

ProPrint: Schnittstelle für ein Qualitätsmanagement digitaler Dokumentenbestände

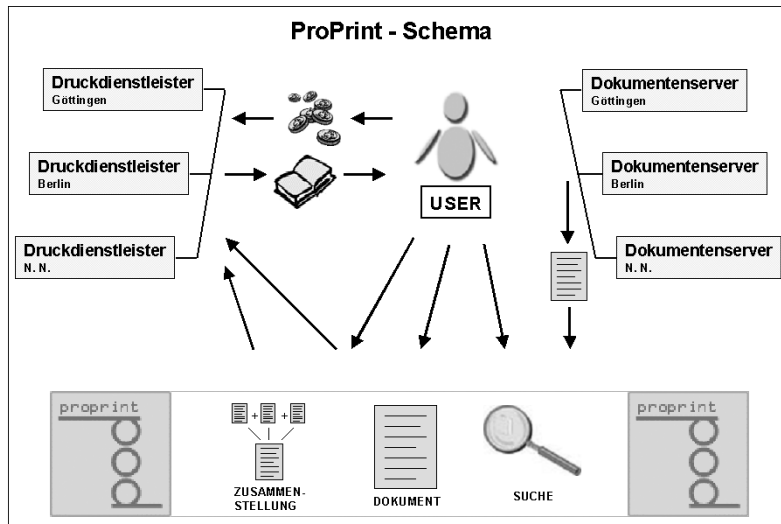
Zusätzlicher Nutzen eines Print-on-Demand-Webservices

Andres Imhof, Matthias Schulz

Seit mehr als fünf Jahren werden in Deutschland an fast jedem Universitätsstandort Dokumentenserver zur Verbreitung und Aufbewahrung digitaler wissenschaftlicher Literatur aufgebaut. Es wurden und werden Promotionsordnungen an Universitäten geändert, um die kostengünstige Abgabe von Dissertationen in Form von digitalen Daten zu ermöglichen, historische Quellen retrodigitalisiert, um diese aus den Tiefen der Magazine zu holen und jedermann zugänglich zu machen. Es werden ferner auf Empfehlung des Wissenschaftsrats Versuche gestartet, wissenschaftliche Zeitschriften statt auf Papier im weltweiten Netz zu etablieren, um Kosten zu sparen. Ziel all dieser Unternehmungen ist der leichte Zugang zur Information von annähernd jedem Ort der Welt und dies zu preisgünstigen Konditionen.

Nun ist statt der bloßen Betrachtung am Bildschirm auch die „althergebrachte“, gedruckte und gebundene Form nach wie vor für die wissenschaftliche Lektüre von den Nutzenden gewünscht. Darum versuchen verschiedene Projekte, die Dokumentenpools auch über „Book-on-Demand“ bzw. „Print-on-Demand“ der interessierten Öffentlichkeit zur Verfügung stellen.

Eins dieser ehrgeizigen Projekte ist ProPrint, ein vom DFN-Verein gefördertes Gemeinschaftsprojekt des Computer- und Medienservice und der Universitätsbibliothek der Humboldt-Universität zu Berlin sowie der Staats- und Universitätsbibliothek Göttingen. Hier wurde ein Webservice entwickelt, der eine einfache Anbindung bestehender Dokumentenserver, der Aufbau einer Informationsdatenbank über lieferbare Titel, die Aufbereitung und Kombination von Dokumenten nach Nutzerwünschen und die Weiterleitung an lokale, zertifizierte Druckdienstleister ermöglicht. Die Nutzer des Print-on-Demand-Webservices holen sich abschließend das gebundene Buch entweder bei sich vor Ort ab und bezahlen bar oder lassen es sich per Post zustellen und begleichen die Rechnung per Überweisung.



Workshop „Print-on-Demand und Dokumentenserver“

Am 17. September 2002 stellte sich das Projekt in Göttingen auf dem Workshop „Print-on-Demand und Dokumentenserver“ vor¹. Zum ersten Mal wurde ProPrint in größerem Zusammenhang mit anderen, thematisch verwandten Unternehmungen diskutiert. Gastvorträge wurden von Herrn Scholze über den Stuttgarter OPUS-Server (<http://elib.uni-stuttgart.de/opus/>), von Herrn Lützenkirchen zu MILESS und MyCoRe aus Essen (<http://miless.uni-essen.de/> und <http://www.mycore.de/>), von Frau Franken (Utrecht) zum Roquade-Projekt (<http://www.roquade.nl>), von Herrn Wätjen (Oldenburg) zu GAP (<http://www.ubka.uni-karlsruhe.de/gap-c>) und von Frau Dr. Reinitzer über Book-on-Demand an der UB Graz (<http://www.kfunigraz.ac.at/ub/>) vorgetragen.

