

ProPrint

Der Print-on-Demand-Service für Dokumenten- und Publikationsserver

Matthias Schulz

matthias.schulz.1@cms.hu-berlin.de

Einleitung

Seit mehreren Jahren werden in Deutschland an fast jedem Universitätsstandort Dokumentenserver zur Verbreitung und Aufbewahrung digitaler wissenschaftlicher Literatur aufgebaut. Es wurden und werden Promotionsordnungen an Universitäten verändert, um die Veröffentlichung von Dissertationen in digitaler Form zu ermöglichen. Ebenso werden historische Quellen digitalisiert, um diese über das Internet zugänglich zu machen. Daneben werden elektronische Zeitschriften aufgebaut, die langfristig den Weg aus der so genannten Zeitschriftenkrise weisen sollen. Ziel all dieser Projekte ist es, Informationen schnell, weltweit und kostengünstig verfügbar zu machen. Damit entstand eine neue Kultur, die des elektronischen Publizierens.

Diese Kultur trennt sich von traditionellen Prozessen der Produktion und Vermittlung von Publikationen. Hauptmerkmal dieser Kultur ist, dass die Erstellung ausschließlich mit Mitteln der elektronischen Datenverarbeitung, mit Computern, realisiert wird. Damit vollzieht sich ein Medienwandel. So ist der Computer mit einer Textverarbeitung nicht an die Abmessung einer DIN A4-Papierseite gebunden, Text-, Bild und Videoverknüpfungen können automatisiert werden. Der Autor kann sein Dokument ausdrucken, es als E-Mail verschicken oder über das World Wide Web veröffentlichen.

Die elektronische Publikation ist nicht mehr an einen Speicherträger gebunden. Aber sie schließt einen Ausdruck auf Papier nicht aus. Anders ist es beim

Buch, es ist eine Endform der Informationsspeicherung. Sollen diese Informationen digital zur Verfügung gestellt werden, muss das Buch aufwendig eingescannt und jeder Buchstabe in Bits und Bytes umgewandelt werden. Aber elektronische Publikationen werden gedruckte Informationen nicht entbehrlich machen – im Gegenteil: Der Wunsch nach dem gedruckten Dokument wächst bei den Nutzern solcher Informationssysteme. Dieses Bedürfnis bezieht sich in vielen Fällen jedoch nicht auf das gesamte Dokument, sondern lediglich auf Ausschnitte, einzelne Aufsätze, Kapitel, Zitate o. ä. Ziel von ProPrint ist es, einen Print-on-Demand-Service für Hochschulschriften in ganz Deutschland anzubieten.

Mit dem Ziel vor den Augen, Wissenschaftler effektiv mit Informationen zu versorgen, entwickelten der Computer- und Medienservice der Humboldt-Universität zu Berlin und die Staats- und Universitätsbibliothek Göttingen (SUB) von 2000-2003 im Rahmen des DFN-

Der Print-on-Demand-Service ProPrint ist ein Gemeinschaftsprojekt der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB) und des Computer- und Medienservices der Humboldt-Universität zu Berlin. Der Service bietet die Möglichkeit, wissenschaftliche Aufsätze, elektronische Dissertationen oder Beiträge aus digitalisierten Büchern online zu recherchieren, individuell zusammenzustellen und als Paperback oder Skript von Druckdienstleistern produzieren zu lassen. ProPrint bietet bereits einen Zugriff auf über viertausend Monographien und Hochschulschriften an. Dabei wird das Netzwerk von Dokumentenservern und die Einbindung lokaler Druckunternehmen mit Print-on-Demand-Ausstattung kontinuierlich erweitert. Ziel ist es, die individuelle Publikation mit dem ProPrint-Service in jeder wichtigen deutschen Universitätsstadt verfügbar zu machen.

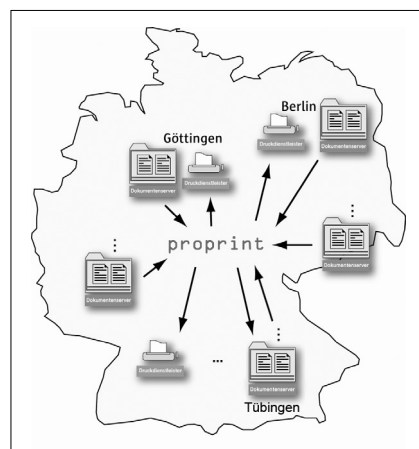


Abb. 1: Schema des ProPrint-Service.

Programms »Einsatz von Netzdiensten im Wissenschaftlichen Informationswesen«, gefördert mit Mitteln des BMBF, den Dienst ProPrint.

Was ist ProPrint?

