Elektronisches Publizieren – Etablierung eines neuen Service für die Universität durch UB und CMS

Susanne Dobratz, dobratz@cms.hu-berlin.de

Kurzfassung: Das elektronische Publizieren durch die Universitätsbibliothek und den Computer- und Medienservice (Rechenzentrum) begann 1997 mit dem Projekt »Digitale Dissertationen«. Heute hat sich aus diesem und aus weiteren Projekten mit dem Dokumenten- und Publikationsserver als zentralem Publikationsort ein festes Dienstleistungsangebot im Bereich des elektronischen Publizierens herausgebildet, welches allen Mitgliedern der Humboldt-Universität zur Verfügung steht.

Von den Digitalen Dissertationen zu einem Dienstleistungsangebot für die Universität

Schnell, kostengünstig, up to date! So riefen wir die Doktoranden 1997 dazu auf, an einem neu entstehenden Angebot zur elektronischen Publikation ihrer Dissertationen teilzunehmen, das innerhalb des Projekts »Digitale Dissertationen«[1] entwickelt wurde. Heute existiert eine gemeinsame Arbeitsgruppe »Elektronisches Publizieren«[2] der Universitätsbibliothek und des Computer- und Medienservice, in der sowohl Mitarbeiter aus beiden Einrichtungen als auch Personal aus den verschiedenen Projekten zu den Themen »elektronisches Publizieren« und »digitale Bibliothek« zusammenarbeiten. Damit ist das elektronische Publizieren zu einer festen Dienstleistung für die Universität geworden, deren Schwerpunkte in der Betreuung technischer und bibliothekarischer Konzeptionen digitaler Publikationsvorhaben, bei der direkten Autorenbetreuung sowie in der Wartung und Weiterentwicklung des Dokumenten- und Publikationsservers liegen. Der Dokumenten- und Publikationsserver wird in Entwicklungsprojekten ständig verbessert und erweitert – mit dem Ziel den Aufbau der digitalen Bibliotheken der Humboldt-Universität durch einen digitalen Publikationsservice zu unterstützen.

1 http://edoc.hu-berlin.de/epdiss/
2 http://edoc.hu-berlin.de/e_info/epub.php.

Der Dokumenten- und Publikationsserver als Kern des elektronischen Publizierens

Den Kern des Dienstleistungsangebots und der weiteren Entwicklungen bildet der Dokumenten- und Publikationsserver (*http://edoc.hu-berlin.de*), der nach besonderen organisatorischen und technischen Spezifikationen aufgebaut wurde.

Er beinhaltet zurzeit ca. 1500 digitale Dokumente, davon zwei rein digital erscheinende Zeitschriften – »SPEPS« (Stochastic Programming E-Print Series) und »kunsttexte.de«, die von Mitgliedern der Humboldt-Universität herausgegeben werden, sowie zwei fortlaufende Bände (»Historisches Forum« und »Historische Literatur«) und drei komplette Tagungsbände mit insgesamt mehr als 200 Beiträgen. Daneben wurden bisher ca. 800 Dissertationen, 200 Habilitationsschriften, 30 Magister- und Diplomarbeiten sowie 150 Öffentliche Vorlesungen der Humboldt-Universität – eine Publikationsreihe der Forschungsabteilung – veröffentlicht. Das Angebot wird durch nachträglich digitalisierte historische Dissertationen berühmter Persönlichkeiten, z. B. von Carl Friedrich Gauß, Max Planck, Heinrich Hertz, Friedrich Hegel und Rudolph Virchow sowie die Zeitschrift des Computer- und Medienservice (ehem. RZ-Mitteilungen) ergänzt.

Was wird auf dem Dokumentenserver veröffentlicht?

