanuskripte, Lehrunterlagen, Fachartikel usw. Ziel des Projekts ist der Aufbau eines elektronischen Publikationsverbunds, der Universitätsangehörige, seien es Lehrende, Lernende oder Forschende, in die Lage versetzt, für sie relevante, veröffentlichte Dokumente aus Forschung und Lehre innerhalb kürzester Zeit auf ihrem Bildschirm einzusehen und gegebenenfalls auszudrucken.

Im ersten bzw. zweiten Meilensteinbericht wurden zunächst die Online-Publikationskonzepte von Verlagen und Universitäten recherchiert. Daraufhin erfolgte die Untersuchung und Analyse der im WWW vorherrschenden Formate. Aus der Recherche der gängigen DTP-Formate im wissenschaftlichen Bereich resultierte die Untersuchung und Analyse der existierenden Konverter vom DTP-Format zum WWW-Format. Aus diesen Ergebnissen erfolgte eine erste grobe Konzeption über das Einbringen, Abspeichern und Retrieval der Dokumente.

Im ersten Teil dieses Berichts werden Art, Auswahl und Zahl der möglichen Dokumente in einem elektronischen Publikationsverbund analysiert und beschrieben. Daran anschließend folgt die organisatorische und technische Verfeinerung des Konzepts zur Behandlung dieser Dokumente.

2 Art und Auswahl der Dokumente

An der Universität Stuttgart wird eine Vielzahl von Publikationen und Dokumenten erstellt. Hierbei lassen sich grob zwei Bereiche charakterisieren:

- vornehmlich nicht im Verlag erscheinende Publikationen wie Studien-, Diplom-, Magisterarbeiten, Dissertationen, Berichte (Reports), Vorlesungsmanuskripte und andere Lehrunterlagen etc.
- Veröffentlichungen, die von kommerziellen Verlagen publiziert werden wie Monographien, Aufsätze in Zeitschriften, Kongreßberichte etc.

Grundsätzlich sollte der elektronische Publikationsverbund für alle Arten von Dokumenten offen sein, die für Lehre und Forschung an der Universität von potentiell Interesse sein können. Andererseits sollte eine völlige Beliebigkeit bei der Dokumentenaus-

wahl vermieden werden, auch wenn das potentielle Interesse von verschiedenen Benutzergruppen innerhalb der Universität sehr heterogen definiert wird. Dokumente erkennbar ephemeren Charakters sind von vornherein auszuschließen. Dazu zählen Hinweise auf Lehrveranstaltungen, Vorträge, Stellenausschreibungen etc., die ggf. an andere Stelle für das WWW aufzubereiten sind.

Die Universität Karlsruhe konzentriert sich mit ihrem Projekt eines Volltext-Veröffentlichungsverzeichnisses (VVV) auf den Bereich der Verlagsveröffentlichungen, um primär für die Forschung relevante Dokumente anzubieten.³ In Stuttgart soll der Rahmen jedoch weiter gesteckt werden, um auch Dokumente anzubieten, die bisher schlecht oder gar nicht nachgewiesen waren. Hierbei handelt es sich primär um sog. „graue“ Literatur außerhalb des kommerziellen Verlagswesens, insbesondere auch um Prüfungsarbeiten. Über die Frage eines unbeschränkten Zugangs zu Diplom- oder Masterarbeiten herrscht noch keine Einigkeit. Dementsprechend unterschiedlich sieht auch die Praxis an den einzelnen Instituten und Fakultäten aus.

Die Fachgesellschaften Chemie, Physik, Mathematik und Informatik beschäftigen sich zur Zeit intensiv mit der Frage nach digitalen Dissertationen.⁴ Davon ausgehend bietet es sich an, zunächst mit diesen Fachbereichen zusammenzuarbeiten, da diese auch zu den Initiatoren des Arbeitskreises Fachinformation auf Universitätsebene gehören und im Bereich elektronisch verfügbarer Dokumente bereits z.T. umfangreiche Vorarbeit geleistet haben. Auf diese Art und Weise läßt sich schnell und effizient eine funktionsfähige Pilotapplikation entwickeln, die dann in einen Regelbetrieb überführt werden kann. Ein System, das über das reine Teststadium hinaus ist, wird wirkungsvoller den interdisziplinären Forschungsschwerpunkten an der Universität Stuttgart zu Gute kommen, vorausgesetzt, daß diese bereit sind, ihre Dokumente uneingeschränkt einzubringen. Hierzu zählen die Bereiche Materialwissenschaft, Energie- und Umweltforschung, Verkehrs- und Kraftfahrwesen, Mikrosystemtechnik u.a.

2.1 Abschätzung der maximal anfallenden Dokumente

Verlagspublikationen der Mitarbeiter der Universität Stuttgart werden bislang nicht vollständig und systematisch erfaßt. Genaue Zahlen liegen daher nicht vor, so daß eine Abschätzung anhand von Vergleichszahlen vorgenommen werden muß.

