

Werner Stephan / Frank Scholze (UB Stuttgart):

Online Publikationsverbund: Erfassung und Organisation elektronischer Hochschulschriften¹

In Forschung und Lehre wird heute bereits eine Vielzahl von Dokumenten mittels elektronischer Werkzeuge in druckreifer Qualität erstellt. Dazu zählen unter anderem wissenschaftliche Prüfungsarbeiten (Studien-, Diplom-, Doktorarbeiten), Forschungsberichte, Konferenzproceedings, Vorlesungsmanuskripte und Lehrunterlagen, d.h. Dokumentarten, die aufgrund ihrer schlechten Erschließung und Verfügbarkeit oftmals als „graue Literatur“ zusammengefaßt werden. Gleichzeitig erfordert die Suche nach relevanten Dokumenten und das Verbreiten von Informationen und Forschungsergebnissen einen immer höheren Zeitaufwand.

Um diese Situation zu verbessern, wurde an der Universität Stuttgart vom Rechenzentrum und der Universitätsbibliothek ein Projekt des Deutschen Forschungsnetzes durchgeführt, das unter dem Namen OPUS (Online Publikationsverbund der Universität Stuttgart) die technischen und organisatorischen Voraussetzungen für ein effizienteres Anbieten und Erschließen von elektronischen Volltext-Dokumenten im Hochschulbereich schafft.² Jeder Universitätsangehörige, sei es Lehrender, Lernender oder Forschender, soll dabei vom eigenen Arbeitsplatz aus Dokumente im World Wide Web veröffentlichen und gleichzeitig auch durch unterschiedliche Suchmöglichkeiten gezielt auf elektronische Texte zugreifen können. Eine Grundidee des Projektes besteht darin, in Zeiten knappster Personalressourcen eine möglichst große Zahl von Arbeitsabläufen zu automatisieren und gleichzeitig dem Verfasser die Vorbereitung - und damit auch die Verantwortung und Kontrolle über die Gestaltung - seines zu veröffentlichenden Dokuments zu überlassen. Im Zusammenhang mit der Tatsache, daß die Autoren zusätzlich ihr Dokument beschreibende Angaben mitliefern ist bereits der betriebswirtschaftliche Begriff des "Outsourcing"³ in diesem Bereich bibliothekarischer Tätigkeiten eingeführt worden.

Veröffentlichen

Zur Umsetzung dieser Grundidee des Projektes, bietet sich derzeit als Ablieferungs- und gleichzeitig Präsentationsformat lediglich PostScript bzw. das daraus weiterentwickelte PDF (Portable Document Format) an. Diese Aussage will weniger im apoktischen als vielmehr pragmatischen Sinne verstanden werden. Das Problem der Langzeitarchivierung und Konvertierbarkeit elektronischer Dokumente ist im Rahmen des gewählten Ansatzes sicher noch nicht gelöst. In diesem Punkt ist vor allem die Entwicklung von SGML⁴ bzw. XML zu verfolgen. Die gewonnene Erfahrung durch