Durch die Gegenüberstellung verschiedener Ansätze, das Abwägen der jeweiligen Vor- und Nachteile und die anschließende Diskussion mit dem fachkundigen Publikum traten essentielle Zusammenhänge, Fragestellungen und Problemfelder zu Tage, von denen die wenigsten bislang im Umgang mit digitalen Dokumenten so deutlich ins Bewusstsein getreten waren².

¹ siehe <http://dochost.rz.hu-berlin.de/proprint/workshop.xml>

² Sie finden eine Zusammenfassung der Diskussion auf der oben genannten Internetseite.

In der Vergangenheit war der Druck von einem Exemplar nicht zu erschwinglichen Preisen möglich, so dass sich dieser Vertriebsweg für wissenschaftliche Literatur nicht anbot. ProPrint gelang es unter Nutzung der technischen Entwicklung der letzten Jahre durch die Vereinfachung, Vereinheitlichung und Automatisierung von Workflows den Kostendruck so zu senken, dass sich Druckdienstleister in der Lage sehen, Einzelausgaben zu marktüblichen Kopierpreisen anbieten zu können. Damit besteht grundsätzlich die Möglichkeit, digitale Materialien, wie sie auf den Servern vieler Hochschulen bereitstehen, in einen derartigen Dienst einzubeziehen.

Die rechtliche Situation erscheint vielleicht nicht immer ausreichend geklärt. Haben die Autoren die Druckrechte abgegeben? Wenn das nicht sicher ist, könnte bestenfalls das Recht auf Kopie eines Exemplars für private oder wissenschaftliche Zwecke die Bereitstellung eines Exemplars erlauben – gerade das aber macht ProPrint möglich.

Auch die technische und inhaltliche Qualität aller Dokumente ist nicht immer auf dem hohen Niveau, das man sich für eine Ausgabe wünscht. In den Anfängen der Entwicklung wurden auch Inhalte auf die lokalen Dokumentenserver gestellt, die nicht unbedingt wissenschaftlichen Maßstäben genügten oder zuvor von entsprechenden Stellen geprüft worden sind. Außerdem sind in den vergangenen fünf Jahren unterschiedliche Archivierungsformate und -versionen verwendet worden. Differierende Formate bringen aber immer auch Kompatibilitätsschwierigkeiten mit sich. Daraus folgt, dass eigentlich jedes „historische“ digitale Dokument noch einmal geprüft und gegebenenfalls aktuellen technischen Standards angepasst werden müsste. Bei den retrodigitalisierten Dokumenten, insbesondere aus den DFG-geförderten Projekten, sieht das sicher anders aus. Bei diesen Digitalisierungsprojekten historischer Quellen wurde eine strenge Vorprüfung der wissenschaftlichen Relevanz vorgenommen.

Die Schwächen bei den älteren digitalen Dokumentenbeständen werden bei der Anbindung eines Print-on-Demand-Service offensichtlich. Der Einsatz eines Webservices wie ProPrint verlangt deshalb den Dokumentenservern ein gehöriges Maß an Standardisierungen ab, um weiterführende Dienstleistungen anbieten zu können. Der Zwang zur Standardisierung, der von ProPrint ausgeht – und das wurde auch in der Diskussion hervorgehoben – ist jedoch schon für sich allein von Wert, weil durch die konkrete Nutzung der Bestände die kontinuierliche Qualitätsverbesserung gefördert wird. Der gemeinsame Einsatz einheitlicher Standards aber ermöglicht auf die Dauer einen reibungslosen Austausch von Dokumenten zwischen Dokumentenservern, ProPrint und Druckdienstleistern.

Standardisierte Kommunikation

Standards sind erforderlich, wenn es um Kommunikation und insbesondere um den konkreten Austausch von Daten geht. Für die Kommunikation wurde mit ProPrint nicht der Weg eines Dokumentenharvesting gegangen, sondern das Modell eines Services-Providers realisiert, der Metadaten sammelt und zentral für die Recherche vorhält. Hierbei wurden die Ergebnisse und Entwicklungen bei OAI angewendet. Das Open Archives Protocol for Metadata Harvesting (OAI-PMH 2.0) bildet die Grundlage, die digitalen Archive verteilter Standorte miteinander zu verbinden, wobei der Zugriff auf die Metadaten durch eine erweiterte OAI-Schnittstelle erfolgt, die es ermöglicht, den Dublin Core und den ergänzenden ProPrint-Metadatenatz abzurufen. Für eine Anbindung benötigen die Dokumentenserver vor allem eine OAI-Schnittstelle. Dann ist ProPrint in der Lage, heterogene Systeme zusammenzuführen und in einer homogenen Umgebung zu steuern.