An der Universität Karlsruhe, die mit 12 Fakultäten gegenüber 14 in Stuttgart etwas kleiner ist, werden die Verlagspublikationen (einschließlich Zeitschriftenaufsätzen) der

³ Das Risiko eines möglichen Urheberrechtskonflikts trägt dabei der Autor. Vgl. <http://www.ubka.uni-karlsruhe.de/vvv/vvvproj.html>. Hier ist die mögliche Gründung einer „University Press“ anzudenken, die sich auf elektronische Publikationen spezialisiert und ihre Autoren umfassender betreut. Vgl. dazu das Beispiel an der GHS Kassel unter <http://www.upress.de/>

⁴ http://www.educat.hu-berlin.de/diss_online/

Universitätsangehörigen seit längerem erfaßt. Diese liegen bei 4000 bis 5000 Titeln pro Jahr.⁵ In Stuttgart ist hier von einem etwas höheren Satz auszugehen. Der Anteil der von dieser Zahl tatsächlich eingehenden elektronischen Volltexte wird jedoch (wahrscheinlich nicht nur in der Anlaufphase) aufgrund der Konkurrenzsituation zu kommerziellen Verlagen deutlich darunter liegen. In Karlsruhe liegt das Verhältnis für 1996 bei 87 elektronisch erfaßten Volltexten gegenüber einer Publikationszahl von 4529 Dokumenten (Stand 10.6.97).

Zu den Verlagsveröffentlichungen kommen die Studienabschlußarbeiten. Ihre Zahl stellte sich für Stuttgart im Studienjahr 1995/96 wie folgt dar:

Art der Arbeit	SS 95:	WS 95/96:	Insgesamt
Diplom	853	820	1673
Magister	162	174	336
Promotion			450
Total			2459

(Quelle: Rechenschaftsbericht des Rektors der Universität Stuttgart)

Auch von den Studienabschlußarbeiten wird mit relativer Sicherheit nur ein Teil in den Publikationsverbund Eingang finden. Die Gründe hierfür reichen von Qualitätsansprüchen bis zu rechtlichen Fragen des fehlenden Urheberrechts für Magister- und Diplomarbeiten, die noch nicht abschließend geklärt sind.

Als Maximalschätzung ergibt sich somit eine Menge von 6000 bis 8000 Dokumenten pro Jahr, die in der Realität jedoch zunächst weit unterschritten werden wird.

2.2 Kurzbeschreibung ausgewählter Projekte anderer Institutionen

2.2.1 VVV⁶ der Universität Karlsruhe

Seit längerem erfaßt die UB Karlsruhe die bibliographischen Angaben der Publikationen der Mitarbeiter der Universität Karlsruhe im sogenannten Veröffentlichungsverzeichnis.

5

Jahr	1996	1995	1994
Dokumente	4529	4970	4575

(Quelle: Veröffentlichungsverzeichnis der Universität Karlsruhe Stand 10.6.97)

⁶ <http://www.ubka.uni-karlsruhe.de/vvv/vvvproj.html>

Im VVV (Volltext-Veröffentlichungs-Verzeichnis) werden diese Metadaten mit den zugehörigen elektronischen Volltexten verknüpft. Diese sind in den Formaten HTML, PS und PDF verfügbar. Nur bei selbständig erscheinenden Dokumenten erfolgt ein manuell eingetragener Nachweis des elektronischen Volltextes im Verbund.

2.2.2 MATEO⁷ der Universität Mannheim

Ein ähnliches Konzept hinsichtlich eines möglichst weiten Rahmens an erfaßten Dokumenten, wie es in Stuttgart angestrebt werden soll, verfolgt die Universität Mannheim mit ihrem Volltextprojekt MATEO (Mannheimer Texte Online). Zur Zeit sind jedoch nur Testdokumente bzw. Teile von Volltexten im HTML-Format abrufbar, ein Teil liegt nur im ursprünglichen DTP-Format (meist Word) vor, so daß die Problematik des Archivierungsformats als noch nicht gelöst betrachtet werden kann. Auch hier erfolgt ein manueller Nachweis der Volltexte im Verbund nur in Auswahl.

2.2.3 Virtueller Medienserver⁸ des Südwestdeutschen Bibliotheksverbunds (SWB)

In der bibliographischen Datenbank des Verbunds ist eine zunehmende Zahl von Titelaufnahmen über Links mit dem Virtuellen Medienserver verknüpft. Die Metadaten werden zentral gehalten und verweisen auf elektronische Volltexte interner und externer Server.

Da die Verbunddatenbank den zentralen bibliographischen Nachweis aller wissenschaftlichen Bibliotheken Baden-Württembergs darstellt, sollten auch möglichst alle elektronischen Dokumente, die im Rahmen des Projekts erfaßt werden, im SWB nachgewiesen werden.

2.3 Analyse des Ist-Zustands ausgewählter Fachbereiche der Universität Stuttgart

2.3.1 Informatik

Die Fakultät Informatik hat hinsichtlich eigener elektronischer Volltexte mit die umfangreichsten Vorarbeiten geleistet und bereits einen Regelbetrieb für Studien-, Diplomarbeiten und Dissertationen eingerichtet, die auf freiwilliger Basis abgeliefert werden. Eine

⁷ <http://www.uni-mannheim.de/mateo/index.html>

⁸ http://www.swbv.uni-konstanz.de/wwwroot/s71800_d.html

verpflichtende Ablieferung existiert für die Fakultätsberichte. Verlagspublikationen wurden wegen befürchteter Urheberrechtsprobleme bislang nicht erfaßt.

Als Pilotanwender nimmt die Fakultät am MeDoc-Projekt⁹ teil. MeDoc ("Multimedia electronic Documents") ist ein Verbundvorhaben, in dem am Beispiel von Informatik-Fachliteratur und auf der technischen Basis des Internet die erste digitale, online zugreifbare Volltext-Bibliothek Deutschlands aufgebaut werden soll. MeDoc hat bereits mit 12 Verlagen für 55 Werke Lizenzvereinbarungen geschlossen. Diese Bücher und Zeitschriften werden seit Oktober 1996 teilweise kostenpflichtig über das Internet angeboten.