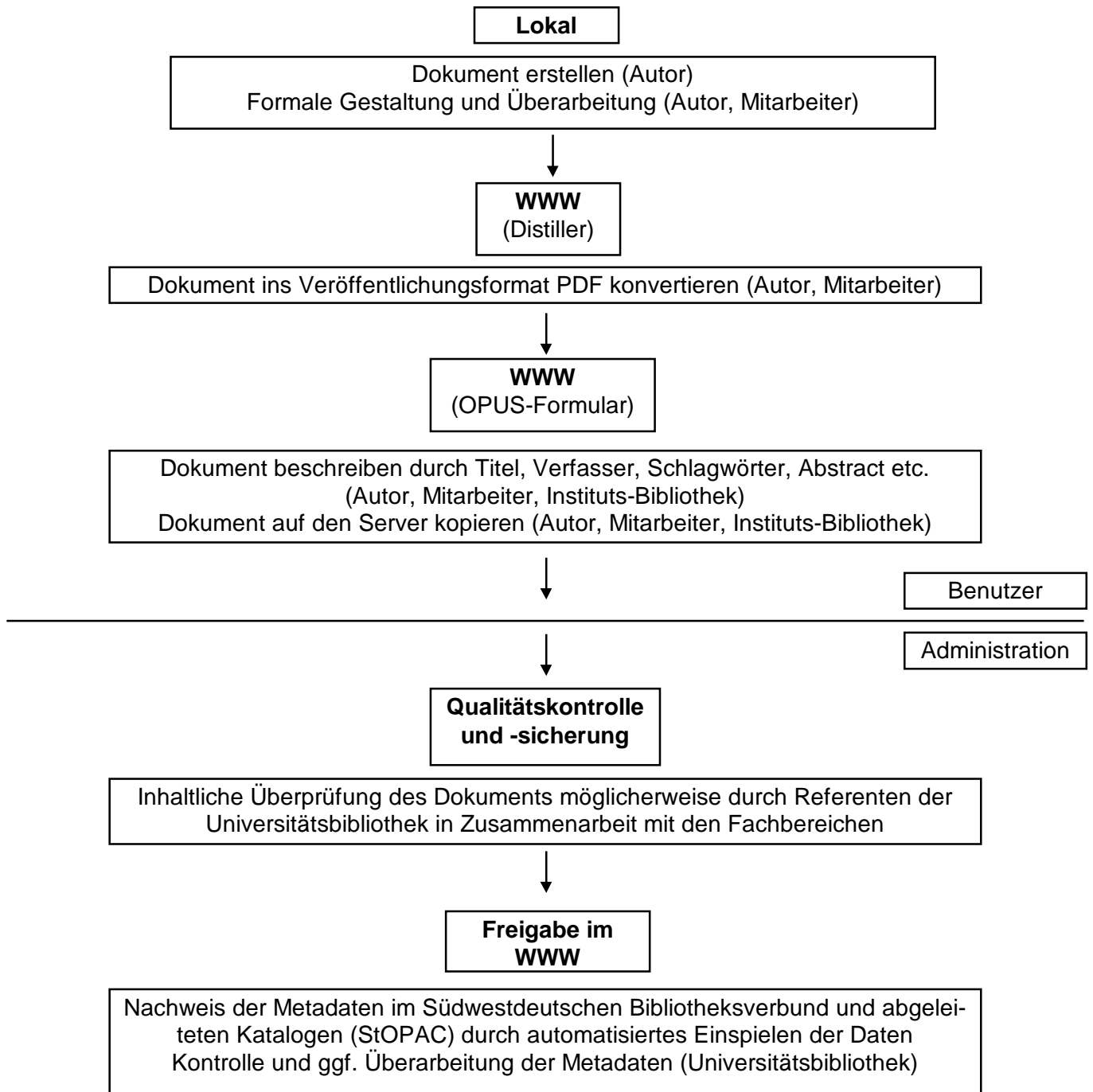
¹ Dieser Artikel entstand - in geringfügig überarbeiteter Form - aus einem Vortrag beim 25. Österreichischen Bibliothekartag, St. Pölten, 15. - 19. September 1998.

² <http://www.uni-stuttgart.de/opus/>

³ Verlagerung von Arbeitsschritten aus dem Unternehmen. Vgl. z.B. Eileen Degenhardt, "Elektronische Dissertationen" in *Bibliotheken*, Diplomarbeit, Fachhochschule Hannover, Fachbereich Informations- und Kommunikationswesen, 1998, S. 78. <http://www.ik.fh-hannover.de/ik/personen/bock/degenhardt/ediss.pdf>

⁴ Vgl. Daniel Ohst, „Dateiformate für das elektronische Publizieren“, Studienarbeit, Humboldt-Universität Berlin, 1998. <http://www2.rz.hu-berlin.de/~h0444saa/didi/formate.html> bzw. die Ergebnisse im Rahmen des Projekts „Digitale Dissertationen“ (DiDi) an der Humboldt-Universität. <http://dochost.rz.hu-berlin.de/epdiss/>