Der Dokumenten- und Publikationsserver bietet jedem Angehörigen der Humboldt-Universität sowie den Nutzern des Erwin Schrödinger-Zentrums in Adlershof die Möglichkeit, die eigenen Publikationen digital zu publizieren und weltweit in entsprechende Retrievalnetzwerke einzubinden. So können Fachkollegen im In- und Ausland gezielt in bibliothekarischen Netzwerken nach Publikationen suchen und die für sie interessanten Veröffentlichungen ohne lange

Abb. 1: Der Dokumenten- und Publikationsserver der Humboldt-Universität zu Berlin

Einbindung des Dokumenten- und Publikationsservers weltweit einem entsprechenden Fachpublikum zur Verfügung gestellt werden. Durch das bloße Bereitstellen digitaler Publikationen durch die Universitätsbibliothek erweitert sich deren klassischer Sammelauftrag zur Speicherung, Katalogisierung und Archivierung aller durch Angehörige der Humboldt-Universität veröffentlichten wissenschaftlichen Arbeiten um die digitalen Publikationen.

Bestellfristen sofort und ohne Einschränkungen lesen.

Um eine hohe inhaltliche Qualität der publizierten Dokumente zu gewährleisten, wurden Dokumentenkategorien eingeführt und Bedingungen an diese Kategorien geknüpft:

1. Publikationen und Publikationsreihen der Humboldt-Universität mit wissenschaftlichen Inhalten wie die Öffentlichen Vorlesungen der Humboldt-Universität oder die Schriftenreihe der Universitätsbibliothek,
2. durch Angehörige der Humboldt-Universität herausgegebene Publikationen und Publikationsreihen wie Sammelwerke, Kongressbände, Forschungsberichte, Zeitschriften (E-Journals),
3. Dokumente, deren Veröffentlichung durch Prüfungsordnungen vorgeschrieben ist (Dissertationen und Habilitationsschriften),
4. Publikationen und Publikationsreihen von Einrichtungen und Personen, die mit der Humboldt-Universität assoziiert sind sowie
5. Dokumente von Studierenden der Humboldt-Universität wie Diplom-, Magister- und Seminararbeiten, wenn deren Publikation durch einen Betreuer bzw. Gutachter empfohlen wird.

Dabei werden aus inhaltlicher, konservatorischer oder Benutzungssicht auch retrodigitalisierte Bestände der Universitätsbibliothek und anderer Einrichtungen mit einbezogen. Die an der Humboldt-Universität entstehenden wissenschaftlichen Publikationen sollen über die

Wodurch zeichnet sich eine elektronische Publikation aus?

Eine Webseite oder eine Textdatei ins Internet zu stellen, ist heutzutage eine leichte Übung. Verbindet man jedoch mit der elektronischen Publikation eines Dokuments ähnliche Ansprüche wie mit der herkömmlichen papiergebundenen Veröffentlichung eines Buches oder einer wissenschaftlichen Zeitschrift (z. B. Langzeitverfügbarkeit und Zitierbarkeit, Integrität und Authentizität), ergibt sich schnell eine Reihe von Anforderungen und Kriterien, deren Umsetzung für den Dokumenten- und Publikationsserver zu den Hauptaufgaben der Arbeitsgruppe »Elektronisches Publizieren« zählt.

Dauerhafte Zitierbarkeit

Um auf die elektronischen Dokumente jederzeit über einen eindeutigen Weg zugreifen zu können, arbeitet der Dokumentenserver mit so genannten URNs, Uniform Resource Names, in Form der National Bibliographic Number (NBN). Durch sie werden die Dokumente mit individuellen und dauerhaften Adressen versehen, welche einen vom aktuellen Rechner und der WWW-Schnittstelle, d. h. vom http-System unabhängigen, unmittelbaren Zugriff auf die Dokumente erlauben.