Im Rahmen des MeDoc-Projekts hat man sich zur Unterstützung des NCSTRL-Systems¹⁰ (Networked Computer Science Technical Reports Library) für die Erschließung von Technical Reports im Volltext entschieden. Das System weist Technical Reports, die mit Metadaten in einem eigenen Format versehen werden, aus den beigetretenen Institutionen nach und arbeitet mit verteilten Servern. Etwa weltweit 80 Informatik-fachbereiche wenden das System derzeit an.

Die zur Erschließung der Volltexte benötigten Metadaten werden vom Verfasser des Textes anhand einer Erfassungsmaske eingegeben, die im WWW aufgerufen werden kann. Die Volltext-Datei (im PostScript-Format) und eine Datei mit einer Inhaltszusammenfassung/Abstract (in ASCII) werden ebenfalls vom Verfasser an den FTP-Server der Fakultät übermittelt. Recherchierbar sind die Volltexte über einen WAIS-Index. Noch nicht realisiert ist ein Hinweis auf die Volltexte innerhalb der bibliographischen Nachweisdatenbanken des Instituts (IFIBIB) und des Südwestverbundes (SWB). In diesen werden lediglich die vom Autor erfaßten und durch Bibliothekare überprüften und ergänzten bibliographischen Metadaten angeboten.

2.3.2 Mathematik

Die Fakultät Mathematik (vertreten bisher nur durch das Mathematische Institut A) bietet auf ihren WWW-Seiten Zusammenfassungen von Diplomarbeiten und Volltexte von Dissertationen (Anfang Juni 97 lagen 3 Beispiele vor) in PostScript. Wenige Preprints liegen als gezippte PS Files vor. Eine Suche auf dem Server der Fakultät ist mittels WAIS möglich.

Ein Hinweis auf die Volltexte ist über bibliographische Datenbanken (StOPAC, SWB) nicht gegeben; Preprints und Diplomarbeiten der Fakultät sind bibliographisch hier überhaupt nicht nachgewiesen.

Die Einrichtung eines zentralen Preprint-Servers (Konrad-Zuse-Zentrum für Informationstechnik Berlin)¹¹ befindet sich in Vorbereitung.

⁹ <http://medoc.informatik.tu-muenchen.de/>

¹⁰ <http://www.ncstrl.org/>

¹¹ <http://www.zib.de/>

2.3.3 Physik

Die Fakultät Physik bietet auf ihrem WWW-Server bibliographische Daten zu Aufsätzen und Preprints in chronologisch bzw. fachlich gegliederten Listen an. Teilweise liegen Abstracts (HTML, PS) vor, teilweise Volltexte (PS gezippt). Die Volltexte der Preprints sind größtenteils auch über den Preprint Server der U.S. National Science Foundation zugänglich.

Bei Diplomarbeiten werden nur zur Bearbeitung anstehende Themen nachgewiesen, nicht jedoch abgeschlossene Arbeiten.

2.4 Verfügbarkeit der Dokumente

Aufbereitung, Speicherung und Erschließung elektronischer Volltexte an der Universität Stuttgart bedürfen der Vereinheitlichung und Verbesserung. Außerhalb der Fakultät Informatik sind noch keine geregelten Verfahren zur Behandlung der Dokumente etabliert, so daß Insellösungen entstehen. Die Vollständigkeit des Dokumentbestandes könnte - auch bei einer Ablieferung der Dokumente auf freiwilliger Basis - stark verbessert werden, wenn entsprechende Empfehlungen der Universitätsgremien vorlägen. Hierzu sind Vorschläge und Entwürfe zu erarbeiten. Auf eine Änderung der Prüfungsordnungen (wie sie an der TU Chemnitz z.T. bereits vollzogen ist)¹² ist hinzuarbeiten.

Erste Sondierungsgespräche mit der Fakultät Informatik und dem Institut für maschinelle Sprachverarbeitung (Fakultät 11) zeigten die prinzipielle Bereitschaft, bereits verfügbare elektronische Dokumente in eine Pilotapplikation einzubringen. Die Universitätsbibliothek wird die in elektronischer Form vorliegenden Beiträge der Reihe *Reden und Aufsätze* der Universität Stuttgart einbringen. Ein Grundbestand an relevanten Dokumenten ist damit gesichert.

In einer Umfrage zur elektronischen Fachinformation, konzipiert von Herrn Kämper (UB) und Frau Bassen (Institut für physikalische Chemie) vom Arbeitskreis Fachinformation, wurden alle Institute der Universität u.a. auch zum gegenwärtigen Stand elektronischer Publikationen befragt. Die Institute der Fakultät Chemie wurden dabei ausführlicher auch nach der Bereitschaft zur zukünftigen Ablieferung elektronischer Volltexte gefragt. Ergebnisse der Umfrage liegen voraussichtlich erst gegen Ende Juni vor. Sie werden die Grundlage einer detaillierteren Verfügbarkeitsanalyse und eventueller weiterer Umfragen sein. Dem Arbeitskreis Fachinformation kommt mittelfristig eine wichtige Funktion hinsichtlich der Koordination und Akzeptanz eines elektronischen Publikationsverbunds zu.

¹² Für die Fakultät für Naturwissenschaften vgl. *Amtliche Bekanntmachungen* der TU Chemnitz Nr. 9 vom 15.3.1995, S. 196.