Für Katalogsysteme die keine OAI-Schnittstelle einrichten können, hält ProPrint zusätzlich als Alternative eine *Allegro*-Datenbank bereit. Diese speziell für bibliothekarische Anforderungen entwickelte Datenbank wurde wegen der vielfältigen Konvertierungsmöglichkeiten von Metadaten systemen in ProPrint implementiert. So können die Datensätze über Allegro in das OAI-Proprint-Metadaten schema konvertiert werden.

Ein Austausch zwischen den Dokumentenservern kann aber nur funktionieren, wenn auch Übereinkünfte zu den Formaten und der Qualität der Dokumente bestehen. ProPrint setzt das PDF-Format von Adobe als Austauschformat ein. Aber PDF ist nicht gleich PDF. In der Erstellung dieses Formates können Fehler unterlaufen, die die weitere Verwendung des Dokumentes ausschließen lassen³, außerdem ergeben sich durch Versionswechsel Kompatibilitätsprobleme.

Welche Fehler auftreten können und welche Aspekte unbedingt für einen standardisierten Austausch erfüllt sein müssen, wurden im Rahmen des ProPrint-Projektes geprüft. Mit Unterstützung der DINI Arbeitsgruppe Elektronisches Publizieren wurden auf Anregungen aus dem Workshop umfangreiche PDF-Testreihen durchgeführt, in denen nach Fehlerquellen und deren technischer Lösung gesucht wurde. Hieran haben sich dankenswerter Weise OPUS Stuttgart und das Universitätsrechenzentrum der TU Chemnitz beteiligt.

3 Für den richtigen Umgang mit dem PDF-Format sei folgende Literatur sehr empfohlen: Merz, Thomas; Drümmer, Olaf: Die PostScript- & PDF-Bibel. Heidelberg 2002, 2. Aufl.

Betrachtet wurden

- die Einbettung von Schriften,
- die PDF-Versionen,
- die jeweiligen PDF-Erstellungstools und
- etwaige Sicherheitseinstellungen.

Zum Abschluss wurde eine Auswahl an fraglichen Dokumenten zum konkreten Druck an einen Druckdienstleister gegeben und auf fehlerhafte Darstellungen überprüft.

Es konnte festgestellt werden, dass der überwiegende Teil der Dokumente korrekt war und problemlos auf allen Systemen geöffnet und gedruckt werden konnte. Kleinere Fehlerquellen in der Einbettung von Schriften verursachten in der Regel keine Schwierigkeiten in der in gewissem Rahmen fehlertoleranten Verarbeitung. Doch wurden auch Dokumente entdeckt, bei denen beispielsweise Nutzungsbeschränkungen mit einem Passwort geschützt waren, was deren weitere Verwendung ausschließt. Auch formale Fehler im Dokument, wie z. B. Reste vom Korrekturmodus in Word oder fehlende Bilder, machten in einzelnen Fällen Probleme.

Ergebnis dieses Testes ist die Implementierung eines Qualitätsmanagement-Workflows in das ProPrint-System. Durch diesen ist es möglich, dass bei einer fehlerhaften Generierung von PDF durch PDFLib automatisch eine Fehlermeldung an den Nutzer und den ProPrint-Administrator gesendet werden. Gleichzeitig setzt das ProPrint-System das Dokument inaktiv.

Der Nutzer wird in einem solchen Fall informiert, dass das Dokument (noch) nicht geliefert werden kann. Der Administrator erhält die generierte Fehlermeldung, um das PDF-Dokument zu prüfen und um sich an den Betreiber des entsprechenden Dokumentenservers zu wenden. Ist das Dokument erneut generiert und geprüft, kann der ProPrint-Administrator das Dokument wieder aktiv setzen und gegebenenfalls den Nutzer darüber informieren. Erst danach kann das Dokument wieder durch eine Sucheingabe in der ProPrint-Datenbank gefunden werden.

Dieser Workflow evaluiert so Dokumentenbestände, deren Qualität in technischer Hinsicht unbekannt ist. Jedes fehlerhafte Dokument wird nur einmal aufgerufen, danach ist es vorerst nicht mehr aktiv. Unter diesem Gesichtspunkt kann ProPrint auch als eine Art Kontrollinstitution für das Qualitätsmanagement von Dokumentenservern wirken.