Agieren im Rahmen einer derzeit praktikablen Lösung kann unserer Meinung nach jedoch ebenso entscheidend zur Lösung längerfristiger Probleme in diesem Sektor beitragen, wie die theoretische Auseinandersetzung unter Testbedingungen. PostScript kann plattformunabhängig aus fast allen Anwendungsprogrammen durch drucken in eine Datei erzeugt werden. Auf dem nur für Angehörige der Universität Stuttgart zugänglichen PC-Campus-Server des Rechenzentrums können diese Dateien dann mit Hilfe des netzwerkfähigen Konvertierungsprogramms Acrobat Distiller in PDF umgewandelt werden. Damit steht allen Universitätsangehörigen, trotz sehr heterogenen technischen Voraussetzungen, eine einheitliche Möglichkeit offen, ihre elektronischen Dokumente für das WWW aufzubereiten. Abb. 1 zeigt die erforderlichen Arbeitsschritte in schematisierter Form:



Für das Einbringen von Dokumenten auf den Dokumenten-Server und die Erfassung einiger bibliographischer Angaben (Metadaten) werden zwei Eingabeformulare im WWW zur Verfügung gestellt. Technisch wurde dies mit Hilfe eines Perl-Skripts und der cgi-lib.pl Library⁵ realisiert. Ein etwas verkürztes und damit einfacher zu handhabendes Formular dient zur Anmeldung von Prüfungsarbeiten, wie Dissertationen, Diplom- und Studienarbeiten etc. (Abb.2); ein ausführliches Formular steht für alle anderen Dokumentarten, wie Zeitschriftenaufsätze, Proceedingsbeiträge etc., zur Verfügung. Die Struktur der Metadaten basiert auf den Vorschlägen der Dublin Core Initiative⁶ und wurde gemeinsam mit dem Bibliotheksservicezentrum Baden-Württemberg entwickelt.⁷

The screenshot shows a Netscape browser window with the title 'Anmelden von Prüfungsarbeiten für den Online Publikationsverbund der Universität Stuttgart'. The page content includes:

- Header:** 'Anmelden von Prüfungsarbeiten für den Online Publikationsverbund der Universität Stuttgart' with a small icon of a folder and a floppy disk.
- Text:** 'Beschreiben Sie hier anhand der vorgegebenen Kategorien die Prüfungsarbeit (Dissertation, Diplom-, Magisterarbeit o.ä.), die Sie in den Publikationsverbund der Universität Stuttgart einbringen wollen. Die mit * gekennzeichneten Felder sind obligatorisch auszufüllen (Pflichtfelder). Wenn Sie Hilfe zu einem Eingabefeld benötigen, klicken Sie bitte auf den Feldnamen. Vielen Dank für Ihre Mithilfe.'
- Form Fields:**
 - Dokumentart:** A dropdown menu with 'Diplomarbeit, Magisterarbeit' selected. Below it, text says: 'Falls Sie hier nicht die für Ihr Dokument passende Dokumentart gefunden haben, benutzen Sie bitte das [Alternative Anmeldeformular von OPUS](#)'.
 - Titel der Arbeit:** A text input field.
 - Verfasser:** A text input field with the label 'Nachname, Vorname'.
 - Schlagerwörter:** A text input field. Below it, text says: 'Bitte recherchieren Sie ausschließlich im kontrollierten [Schlagwortkatalog von OPUS](#) und übernehmen Sie die zutreffenden Begriffe (max. 50 per "Kopieren und Einfügen" (Copy and Paste). Bitte trennen Sie die einzelnen Begriffe durch Leerzeichen Komma Leerzeichen (Schlagwort1, Schlagwort2)'.