3 Verfeinerung des organisatorischen Konzepts

Das Organisationsmodell der Fakultät Informatik läßt sich bei entsprechender Modifikation auch auf andere Fakultäten übertragen. Davon ausgehend kann das bereits im Meilensteinbericht vom 1. April 1997 umrissene Konzept des Projekts organisatorisch und technisch weiter verfeinert werden.

Das Konzept gliedert sich in fünf Teile:

1. Vorbereitung der Dokumente
2. Einbringen der Dokumente auf den Dokumenten-Server und Erfassung der Metadaten
3. Speicherung und Bearbeitung der Metadaten
4. Retrieval der Dokumente
5. Aufbereiten und Ausgabe der Dokumente

3.1 Vorbereitung der Dokumente

Das Dokument soll vom Verfasser (bzw. Dokumentlieferanten) nach PDF konvertiert werden, um in diesem Format als Volltext verfügbar zu sein. Die entsprechende Software ist per Campuslizenz allen Instituten zur Verfügung zu stellen. Sollte die Konvertierung nicht geleistet werden können, so ist in Ausnahmefällen eine nachträgliche Konvertierung bei der Bearbeitung der Metadaten vorzusehen. Der Abstract (bzw. - falls ein solcher nicht vorliegt - das Inhaltsverzeichnis, die Einleitung, das Schlußkapitel/Fazit) soll vom Verfasser zur späteren Aufbereitung nach ASCII konvertiert werden. Hierzu sind Merkblätter zu erstellen, die von den entsprechenden Universitätsgremien als Empfehlung oder Richtlinie verabschiedet werden sollten.

Eventuell muß die Arbeit zusätzlich noch in einem „neutralen“ Format zur Archivierung abgespeichert werden, da PDF kein Standard ist. SGML wäre in diesem Zusammenhang wünschenswert, jedoch sind hierfür die technischen Voraussetzungen nicht gegeben. Ein Workshop zum Problem möglicher Formate bei der Tagung Dissertationen Online (Berlin, 17. - 18. März 1997) zeigte, daß die meisten Projekte im Bereich elektronischer Volltext-Dissertationen PDF bzw. PS als pragmatische Lösung bevorzugen.¹³ Ein sog. „neutrales“ Format böte hingegen die Sicherheit, daß die Dokumente jederzeit in ein aktuelles Format konvertiert werden können. Die Frage, welches Format als Standard für eine Langzeitarchivierung in Frage kommt ist noch nicht gelöst. Entsprechende Projekte auf europäischer Ebene (DLM-Forum,¹⁴ NEDLIB) sind erst in der Anlaufphase. Das Problem einer möglichen späteren Konvertierbarkeit soll jedoch weiter verfolgt werden.

¹³ http://www.educat.hu-berlin.de/diss_online/formate.html

¹⁴ <http://www.ispo.cec.be/infosoc/promo/dlmfor1.html>

3.2 Einbringen der Dokumente auf den Dokumenten-Server und Erfassen der Metadaten

Das Einbringen der Dokumente auf den Dokumenten-Server erfolgt über ein HTML-Formular mit Angabe von Metadaten. Technisch läßt sich dies auf unterschiedliche Weise realisieren.

3.2.1 „Upload“ der Dateien auf den Dokumenten-Server

Der Volltext und die ASCII-Datei (Abstract, Inhaltsverzeichnis, Einleitung, Fazit o.ä.) müssen zur weiteren Verarbeitung bzw. Organisation der Daten vom lokalen Rechner auf einen zentralen Server kopiert werden. Für diesen sogenannten Upload existieren drei Möglichkeiten.

Upload über anonymous ftp

Mittels eines FTP-Klienten loggt man sich „anonym“ auf einen FTP-Server ein, indem als Username „anonymous“ und als Passwort die email-Adresse eingegeben wird. Dann können die Dateien mit diesem FTP-Klienten auf ein Verzeichnis (meist „incoming“ genannt), welches Schreibberechtigung für jedermann besitzt, kopiert werden. Diese Möglichkeit ist etwas umständlich und nur für Personen mit entsprechenden Fachkenntnissen geeignet. Zudem kann nicht davon ausgegangen werden, daß jeder einen FTP-Klienten besitzt.

Upload über HTML-Formular

Im HTML-Formular kann ein Feld vom Typ „File“ definiert werden. Im Formular erscheint dann ein Browse-Button, mit dem auf dem lokalen Rechner die gewünschte Datei gesucht und in das Formular übernommen werden kann (siehe Abbildung 3.1). Dies ist eine sehr einfache und komfortable Lösung,

Upload auf einen FTP-Server mittels Browser

Bei Verwendung des Browsers *Netscape* (ab Version 3.0) wird im Menü *File* der Befehl *Upload File* aktiviert, wenn im Feld *Location* die Adresse eines FTP-Servers angegeben ist (siehe Abbildung 3.2). Beim Anklicken dieses Befehls können die Verzeichnisse des lokalen Rechners nach der gewünschten Datei durchsucht und auf den FTP-Server kopiert werden. Unter Windows ist es sogar möglich, die Datei vom Datei-Manager per Drag and Drop auf den Browser zu „ziehen“ und diese so auf den FTP-Server zu kopieren. Diese Möglichkeit ist ebenfalls sehr einfach in der Handhabung, jedoch derzeit nur mit Netscape ab Version 3.0 möglich.