Checkliste für digitale Dokumente

Für die Anbindung eines Print-on-Demand Services ergeben sich als Ergebnis aus den Testreihen folgende Empfehlungen:

- *Einbetten von Schriften und Vorkontrolle der Dokumente:* Wenn PDF als Archivierungsformat verwendet werden soll, sollte es zwingend vorgeschrieben sein, dass grundsätzlich alle verwendeten Schriften während der Umwandlung in das PDF-Format über den Distiller oder anderen Software-Tools eingebettet werden. Dies erhöht zwar die Dateigröße und ist wegen der längeren Ladezeit weniger für die Webbetrachtung geeignet, aber zur Archivierung von wissenschaftlichen Publikationen sollte auf jeden Fall gewährleistet werden, dass die PDF-Dateien auf jeder Plattform gelesen werden können. Ein zwingend notwendiger Schritt vor dem Einstellen der Dokumente auf dem Server ist eine vorgeschaltete gründliche Kontrolle durch den Betreiber des Dokumentenservers oder eine andere Institution.
- *Keine Sicherheitseinstellungen bei PDF-Dokumenten:* Um die Archivierung von Dokumenten im PDF-Format zu ermöglichen, sollten zukünftig Sicherheitseinstellungen und Benutzungsbeschränkung grundsätzlich ausgeschlossen werden. Nutzungseinschränkungen können in jedem Fall umgangen werden, indem die Datei über Ghostscript in eine Postscript-Datei gewandelt und diese anschließend durch den Acrobat Distiller oder einem anderen Tool zu einem neuen PDF-Dokument generiert wird. Die Nutzungsbeschränkungen sind deshalb nutzlos und behindern den sinnvollen Einsatz zusätzlicher elektronischer Dienstleistungen. Die Sicherheitseinstellungen sind grundsätzlich auszuschließen.
- *Kontrolle der Farbigkeit von Texten:* Es sollte aus der Sicht der zukünftigen Leser bei der Abgabe der Dokumente darauf geachtet werden, dass keine Schrift unnötiger Weise in Farbe erscheint. Automatische Funktionen in MS Word formatieren z. B. eine URL blau. Diese Funktionalitäten sind auszuschalten. Die Autoren sind darauf aufmerksam zu machen und sollten aufgefordert werden, ein neues PDF-Dokument zu liefern, denn jede dieser Seiten werden in einem Print-on-Demand-Dienst als Farbseite erkannt und als solche bei dem Druckdienstleister beauftragt. Dies erhöht den Preis der Seite um ein Vielfaches, ohne dass ein ersichtlicher Nutzen entstanden wäre.
- *Format bei Print-on-Demand:* Den meisten Dokumenten ist gemein, dass die Schriftgröße jeweils zu gering ist, so dass ein Ausdruck auf dem Format DIN A5 nicht sinnvoll erscheint. Der Text wäre nicht mehr angenehm lesbar. Darum sollte man beim Ausdruck grundsätzlich das Format DIN A4 vorsehen. Sollte der Ausdruck in DIN A5 gewünscht sein, müsste zukünftig die Schriftgröße beim Seitenlayout entsprechend berücksichtigt werden.

Zusammenfassung

Die Entwicklung digitaler Dokumentenarchive steht noch in den Anfängen. Wenn eine solide Basis für spätere Webservices bzw. zusätzliche Dienstleistungen oder Anforderungen – man denke z. B. auch an den barrierefreien Zugang zur Information für Sehbehinderte – aufgebaut werden soll, muss ein Qualitätsmanagement eingesetzt werden, das in der Lage ist, jede Schnittstelle nach außen auf internationale Standards abzustimmen. Erst danach sind Archive präpariert, auf neue Entwicklungen schnell reagieren zu können.

Ein Webservice wie ProPrint kann hierzu Hilfestellung leisten und dies ergänzend zu seiner ohnehin wertschöpfenden Dienstleistung. Er verschafft nicht nur einen umfangreichen Vertriebsraum für verteilte Dokumentenserver, er prüft auch die Beschaffenheit der Dokumentenbestände und fördert die Einführung von internationalen Standards bei allen beteiligten Partnern. Damit ist eine dauerhafte Grundlage für ein Qualitätsmanagement geboten.

Mit dem Angebot einer qualitativ hochwertigen und doch preiswerten Druckausgabe wird vor allem die Attraktivität für DoktorandInnen und WissenschaftlerInnen erhöht, ihre Publikationen auf elektronischen Weg medienneutral zu publizieren.