Abb. 2: Anmeldeformular für Prüfungsarbeiten (Ausschnitt)

Die DC-Metadaten werden in einer mSQL-Datenbank⁸ gespeichert. Dieses von David Hughes entwickelte Shareware-Datenbanksystem ist für Hochschulen frei verfügbar. Das WWW-Gateway W3-mSQL ermöglicht es, die gesamte Funktionalität von mSQL zu nutzen, ohne cgi-Skripts programmieren zu müssen. Dazu müssen lediglich Befehle der mSQL-Skriptsprache Lite in HTML-Dokumente eingebettet werden.

⁵ <http://www.bio.cam.ac.uk/cgi-lib/>

⁶ http://purl.oclc.org/metadata/dublin_core/

⁷ http://www.swbv.uni-konstanz.de/wwwroot/s71800_d.html

⁸ <http://www.Hughes.com.au/>

Retrieval

Je nach Anforderungen und Vorwissen eines Anwenders stehen für die Recherche nach elektronischen Dokumenten eine ganze Reihe von sich ergänzenden Instrumenten zur Verfügung. Hier ist zunächst die Recherche in der OPUS-Datenbank zu nennen. Das Suchformular lehnt sich an die Masken der diversen OPACs der Bibliotheken, insbesondere jedoch die des Stuttgarter Online Public Access Catalog (StOPAC)⁹ an, so daß sich dem Suchenden eine vertraute Suchfunktionalität bietet. Über diese gewohnte Funktionalität hinaus werden hier jedoch zwei Suchmodi angeboten. Zum einen die exakte Suche, bei der die Suchbegriffe einschließlich aller Artikel, Präpositionen etc. genau in der Reihenfolge eingegeben werden müssen, in der sie im jeweiligen Suchfeld vorkommen. Zum anderen die Teilwortsuche bei der eine automatische Rechts- und Linkstrunkierung des Suchbegriffs vorgenommen wird. In beiden Modi ist eine zusätzliche Inrentunkierung, d.h. das Ersetzen einzelner Buchstaben durch Wildcards, möglich. Die inhaltliche Suche ist entweder über SWD-Schlagwörter¹⁰ oder einzelne Fachklassifikationen möglich, wobei die Schlagwortnormdatei und derzeit fünf Fachklassifikationen¹¹ dem Verfasser bei der Beschreibung seines Dokuments online zur Verfügung stehen.

Die zweite im Rahmen des Projekts realisierte Suchmöglichkeit ist die flache Volltextsuche über alle enthaltenen Dokumente. Zu diesem Zweck wird beim Einbringen der PDF-Dokumente bereits automatisch der Text im ASCII-Format extrahiert.¹² Aus diesem wird mit Hilfe des von der University of Arizona entwickelten, ebenfalls frei verfügbaren Suchwerkzeugs Glimpse¹³ ein durchsuchbarer Index generiert. Ein Vorteil dieses ansonsten recht unspezifischen Suchinstruments liegt in der Möglichkeit, Tippfehler bei der Formulierung der Suchbegriffe zu berücksichtigen, d.h. nach ähnlich geschriebenen Wörtern zu suchen.

Kooperation mit dem Bibliotheksservicezentrum Baden-Württemberg

Zusätzlich zu diesen im Rahmen des Projekts geschaffenen Möglichkeiten werden die bibliographischen Metadaten automatisch in den Südwestdeutschen Bibliotheksverbund (SWB) eingespielt. Ein derartiger Nachweis in der zentralen bibliographischen Datenbank der Region ergibt sich beinahe zwingend aus der Forderung, elektronische Dokumente trotz oder gerade wegen ihres informationswissenschaftlichen Mehrwerts nicht grundsätzlich anders zu behandeln und nachzuweisen als konventionelle Publikationen. Wäre dies der Fall, würde wieder eine Situation unübersichtlicher Partikularnachweise entstehen, die im Zeitalter der Online-Bibliothekskataloge als historisch zu gelten hat. Aus diesen Gründen müssen, gerade jetzt zu Beginn der Entwicklung neuer Publikationsformen in Bibliotheken, leistungsfähige Verfahren entwickelt werden, die eine vollständige Einbindung elektronischer Dokumente in die bestehende Infrastruktur der Bibliotheken ermöglicht, ohne unverhältnismäßig großen zusätzlichen Personalaufwand zu verursachen.