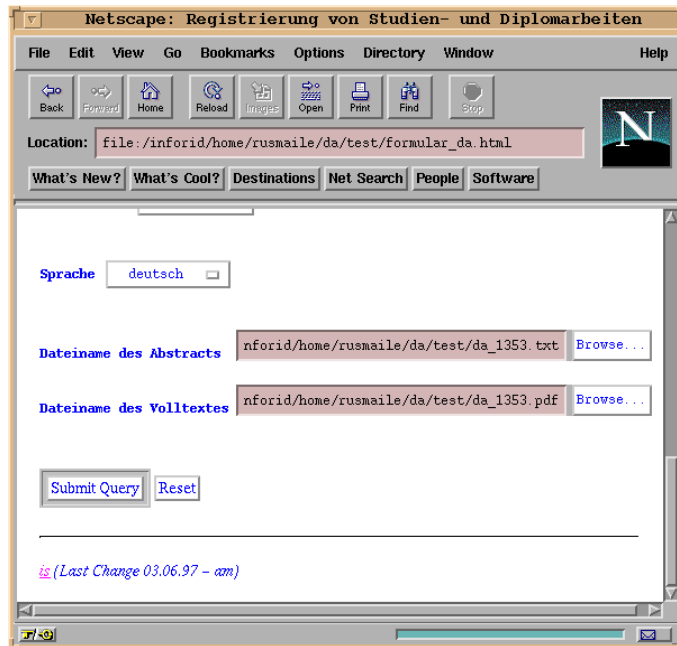


Abbildung 3.1: HTML-Formular für File-Upload

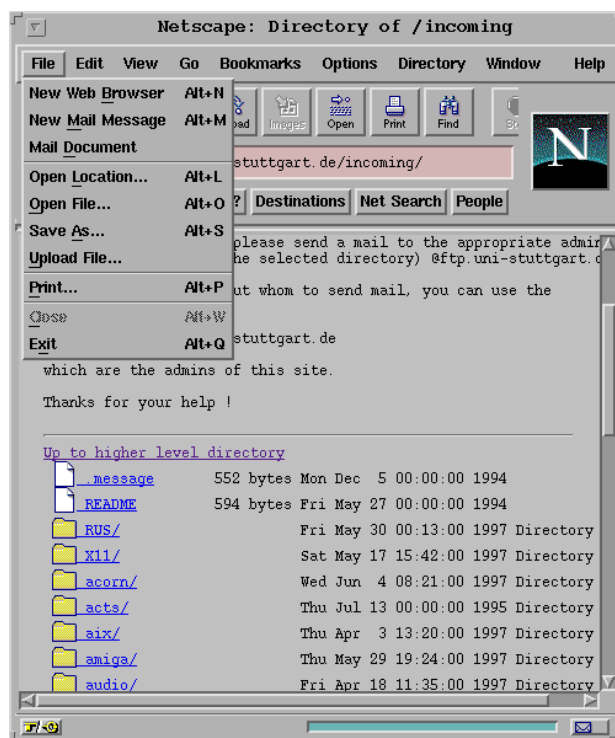


Abbildung 3.2: Aktivierung des Befehls „Upload File“ in Netscape 3.0

3.2.2 Dublin Core als Erfassungsschema der Metadaten

Als Grundlage für die Metadaten könnte der „Dublin Core Metadata Element Set“ (kurz Dublin Core bzw. DC) dienen. Die fünfzehn Basiselemente des Dublin Core dienen zur Vereinfachung der Beschreibung, der Organisation und dem Wiederfinden von Netzwerk-Ressourcen. LOC (Library of Congress), OCLC (Online Computer Library Center) und eine Reihe von Nationalbibliotheken setzen sich in Projekten mit Dublin Core auseinander bzw. stehen kurz vor dessen Einführung. Das Bibliotheksservicezentrum das die Verbunddatenbank des Südwestdeutschen Bibliotheksverbunds (SWB) betreibt nimmt an der Entwicklung des DC aktiv teil.¹⁵ Da durch den Verbund aus bibliothekarischer Sicht der organisatorische und konzeptionelle Rahmen bestimmt wird, in dem sich das Projekt bewegt, ist eine derartig offene und weit verbreitete Struktur der Metadaten vorzuziehen. Die fünfzehn Elemente des DC decken voraussichtlich die meisten der Metadaten-Attribute ab, die benötigt werden, um die einfache Beschreibung eines Dokuments zu ermöglichen. Bei nur fünfzehn Elementen werden wahrscheinlich oftmals Mehrfachbelegungen mit unterschiedlichen Inhalten auftreten. Um diese differenzieren, interpretieren und voneinander unterscheiden zu können ist die Nutzung von Qualifizierungselementen (qualifiers) vorgesehen. Die Qualifizierungselemente bestehen aus Namen und Wert. Beispielsweise gibt es ein Qualifizierungselement *Scheme*, dessen Wert den Inhalt des Basiselements *Subject* spezifiziert bzw. zu interpretieren erlaubt, z.B. als freies Schlüsselwort (uncontrolled keyword), als RSWK-Schlagwort, als DDC-Nummer (Dewey Decimal Classification) usw. Sowohl Namen als auch Wert des Qualifizierungselements sollten dabei einem kontrollierten Vokabular entstammen, um die maschinelle Interpretation von Daten unterschiedlichster Herkunft zu erleichtern.