Der Vorteil dieser Methode liegt darin, dass auf ein Dokument nicht mehr über den unmittelbaren Aufruf einer URL (z. B. *http://edoc.hu-berlin.de/dissertationen/agrar/otto-grit/HTML/*

index.html) zugegriffen wird. Stattdessen wird es über einen eindeutigen und vom Übertragungsprotokoll unabhängigen Identifikator angesprochen (für das Beispiel *urn:nbn:de: kobv:11-10011214*), der sich auch dann nicht mehr ändert, wenn sich zum Beispiel der Speicherort der Datei auf einen anderen Rechner verschiebt. Da sich ein Dokument letztendlich natürlich immer hinter einer konkreten URL verbirgt, ist ein so genannter Resolving-Mechanismus notwendig, der eine URN in die jeweils aktuelle http-Adresse des Dokuments auflöst. Dieser Dienst wird durch Die Deutsche Bibliothek (DDB) in Frankfurt/Main realisiert. Der Resolver-Mechanismus kann unter Verwendung einer speziellen URL (für die angegebene URN z. B. *http://nbn-resolving.de/urn/resolver.pl?urn: nbn:de:kobv:11-10011214)* oder mithilfe eines durch die DDB zur Verfügung gestellten Browser-Plugins[3] genutzt werden. Letzteres erlaubt die direkte Eingabe einer URN in die Adresszeile des Browsers. In beiden Fällen wird man dann zum tatsächlichen Dokument weitergeleitet. Sollte sich zum Beispiel der Rechnername des Servers oder ein anderer Bestandteil der URL einmal ändern, muss nicht in allen Bibliothekskatalogen die URL geändert werden. Lediglich der Resolving-Rechner benötigt die neue Zurodnung zwischen URN und URL und muss diese in seiner internen Datenbank vorhalten.

Abb. 2: Suche nach »Miniaturschwein« in der NDLTD Suchmaschine

Weltweite Verfügbarkeit und Recherchierbarkeit

Das Auffinden der elektronischen Dokumente ist über Bibliothekskataloge, Recherchen in bibliographischen Metadaten, Suchtermen innerhalb der Strukturen der elektronischen Dokumente und über alphabetische und klassifikatorische Ordnungen sowie dynamisch erstellte Listen und Indizes möglich.

Die Veröffentlichungen werden auch durch Anwendung neuester technischer Entwicklungen, wie der des Protokolls der Open Archives Initiative[4] in nationale und internationale Retrievalnetzwerke einbezogen. So können alle digitalen Dokumente, die auf dem Dokumenten- und Publikationsserver der Humboldt-Universität veröffentlicht wurden, auch bei der DDB[5] gefunden werden. Ebenso werden sie an verschiedenen Stellen der Welt über die OAI-Schnittstelle in Suchmaschinen eingebunden, so zum Beispiel bei der Networked Digital Library of Theses and Dissertations (NDLTD)[6], dem VTLS OAI Union Catalog[7], dem ARC Search Service[8] oder auch beim infoball-Suchportal[9]. Digitale Publikationen des edoc-Servers sind auch über die Suchmaschine Google[10] auffindbar.

Es wird ebenso eine Volltextsuche angeboten, die das Durchsuchen des gesamten Dokumenteninhaltes ermöglicht. Damit werden bessere Trefferquoten erzielt.

In Vorbereitung ist auch eine Suchmöglichkeit, die den Anforderungen der Nutzer dadurch noch besser gerecht wird, dass sie Strukturinformationen des Dokumentes mit einbezieht. Dies wird dadurch möglich, dass alle Dokumente, die für eine Langzeitverfügbarkeit ausgewählt wurden, in einem speziellen XML-Format vorgehalten werden und somit Infor-

3 siehe http://www.nbn-resolving.de
4 Siehe Informationen unter http://www.openarchives.org oder [3]
5 http://deposit.ddb.de
6 http://jin.dis.vt.edu/fedsearch/ndltd/support/search-catalog.html
7 http://rocky.dlib.vt.edu/~etdunion/cgi-bin/index.pl
8 http://arc.cs.odu.edu/
9 http://www.infoball.de
10 http://www.google.de

Abb. 3: Anzeige der Fundstelle – eine Abbildung, die ein Miniaturschwein zeigt.

mationen über Überschriften, Abbildungen, Tabellenbeschriftungen, Keywords u. ä. enthalten.