⁹ <http://www.uni-stuttgart.de/ub/OpacStOPAC/suchmaske.html>

¹⁰ Eine bibliothekarische Prüfung der vom Verfasser vergebenen SWD-Schlagwörter und Überarbeitung nach RSWK erfolgt derzeit nicht.

¹¹ Thesaurus Sozialwissenschaften, NASA-Thesaurus, MSC - Mathematics Subject Classification, PACS - Physics and Astronomy Classification Scheme, CRCS - Computing Reviews Classification System

¹² Hierzu dient die frei verfügbare Xpdf-Komponente pdf2text. <http://www.foolabs.com/xpdf/>

¹³ <http://glimpse.cs.arizona.edu/>

Zu diesem Zweck wurde eine gemeinsam mit dem Bibliotheksservicezentrum Baden-Württemberg (BSZ) festgelegte Dublin Core-Konvention auf das SWB-Format abgebildet. Der DC-Metadatenatz liegt dabei im HTML-Format vor, gemäß folgendem Beispiel:

```
<HTML>
<HEAD>
<META NAME="DC.Title" CONTENT="Dynamische adaptive Lastbalancierung
f&uuml;r gro&szlig;e, heterogen konkurrierende Anwendungen">
<META NAME="DC.Creator.PersonalName" CONTENT="Becker, Wolfgang">
<META NAME="DC.Subject" CONTENT="(SCHEME=SWD) Parallelrechner , Last-
teilung ">
<META NAME="DC.Publisher" CONTENT="Universit&auml;t Stuttgart">
<META NAME="DC.Publisher" CONTENT="Fakult&auml;t Informatik">
<META NAME="DC.Date.Creation_of_intellectual_content" CONTENT="1995">
<META NAME="DC.Date.Creation_of_present_form" CONTENT="03-Nov-1997">
<META NAME="DC.Type" CONTENT="(SCHEME=DC Standard Object Types)
PhDThesis">
<META NAME="DC.Format" CONTENT="(SCHEME=imt) application/pdf">
<META NAME="DC.Identifier" CONTENT="(SCHEME=url)
http://www.uni-stuttgart.de/opus/volltexte/14/5/index.html">
<META NAME="DC.Source" CONTENT="SWB-IDNR/5322216 ">
<META NAME="DC.Source" CONTENT="OPUS-IDN/5">
<META NAME="DC.Language" CONTENT="(SCHEME=ISO639-2) ger">
</HEAD>
<BODY></BODY>
</HTML>
```

Durch einen am BSZ entwickelten Konverter wird folgende Aufnahme generiert und in die Verbunddatenbank eingespielt:

```
200eBecker, Wolfgang
320*Dynamische adaptive Lastbalancierung für große, heterogen konkurrierende
  Anwendungen
334 Computerdatei
359 Wolfgang Becker
410 [S.I.]
412 Universität Stuttgart , Fakultät Informatik
425 1995
433 Computerdatei
501 http://www.uni-stuttgart.de/opus/volltexte/14/5/index.html
504 dt.
519 Stuttgart, Universität, Dissertation, 1995
574 hs
575 cofz
adr S1UB,opus
url http://www.swbv.uni-konstanz.de/depot/dokersch/6400000/6428000/
  6428823k.html
txt Volltext
end
```

Die Schlagwörter werden dabei gegenwärtig noch in den Lokaldatenbereich der Datenbank übernommen. Langfristig ist beim Erfassen der Schlagwörter eine Verknüpfung mit ihrer jeweiligen SWD-Identnummer geplant, die ein Einspielen der Daten in den Titelstammbereich ermöglichen würde.