Nicht alle der fünfzehn Kategorien des DC müssen als obligatorisch für den Verfasser bzw. Dokumentlieferanten definiert werden. Als Minimalanforderung bei der Dokumentablieferung genügen:

- Verfasser
- Titel
- Datum
- Dokumentart (Type)
- Sprache
- Name des Dokument- und Abstract/Inhalts-Files

¹⁵ Vgl. dazu http://www.swbv.uni-konstanz.de/wwwroot/metadaten/kv_dc000.html und den Aufsatz von Diann Rusch-Feja in: *Bibliotheksdienst* Heft 4, 1997 (http://www.dbi-berlin.de/dbi_pub/bd_art/97_04_08.htm)

Inhaltliche Erschließung

Ein wichtiger Teil der Metadaten ist die inhaltliche Erschließung durch Klassifikationen bzw. Schlagwörter. Gegen frei vergebene Schlagwörter spricht die mangelnde Kontrolle und der damit verbundene Informationsverlust. Allgemeine Klassifikationen, Notationen und Schlagwortsysteme (wie DK, Eppelsheimer o.ä.) sind für den Dokumentlieferanten in der Regel nicht leicht anwendbar. Hier ist ein Kompromiß zu finden, der keinen zu hohen Aufwand vom Dokumentlieferanten einerseits erfordert, andererseits jedoch auch qualitativ gleichbleibende Metadaten liefert. Als mögliche Lösung bietet sich eine Verschlagwortung nach RSWK an. Dieses Regelwerk liegt der Sacherschließung des Südwestverbundes und damit auch der UB Stuttgart zugrunde, die Konsistenz der Metadaten wäre in diesem Punkt daher gewährleistet. Auch die manuelle Überprüfung und Nachbearbeitung der Schlagwörter durch die UB wäre theoretisch möglich. Bei der Eingabe der Metadaten müßte dem Dokumentlieferanten jedoch die Schlagwortnormdatei (SWD) als Hilfsdatei zur Verfügung stehen, aus der er sich die passenden Schlagworte in das Erfassungsformular übernehmen kann.

Als zusätzliches Sacherschließungselement sind fachspezifische Klassifikationen denkbar, wobei die Zahl der verwendeten Systeme begrenzt sein sollte. Sinnvoll wäre hier ein zusätzliches Klassifikationssystem pro Fakultät, so daß maximal 14 Systeme zum Einsatz kommen.

Erfassungsformular des Bibliotheksservicezentrums

Das Bibliotheksservicezentrum hat bereits ein Erfassungsformular auf der Basis des Dublin Core entwickelt, das auf die Verhältnisse in Stuttgart angepaßt werden kann. URL: http://www.swbv.uni-konstanz.de/wwwroot/metadaten/tp_dc00.html (siehe Abbildungen 3.3 - 3.5)

Template 0: Dublin Core in "CaroLine" (einfache Erfassungsschablone)

File Edit View Go Bookmarks Options Directory Window Help

1 TITLE (Titel des zu beschreibenden Dokuments in Vorlageform):

Voller Wortlaut des Titels

Sprache des Titels :

deutsch englisch französisch spanisch italienisch

Zusatz zum Titel

Sprache des Zusatzes zum Titel :

deutsch englisch französisch spanisch italienisch

2 CREATOR (Name des Verfassers in der Form *Familienname, Vorname*):

1. Verfasser Autor oder Körperschaft

2. Verfasser Autor oder Körperschaft

3. Verfasser Autor oder Körperschaft

3 SUBJECT : Schlagwörter (Stichworte zum Thema des Dokuments, mehrere getrennt durch ", "):

Deutsches Vokabular nach [SWD](#)

verwendetes

Stichwort-System
Sprache der Stichwörter:

deutsch englisch französisch spanisch italienisch

3 SUBJECT : Klassifikation (Notationen zum Thema des Dokuments, mehrere getrennt durch ", "):

Abbildung 3.3: tp_dc00.html Seite 1

Template 0: Dublin Core in "CaroLine" (einfache Erfassungsschablone)

File Edit View Go Bookmarks Options Directory Window Help

4 DESCRIPTION (Abstrakt, Beschreibung des Inhalts):

5 PUBLISHER (Verleger, Herausgeber, Universität etc.):

6 CONTRIBUTORS (Name einer weiteren beteiligten Person in der Form *Familienname, Vorname*):

Funktion der 1. Person

Funktion der 2. Person

Funktion der 3. Person

7 DATE (creation date = Datum, zu dem das Dokument in der gegenwärtigen Form zugänglich gemacht wurde):

19970224 (ANSI.X3.30-1985)

8 TYPE (Art des Dokuments):

Dissertation

Abbildung 3.4: tp_dc00.html Seite 2

The screenshot shows a web browser window with the title "Template 0: Dublin Core in 'CaroLine' (einfache Erfassungsschablone)". The browser's menu bar includes "File", "Edit", "View", "Go", "Bookmarks", "Options", "Directory", "Window", and "Help". The main content area contains several sections for metadata entry:

- 9 FORMAT** (Datentechnisches Format des Dokuments, MIME type): A radio button is selected for "text/html" with a "MIME Type" label.
- 10 IDENTIFIKATION** (ISBN, ISSN, URL o.ä. des vorliegenden Dokuments betr. eindeutiger Identifikation): Two input fields are present. The first is labeled "URL" with a diamond icon and "URN" with a diamond icon. The second is labeled "ISBN" with a diamond icon and "ISSN" with a diamond icon.
- 11 SOURCE** (SWB-ID-Nr der Titelaufnahme des vorliegenden Dokuments in der SWB-Verbunddatenbank): A radio button is selected for "SWB-ID-Nr (8-stellig)".
- 11 SOURCE** (Werk, gedruckt oder elektronisch, aus dem das vorliegende Dokument stammt): Two input fields are present. The first is labeled "Verbale Form des Werks". The second is labeled "ISBN oder ISSN".
- 12 LANGUAGE** (Sprache des Inhalts des Dokuments): A radio button is selected for "deutsch". Other options include "englisch", "französisch", "spanisch", and "italienisch". There is also a "Liste mit anderen Sprachen" link.