Die XML-Suchmaschine kann speziell in den semantischen Tags suchen. So kann zukünftig beispielsweise die Suche nach relevanter Literatur auf Überschriften oder Abbildungsbeschriftungen (im Beispiel das Tag <ueberschrift>) eingeschränkt werden. Damit werden nur solche Dokumente gefunden, die das entsprechende Suchwort in einer Überschrift bzw. einer Abbildungsbeschriftung enthalten. Bei wissenschaftlichen Arbeiten hat dies den Vorteil, dass eine Fundstelle wesentlich mehr Relevanz hat, da man davon ausgeht, dass bei der Erwähnung eines Suchwortes in der Überschrift im nachfolgenden Kapitel oder Absatz auch wirklich nähere Erläuterungen zum Suchwort stehen. Bei einer bloßen Stichwortsuche im Volltext werden auch Fundstellen angezeigt, die ggf. nur eine Aufzählung von Begriffen enthalten, ohne wirklich Informationen zum Suchwort zu liefern.

Von besonderem Interesse ist unter anderem die Suche nach bestimmten Abbildungen. Am Beispiel des Suchwortes »Miniaturschwein« lässt sich anhand der Trefferliste und der eigentlichen Fundstellen im Dokument sehen, dass nur die relevanten Treffer in Abbildungsbeschriftungen aufgeführt worden sind (s. Abb. 3).

Einbindung neuer Medientypen

Das elektronische Publizieren ermöglicht es, Dokumente um zusätzliche Medien zu erweitern. So gibt es bereits einige Dissertationen und Habilitationsschriften, die sich dieser Möglichkeit bedienen und zur besseren Veranschaulichung der wissenschaftlichen Ergebnisse Audio- und Videodateien und Animationen beinhalten.[11]

Gesicherter Inhalt, Autor und Veröffentlichungszeitpunkt

Um jederzeit beweisen zu können, dass die über den Dokumenten- und Publikationsserver bereitgestellten Dokumente technisch und inhaltlich nicht von außen verändert wurden und dass sowohl der angegebene Autor als auch der Veröffentlichungszeitpunkt authentisch sind, werden Verfahren zur Sicherung des Servers und der Dokumente eingesetzt, die sich digitaler Signaturen und Zeitstempeln bedienen.

Hier wird im Wesentlichen durch einen kryptografischen Algorithmus eine eineindeutige Zeichenkette für jedes Dokument (ein sogenannter Hashcode) erzeugt. Diese wird durch den Betreiber des Dokumentenservers verschlüsselt sowie mit einer digitalen Signatur und einem digitalen Zeitstempel versehen.

Sollte im Internet eine weitere Variante des digitalen Dokuments auftauchen, können beide Dokumente über ihren Hashcode miteinander verglichen werden. Der Hashcode für das Dokument auf dem edoc-Server wird durch Entschlüsselung der signierten Variante erzeugt. Zur Erzeugung des Hashcodes für das neue Dokument wird der gleiche kryptografische Algorithmus benutzt wie bei dem über den Dokumentenserver bereitgestellten Dokument. Gibt es Unterschiede, so kann man davon ausgehen, dass das neue Dokument eine veränderte Version und der Inhalt nicht vollständig identisch mit dem Dokument ist, das der Autor auf dem edoc-Server deponiert hat. Mithilfe des digitalen Zeitstempels ist nachweisbar, zu welchem Zeitpunkt das echte Dokument vorgelegen hat.

Für die Signierung der Dokumente wird eine qualifizierte digitale Signatur mit Anbieterakkreditierung im Sinne des Deutschen Signaturgesetzes beim Anbieter Telesec GmbH[12] genutzt.