Im Verbundrahmen können die Titelaufnahmen nach bibliothekarischen Gesichtspunkten (RAK-NBM) überarbeitet werden, um dann in den aus der Verbunddatenbank abgeleiteten regionalen und lokalen Online Katalogen (in diesem Falle dem StOPAC) ebenfalls für die Recherche zur Verfügung zu stehen.

Eine weitere zukunftsorientierte Suchoption ergibt sich aus der Möglichkeit, Indexdateien der Volltexte mit Dublin Core-Metadaten auszustatten und für Suchmaschinen anzubieten, um gleichzeitig elektronische Volltexte und andere wissenschaftliche Netzressourcen¹⁴ strukturiert suchen zu können. Hier bleibt die Entwicklung im Bereich der WWW-Suchmaschinen abzuwarten.¹⁵

Ausgabe

Über die im Rahmen des Projekts geschaffenen Suchmöglichkeiten bzw. die WWW-Recherchemaske der SWB-Verbunddatenbank¹⁶ oder des StOPAC¹⁷ gelangt man zu einer ausführlichen Beschreibung des Dokuments, einschließlich Kurzfassung (Abstract). Sie soll dem Benutzer als Entscheidungshilfe dienen, ob sich das Laden bzw. Speichern des Volltextes auf dem lokalen Rechner lohnt. Von dort verweisen nun ein bzw. mehrere Links auf den Volltext im PDF-Format.

Zum Betrachten, Ausdrucken, Navigieren und Suchen im Einzeldokument benötigt man den frei verfügbaren Acrobat Reader¹⁸ oder Ghostview und Ghostscript.¹⁹ Wenn im Browser das PDF-Plug-In installiert ist, so wird das Dokument innerhalb des Browsers durch den Acrobat Reader angezeigt. Ansonsten muß der Browser so konfiguriert werden, daß der Reader (als Helper-Applikation) automatisch mit dem Dokument als externes Fenster geöffnet wird.

Qualitätssicherung

Die technische Voraussetzung für ein mögliches inhaltliches Qualitätssicherungsverfahren besteht darin, die eingebrachten Dokumente nicht direkt im WWW zugänglich zu machen, sondern sie statt dessen, ebenso wie die sie beschreibenden Metadaten, in einem gesicherten Bereich des Servers zwischenspeichern. (Abb. 3)

¹⁴ Vgl. verschiedene Erschließungsprojekte z.B. Deutscher Bildungs-Server <http://dbs.schule.de/indexe.html>, Social Science Information Gateway (SOSIG) <http://www.sosig.ac.uk/> und Nordic Metadata Project/Nordic Web Index <http://linnea.helsinki.fi/meta/>

¹⁵ Erste Ansätze einer DC-basierten Suchmöglichkeit bei einer allgemeinen Suchmaschine, die derzeit wissenschaftlichen Ansprüchen jedoch noch nicht genügt, zeigt beispielsweise Fireball. http://www.fireball.de/meta_daten.html.