At the bottom of the form, there are two buttons: "Create metadata" and "Reset". Below the form, there is a copyright notice: "(C) Copyright : BSZ - Bibliothekservice-Zentrum Baden-Württemberg", "CaroLine : Constance advanced research open Library network", and "Stand: 11.03.1997 --- Thomas Dierig --- E-Mail: tdierig@mail.swbv.uni-konstanz.de".

Abbildung 3.5: tp_dc00.html Seite 3

3.2.3 Weitere Arbeitsschritte

Zusammen mit der Erfassung der Metadaten sollte der Verfasser des Dokuments sein Einverständnis zur elektronischen Speicherung geben. Dieser Arbeitsvorgang sollte von jedem Rechnerarbeitsplatz der Universität aus durchgeführt werden können. Die Dokumentlieferanten müssen eindeutig identifiziert werden können, um Mißbrauch auszuschließen. Zusätzlich sollte die Ablieferung per Diskette bei einer zentralen Einrichtung (möglicherweise der Dissertationenstelle an der UB) oder an den Institutsbibliotheken mit entsprechendem Fachpersonal möglich sein. Dabei sind die entsprechenden Metadaten auf einem Formblatt, das dem HTML-Formular entspricht, mit abzugeben. Auf eine Vergabe von Schlagwörtern müßte ggf. verzichtet werden, wenn diese nicht bei der Abgabe durch Recherchieren in der SWD geleistet werden kann.

3.3 Speicherung und Bearbeitung der Metadaten

3.3.1 Speicherung

Für die Speicherung der Metadaten existieren zwei Möglichkeiten. Die Daten können in einer Datenbank abgespeichert werden, damit bei der Recherche gezielt nach Autor, Titel etc. gesucht werden kann. Es werden sowohl relationale Datenbanksysteme (RDBMS) als auch Volltextdatenbanksysteme geprüft. Als Beispiel für ein RDBMS wird derzeit ORACLE untersucht. ORACLE scheint viele der für das Projekt benötigten Anforderungen, wie

- WWW-Gateway
- Speicherung von Volltexten (ORACLE Text-Server)
- Volltextrecherche (ORACLE Text-Server)

zu erfüllen, jedoch müssen diese Eigenschaften noch genaueren Tests unterzogen werden.

Bei einem Workshop der Firma Dataware wurde das Volltextdatenbanksystem *BRS* vorgestellt. *BRS* speichert die Dokumente im ursprünglichen Format und in einem nur aus Text bestehendem Format, welches u.a. für die Indizierung benötigt wird. Der Nachteil dieses Systems liegt darin, daß lediglich ein eng begrenzter Ausschnitt von DTP-Systemen (hauptsächlich Word-Derivate) unterstützt wird.

Die zweite Möglichkeit besteht darin, die Metadaten im Header eines HTML-Dokuments abzuspeichern. Die Metadaten könnten im Header der HTML-Seite, die aus dem Link auf den Volltext, dem Abstract und einigen Metadaten besteht (im folgenden „Frontdoor“ genannt), stehen.

Beispiel für einen HTML-Header nach Dublin Core:

HTML-Header aus http://www.swbv.uni-konstanz.de/wwwroot/metadata/tp_dc00.html

```
<HEAD>
<TITLE>Template 0: Dublin Core in "CaroLine" (einfache Erfassungsschablone)</TITLE>
<!-- Datei: tp_dc00.html   Dierig / 05.03.1997 -->
<!-- Aufruf in: kv_dc000.html -->

<META NAME="package"      CONTENT="(TYPE=begin)(VERSION=1.0) Dublin Core">
<META NAME="DC.title"     CONTENT="Dublin Core Metadata Creator in CaroLine">
<META NAME="DC.creator"   CONTENT="(TYPE=name) Goeft, Joachim">
<META NAME="DC.publisher" CONTENT="BSZ - Bibliotheksservice-Zentrum Baden-
                          Württemberg">
<META NAME="DC.subject"   CONTENT="Metadata creation, Dublin Core, User support">
<META NAME="DC.date"      CONTENT="(TYPE=creation) 19970515">
<META NAME="DC.type"      CONTENT="Service">
```

```
<META NAME="DC.format"    CONTENT="(SCHEME=imt) text/html">
<META NAME="DC.language"  CONTENT="(SCHEME=ISO.639) de">
<META NAME="DC.rights"    CONTENT="Public domain">
<META NAME="package"      CONTENT="(TYPE=end)(VERSION=1.0) Dublin Core">
</HEAD>
```

Diverse Suchmaschinen suchen im HTML-Header, jedoch gab es bisher keine Möglichkeit, gezielt auf Zeilen, wie z.B. DC.title, zuzugreifen. An der TU München wurde kürzlich eine spezielle auf Harvest basierende Suchmaschine entwickelt, mit welcher es möglich ist, zeilenweise aus dem HTML-Header zu lesen. Jedoch ist das Eingabeformular für die Metadaten und entsprechend auch die Suchmaschine zu stark auf mathematische Publikationen ausgerichtet. Zudem ist der Source-Code sehr umfangreich und nicht ohne weiteres an die Stuttgarter Anforderungen anzupassen. Ein weiterer Nachteil sind die möglicherweise langen Antwortzeiten bei einem großen Dokumentenbestand, da die Dokumente sequentiell durchsucht werden müssen. Im Gegensatz dazu können in einer Datenbank beliebige Felder oder auch Kombinationen von Feldern indiziert werden.