11 z. B. http://edoc.hu-berlin.de/dissertationen/riede-tobias-2000-06-26/

12 http://www.telesec.de/ (Suche nach Zertifikat mit Namen Dig. Dissertation)

Technologie des Servers

Leitlinien[13] regeln den Betrieb des Dokumentenservers. Sie schreiben die organisatorischen und technischen Rahmenbedingungen für den Betrieb des Servers fest. Ziel ist es, das Angebot, digitale Publikationen von Universitätsangehörigen über den Dokumenten- und Publikationsserver bereitzustellen, zu erweitern. Anders als bei einem »normalen« WWW-Server wird bei der Betreuung, Wartung und Weiterentwicklung des Dokumenten- und Publikationsservers besonderer Wert auf die Nutzung neuester Technologien beim elektronischen Publizieren gelegt.

Aus diesem Grunde betreut die Arbeitsgruppe »Elektronisches Publizieren« gern jeden interessierten Autor oder Herausgeber bei der Vorbereitung seiner Publikation für die digitale Veröffentlichung und erarbeitet mit ihm gemeinsam individuelle Lösungen und technische Strategien für seine Publikationsvorhaben. Dabei liegen die Schwerpunkte des Supports speziell in der Programmierung von Dokumentvorlagen und XML-Dokumenttypdefinitionen (DTD) für verschiedene Dokumentarten. Diese liefern Strukturvorschriften für den Aufbau von bestimmten Dokumentarten, so beschreibt z. B. die DiML.DTD (Dissertation Markup Language) den Aufbau und die Abfolge von Strukturelementen wie Überschriften, Textkörpern, Absätzen, Abbildungen, Abbildungsbeschriftungen in digitalen Dissertationen.

Allerdings liegt hier das XML-Dateiformat zugrunde, das nur wenige gängige Textverarbeitungsprogramme unterstützen. XML = eXtensible Markup Language ist ein Format, welches vom W3C[14] standardisiert wurde, anbieterunabhängig ist und auf verschiedenen Hard- und Softwareplattformen ohne Spezialprogramme verarbeitbar ist. Es genügt ein einfacher UNICODE[15]-fähiger Editor. XML-Dokumente haben den Vorteil, dass in ihnen semantische Informationen, wie z. B. Überschriften, durch so genannte Tags (Beispiel: <ueberschrift>Beispiel</ueberschrift>) gekennzeichnet werden können.

Was müssen Autoren tun?

Als Gegenstück zur DiML.DTD gibt es speziell dafür entwickelte Dokumentvorlagen für Microsoft Word, OpenOffice sowie Staroffice. Damit können Doktoranden, die die Möglichkeit nutzen wollen, ihrer Publikationspflicht für die Dissertation durch die elektronische Publikation nachzukommen, ihre Dokumente entsprechend den Regelungen der Universitätsbibliothek aufbereiten und damit die Grundlage dafür schaffen, dass ihre Arbeit in hoher technischer Qualität im Internet veröffentlicht werden kann.

Da nicht jeder Autor ein Spezialist im Umgang mit seinem Textverarbeitungsprogramm sein kann, bietet die Arbeitsgruppe »Elektronisches Publizieren« monatliche Word-Kurse an, in denen nicht nur die Benutzung der Dokumentvorlage gelehrt wird, sondern auch eine Einführung in die wichtigsten Funktionen des Textverarbeitungssystems erfolgt. Oft können damit Dokumente einfacher, schneller und ohne Formatierungsprobleme erstellt werden.

Die digitalen Dokumente werden dann natürlich nicht im Ursprungsformat auf dem Dokumenten- und Publikationsserver veröffentlicht, sondern nach XML konvertiert. Der Vorteil liegt darin, dass diese Version als Archivformat genutzt wird und dass sich daraus beliebige andere Erscheinungsformen der Arbeiten, wie z. B. eine HTML-Version für das normale Darstellen in einem WWW-Browser, automatisch generieren lassen.