¹⁶ <http://www.swbv.uni-konstanz.de/CGI/cgi-bin/opacform.cgi>

¹⁷ <http://www.uni-stuttgart.de/ub/OpacStOPAC/suchmaske.html>

¹⁸ <http://www.adobe.com/prodindex/acrobat/readstep.html>

¹⁹ <http://www.cs.wisc.edu/~ghost/>

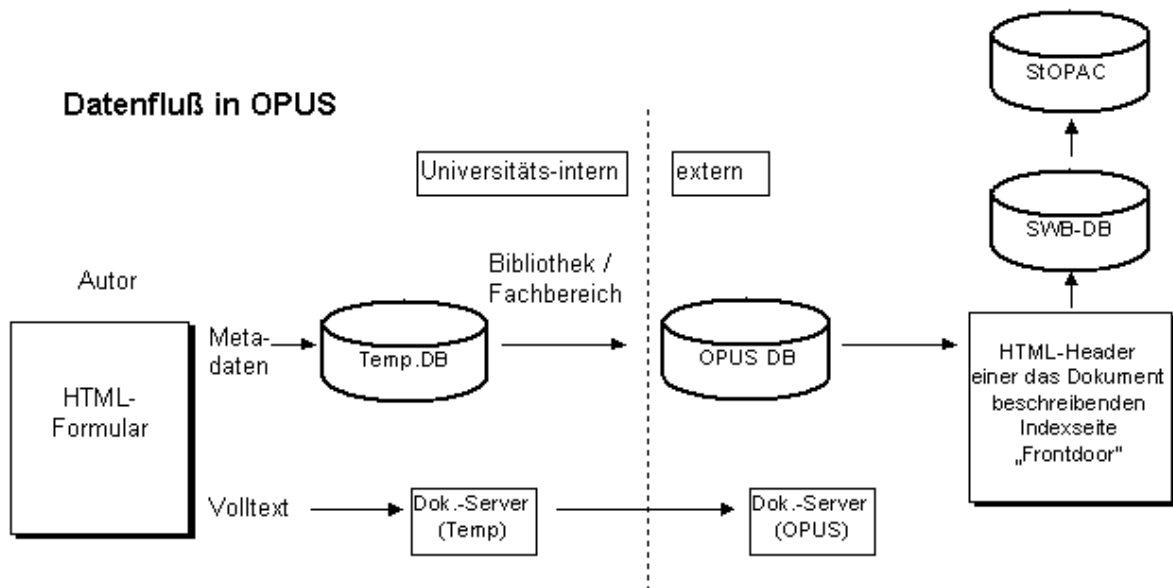


Abb. 3: Datenfluß im Online Publikationsverbund der Universität Stuttgart

Die organisatorische und inhaltliche Durchführung der so möglich gewordenen Qualitätsprüfung wird gegenwärtig noch diskutiert. Zwei prinzipielle Modelle, die jedoch auch Mischformen zulassen, zeichnen sich ab: Die Prüfung elektronischer Dokumente durch Mitarbeiter der Institute bzw. Lehrstühle oder durch die Fachreferenten der Universitätsbibliothek. Alternativ dazu ist die Schaffung einer entsprechenden Organisation zur Sicherung von Qualitätskriterien (Herausgebergremium, Fachbeirat, Lektorat) innerhalb der Hochschule anzudenken. Entsprechende Modelle werden beispielsweise im Rahmen des Konzepts für die Digitale Bibliothek Nordrhein-Westfalen²⁰ erwogen.

Hochschulschriftenverlag

Derartige Überlegungen zur Qualitätssicherung gehen jedoch gleichzeitig von einer stärkeren Betreuung der Autoren bei Vorbereitung und Gestaltung ihrer Dokumente aus und besitzen daher eine nicht zu vernachlässigende verlegerische Dimension. Um diese professionell umsetzen zu können, ist die Zusammenarbeit von Bibliothek und Rechenzentrum mit einem bestehenden Verlag denkbar und möglich. Die im Frühjahr 1997 in Kooperation durch die Gesamthochschule Kassel und den vdf Hochschulverlag Zürich gegründete *Kassel University Press*²¹ ist hierfür ein Beispiel. Lektorielle Bearbeitung von Manuskripten, Gestaltung der äußeren Form (Layout) der Produkte und Marketingkonzepte sind traditionelle Arbeitsfelder eines Verlages, die sich bei einem Hochschulschriftenverlag idealerweise mit den Kompetenzen der Bibliothek als Servicezentrum bei der Erschließung und Aufbereitung von Informationen und ihren Kontakten innerhalb der Universität ergänzen. In diesem Rahmen ließe sich auch das oben bereits angesprochene Herausgebergremium realisieren, das sich aus Hochschullehrern, Verlegern und Bibliothekaren zusammensetzen sollte. Auf diese Weise könnten auch zahlreiche Publikationsunternehmungen einzelner Institute (meist in Form von Dissertations-Schriftenreihen) koordiniert und

²⁰ <http://www.ub.uni-bielefeld.de/digibib-nrw/konzept.htm>

²¹ <http://www.upress.de/>

professionalisiert werden. Dies würde insgesamt wiederum zu einer positiveren und geschlosseneren Außenwirkung der Universität im Sinne einer *Corporate Identity* beitragen.