3.3.2 Bearbeitung

Die Überarbeitung der Metadaten erfolgt durch Bibliothekare an der UB (bei Dissertationen) bzw. an den dafür ausgestatteten Institutsbibliotheken (bei Prüfungsarbeiten und sonstigen Dokumenten). Die dazu benötigten Daten sollten bei Dissertationen und anderen Prüfungsarbeiten aus dem Papierexemplar übernommen werden, liegt kein Papierexemplar mehr vor, muß ggf. mit dem Viewer das PDF-Dokument geöffnet werden. Eine weitergehende Automatisierung der Arbeitsvorgänge würde zwar den Personalaufwand reduzieren, sich jedoch langfristig negativ auf die Qualität und Konsistenz der Daten auswirken. Zudem müßte die Möglichkeit, Dokumente einzubringen wesentlich stärker beschränkt werden.

Ob die Metadaten automatisch ins Format des SWB konvertiert werden können ist noch offen. An einer Umsetzungskonvention DC-SWB wird von Seiten des Bibliotheksservicezentrums gearbeitet. Eine manuelle Erfassung der bibliographischen Angaben im Verbund wäre arbeitsintensiv, eine Verknüpfung der elektronischen Volltexte mit diesem Nachweisinstrument und seinen Derivaten erscheint jedoch unverzichtbar. Ein Katalogisieren derartiger Dokumente ist durch die RAK-NBM bzw. RAK-UW möglich und wird bereits von anderen Bibliotheken, wie z.B. der UB Mannheim, im Verbund realisiert.

3.4 Retrieval der Dokumente

Zur schnellen Orientierung (Browsing) sollen - innerhalb der Fakultäten chronologisch sortierte - Listen der Dokumente im WWW angeboten werden. Zu überlegen ist, ob eine zusätzlich Browsing-Möglichkeit bestehen sollte, d.h., daß die Arbeiten beispielsweise nach Autor oder Titel sortiert zur Verfügung stehen.

Das Suchen nach bestimmten Arbeiten erfolgt ebenso wie die Eingabe der Metadaten über ein HTML-Formular. Dort können die gesuchten Begriffe (Autor, Schlagworte etc.) eingetragen werden. Diese Daten werden an ein CGI-Programm geschickt, welches in der Datenbank die entsprechenden Arbeiten heraussucht und diese dann als Link auf dem Bildschirm ausgibt.

Die zweite Möglichkeit besteht darin, in einer Volltextrecherche über den Abstract und die Metadaten diejenige Arbeiten herauszufinden, die die angegebenen Worte enthalten. Die Ergebnisse einer Suche werden auch hier als Link auf die Arbeit ausgegeben.

Ein zusätzliches wichtiges Rechercheinstrument stellen - wie bereits erwähnt - die bibliographischen Datenbanken SWB und StOPAC dar, zu denen eine Verknüpfung hergestellt werden soll.

3.5 Aufbereiten und Ausgabe der Dokumente

Nach der erfolgreichen Suche nach Dokumenten, wird zunächst ein Link auf dem Bildschirm ausgegeben, wie oben schon erwähnt. Dieser Link zeigt jedoch nicht sofort auf den Volltext, sondern auf ein aus dem Erschließungsfile (Abstract, Inhaltsverzeichnis usw.) und den Metadaten generiertes HTML-Dokument (siehe Abbildung 3.6).

Dieses HTML-Dokument wird von einem CGI-Programm erzeugt, welches die Metadaten aus der Datenbank und den Abstract/Inhaltsverzeichnis etc., der in ASCII vorliegt, einliest. Dieses HTML-Dokument enthält den Link auf den Volltext. Der Benutzer kann nach Durchsicht des Abstract und der Metadaten entscheiden, ob der Volltext interessant sein könnte oder nicht.



Abbildung 3.6: Beispiel für die Aufbereitung von Abstract und Metadaten

4 Fazit

Ein Dokumentenerstbestand für eine Pilotapplikation ist akquiriert und als Datenbestand verfügbar. Die Arbeitspakete A1 und A2 sind damit abgeschlossen, notwendige Bedarfsanalysen werden während der gesamten Projektlaufzeit weiter betrieben. Die Erschließung des Datenbestandes durch Katalogisierungs- und Klassifikationsschemata (Arbeitspaket A3) ist umrissen worden, muß jedoch noch weiter verfeinert werden. Das zugrundeliegende Organisationsmodell, das im 1. und 2. Meilensteinbericht vorgestellt wurde, konnte nach bibliothekarischen Gesichtspunkten verfeinert werden. Es wird im Projektverlauf durch Analysen und Praxistests weiter modifiziert und optimiert werden. Einzelne Komponenten, wie z.B. Upload von Files, Generierung der Frontdoor aus den Metadaten und der ASCII-Datei etc. wurden prototypisch implementiert. Verschiedene

konzeptionelle Ansätze zur Speicherung der Metadaten (RDBMS, Volltextdatenbanksystem, Harvest-basierte Suchmaschine) wurden analysiert und vergleichend bewertet. Auf dieser Grundlage soll die endgültige Auswahl eines Systems erfolgen.

Über den aktuellen Stand des Projekts wird auch auf der schon seit längerem bestehenden Homepage berichtet. URL: http://www-ks.rus.uni-stuttgart.de/B4M/b4m_elpub.html
Der Arbeitstitel für den zukünftigen **Online Publikationsverbund** der **Universität Stuttgart** lautet OPUS.