Digitale Konferenzbände, die besondere Chance für Tagungsorganisatoren und -teilnehmer

Besonders gut eignet sich das elektronische Publizieren zur Veröffentlichung von Konferenzbänden, da hier die Möglichkeit besteht, bereits vor Beginn der Konferenz einen digitalen Konferenzband in der Form anzubieten, die es dem Leser ermöglicht, seinen individuellen Konferenzband mit ausgewählten Beiträgen zusammenzustellen und über den angebotenen Print-On-Demand-Service herstellen zu lassen. So muss sich der Leser nicht mit einem großen, gedruckten, das gesamte Themenspektrum der Tagung abdeckenden Konferenzband zufrieden geben, sondern kann gezielt über die Nutzerschnittstelle des Dokumenten- und Publikati-

13 http://edoc.hu-berlin.de/e_info/leitlinien.php
14 World Wide Web Consortium (http://www.w3.org)
15 UNICODE ist eine Weiterentwicklung des ASCII Standards.

onsservers einen seinen Interessen entsprechenden Konferenzband zusammenstellen und wird dann automatisch zum Warenkorb des Proprint-Dienstes geleitet[16].

Dieser Dienst schließt verschiedene Dokumentenserver zusammen, momentan den der Staats- und Universitätsbibliothek Göttingen, den der Universität Graz und den der Humboldt-Universität. Der Dienst stellt die ausgewählten Konferenzartikel dann zu einem Dokument zusammen, welches der Leser wahlweise als PDF-Dokument herunterladen oder als Druckauftrag zu einem der angeschlossenen Druckdienstleister weiterleiten kann. Dort wird gegen eine Druckgebühr dann ein Druckexemplar erstellt und dem Leser zugestellt. Dadurch, dass der Leser nur die für ihn relevanten Artikel für den Band ausgewählt hat, erhält er einen individuellen Tagungsband.

Abb. 4: Beispiel für einen digitalen Konferenzband: »Spätmittelalterliche Wallfahrt im mitteldeutschen Raum«

Ausblick

Die Dienstleitungen der Arbeitsgruppe »Elektronisches Publizieren« werden, wie hier an einigen Beispielen demonstriert, auch in Zukunft ausgebaut. Ziel ist es, für die Humboldt-Universität eine digitale Bibliothek aufzubauen, in der die Angehörigen der Universität auf der einen Seite ihre Publikationen weltweit veröffentlichen und zur Verfügung stellen können. Auf der anderen Seite sollen sie in den verschiedenen weltweit verfügbaren digitalen Ressourcen, seien sie frei verfügbar oder von der Bibliothek der Humboldt-Universität käuflich erworben, recherchieren und so auf die für ihren Forschungs- und Lehrbereich relevante Literatur Zugriff erhalten. Letzterem Ziel dient das im Aufbau befindliche Bibliotheksportal der Humboldt-Universität[17] (Metalib).

Literatur

[1] International Organization for Standardization (ISO). ISO 8879: 1986. Information Processing – Text and Office Systems – Standard Generalized Markup Language (SGML). International Organization for Standardization. Ref. No. ISO 8879: 1986 (E). Geneva/New York, 1986.

[2] OAI Technical Committe: Open Archives Initiative Protocol for Metadata Harvesting – Version 2.0, http://www.openarchives.org/OAI/openarchivesprotocol.html (10.02.2003)

[3] Clifford A. Lynch: »Metadata Harvesting and the Open Archives Initiative«, ARL: A Bimonthly Report on Research Library Issues and Actions 217 (August 2001), http://www.arl.org/newsltr/217/mhp.html

[4] DCMI Elements and Element Refinements – a current list, http://www.dublincore.org/usage/terms/dc/current-elements/ (28.01.2003)

[5] Signaturverordnung (SIGV), Bundesgesetzblatt Teil I G5702, Nr. 59, Bonn, 21.11.2001, http://www.sicherheit-im-internet.de/download/SIgVO_dr.pdf (28.01.2003)

16 http://edoc.hu-berlin.de/proprint
17 http://digibib.ub.hu-berlin.de