Die Bibliothek als Kompetenz- und Servicezentrum

Die Universitätsbibliothek Stuttgart sieht die Erfassung und Aufbereitung elektronischer Hochschulschriften als einen Baustein ihres neu zu schärfenden Dienstleistungsprofils. Dabei ist vor allem auf die Integration bereits bestehender Dienste zu achten. Einen besonderen Service leistet die Übersetzungsstelle der Bibliothek, die auf Anforderung für Institute der Universität Übersetzungen wissenschaftlicher Literatur aus dem Russischen anfertigt. Diese Übersetzungen werden, vorbehaltlich der Klärung urheberrechtlicher Fragen, in Zukunft als elektronische Texte im Volltextinformationssystem der Universität zugänglich sein. Zudem wird ein Bestand an wichtigen und nachgefragten bereits erstellten Übersetzungen sukzessive eingebracht, der so einem weitaus größeren Nutzerkreis als bislang zur Verfügung gestellt werden kann.

Die elektronischen Dokumente des Volltextinformationssystems sind als Verbesserung der Informationsdienstleistungen der Bibliothek und als Mehrwert innerhalb ihrer Online-Kataloge zu sehen. Elektronische Zeitschriften und Datenbanken ergänzen dieses Bild ebenso wie konventionelle Informationsmittel. Das sich stetig verbessernde (sich jedoch auch stetig ändernde) Informationsangebot muß den Nutzern und Anwendern vermittelt werden. Dies sieht das traditionelle Selbstverständnis der Bibliotheken ohnehin vor, wobei Anspruch und Wirklichkeit in vielen Fällen nicht deckungsgleich sind. Andererseits muß darüber hinaus jedoch eine (pointiert ausgedrückt) „informationswissenschaftliche Entwicklungshilfe“ geleistet werden, um es den Universitätsangehörigen zu ermöglichen, einfacher und schneller an die benötigten Informationen heranzukommen. Auch diese Forderung ist nicht neu, bislang jedoch vielerorts nur unzureichend verwirklicht worden. Ein dauerhaftes Einbinden von Bibliothekaren bzw. Informationswissenschaftlern in Propädeutika ist dringend vonnöten, um das Ausbildungsangebot der Hochschulen auf einem in der heutigen Informationsgesellschaft besonders wichtigen Feld zu verbessern. Im Zuge des elektronischen Publizierens von Hochschulschriften kommen weitere Inhalte (wie beispielsweise das Erstellen, Strukturieren und Konvertieren von elektronischen Dokumenten) hinzu. An manchen amerikanischen Universitäten sind derartige Kurse und Workshops unter Beteiligung von Bibliothekaren bereits eine Selbstverständlichkeit.²² Verfolgt man diesen Gedanken der Autorenbetreuung konsequent zu Ende, schließt sich hier wieder der Kreis zum bereits erwähnten Hochschulschriftenverlag.

Serviceleistungen sind generell auf die enge Rückkopplung von Anbietern und Nutzern angewiesen. Dies gilt für alle Bereiche bibliothekarischer Informationsvermittlung, insbesondere jedoch im Bereich elektronischen Publizierens, da hier jeder Universitätsangehörige potentieller Nutzer und Anbieter sein kann. In diesem Sinne kann der Publikationsverbund der Universität Stuttgart als Baustein eines Dienstleistungsangebots verstanden werden, gleichzeitig jedoch auch als Katalysator, um dieses Angebot insgesamt zu überdenken und zu verbessern.

²² Vgl. z.B. die Virginia Polytechnic Institute and State University. <http://scholar.lib.vt.edu/theses/theses.html>