- ProPrint bietet eine Print-on-Demand-Komponente für jeden Dokumenten- und Publikationsserver. Für den Nutzer des Dokumenten- und Publikationsservers (edoc-Server) erweitern sich die Möglichkeiten des Zugriffs auf die Publikationen. Sie sind nicht nur elektronisch, sondern auch in Papierform verfügbar. Wertvolle digitalisierte Publikationen, die am Bildschirm sehr schlecht zu lesen sind, können als gedruckte Version mit ProPrint bestellt werden. Elektronische Zeitschriften und Tagungsbände können mit ProPrint auch im Print-on-Demand-Verfahren gedruckt und vertrieben werden.
- ProPrint bietet Unterstützung für den Workflow in Bibliotheken an. Mit ProPrint kann die technische Qualität der elektronischen Dokumente automatisch auf Langzeitarchivierbarkeit geprüft werden. Für den Betreiber von Dokumenten- und Publikationsservern ist die Frage nach der Qualität der elektronischen Publikationen in mehrfacher Hinsicht bedeutend. So der Langzeitverfügbarkeit, der Druckbarkeit und deren Qualität. Durch die automatische technische Kontrolle der PDF-Dateien bietet Pro-Print eine Unterstützung für den Workflow in der Bibliothek (siehe Abschnitt Standardisierung).
- ProPrint kann Dokumenten- und Publikationsserver virtuell zusammenschließen. Damit braucht der Suchende nur eine einzige Suchmaske zu benutzen. ProPrint ist ein Service-Provider im Sinne von OAI. Mit der Anwendung von OAI-PMH können Dokumenten- und Publikationsserver zusammengeschlossen werden. Die Besonderheit an ProPrint besteht in seiner Möglichkeit, die gefundenen Publikationen in gedruckter Form zu bestellen. Die Logistik des Versands und der Rechnungsstellung wird dabei von ProPrint geregelt.

Die softwaretechnische Grundlage für ProPrint wird durch ein LAMP-System realisiert. Dieses System setzt sich aus folgenden Komponenten zusammen: Linux als Betriebssystem, dem Apache-Webserver, einer MySQL-Datenbank und PHP als Programmiersprache. Mit dieser Entscheidung wurde besonders auf die Nutzbarkeit des Systems geachtet, da alle verwendeten Softwarekomponenten frei zur Verfügung stehen. Darüber hinaus werden konsequent Standards sowohl für die Kommunikation zwischen den Dokumentenservern als auch für Dokumentformate und Metadaten verwendet.

Um Nutzern Recherchemöglichkeiten über Dokumentenservergrenzen hinaus anzubieten, können verschiedene Techniken zum Einsatz kommen. Genannt seien die verteilte Suche und das Harvesting. Beim ProPrint-System wird das Harvesting von Metadaten mit dem *Open Archives Protocol for Metadata Harvesting* (OAI-PMH) verwendet.

Das OAI-Protokoll ermöglicht einen effizienten Austausch von Metadaten und impliziert eine funktionale Aufteilung in Anbieter von (Dokumenten und) Metadaten, so genannte *Data Provider*¹, und darauf aufbauende Dienste (*Service Provider*²). Die zentrale ProPrint-Suchmaschine ist im Sinne von OAI der Service Provider und die einzelnen Dokumentenserver sind Data Provider.

Voraussetzung für die OAI-Kompatibilität eines Dokumentenservers ist dessen Fähigkeit, Dublin Core als Metadatenformat auszuliefern. Für die Beschreibung weiterer Metadaten, die für den ProPrint-Dienst erforderlich sind, wurden innerhalb dieses Metadatenformates weitere Elemente mit einem gesonderten Namensraum definiert. Dieser Namensraum enthält auch Elemente des DIEPER-Metadatenatzes³. Die ProPrint-Erweiterung des Metadatenformates, der ProPrint-Metadatenatz, umfasst Informationen für besondere Seitenformate, Vertriebsinformationen und Gliederungen eines Dokumentes (Kapitel, Unterkapitel). Die Entwicklung der Elemente für die Beschreibung von Teilstrukturen konnte nicht von physisch vorliegenden Papierseiten einer gedruckten Publikation ausgehen, wie dies das Metadatenformat des

Göttinger Digitalisierungszentrums (GDZ) für Digitalisate vorsah. Da bei den digitalen Dokumenten der Dokumentenserver der Humboldt-Universität und der SUB-Göttingen keine physischen Papierseiten vorlagen, wurde das Metadatenformat für Teilstrukturen auf logischen Einheiten (Kapiteln, Unterkapiteln ...) aufgebaut.⁴

Über die zentrale ProPrint-Suchmaschine haben die Wissenschaftler und Studierenden so einen unbegrenzten Zugang zu elektronischen Dokumenten.

Der ProPrint-Dienst kann aber nicht nur über den zentralen Zugang genutzt werden. Jeder angeschlossene Dokumentenserver kann einen so genannten *Pro-Print-Button* auf seinen Internetseiten anbringen. Damit erweitert sich das Dienstleistungsangebot der Bibliothek um einen Print-on-Demand-Dienst.

¹ Als Data Provider wird eine OAI-kompatible Schnittstelle zu einer Datenbank verstanden, in der sich Metadaten über Dokumente oder andere digitale Objekte befinden und die über eine HTTP-Verbindung erreichbar ist. Sie muss in der Lage sein, OAI-Anfragen entsprechend der Protokolldefinition korrekt zu beantworten.

² Der Service Provider bietet mithilfe von Daten, die er unter Nutzung des OAI-PMH gesammelt hat, einen (innerhalb der Protokollspezifikation nicht näher definierten) Dienst an. Der aus Sicht des Protokolls relevante Teil des Service Providers, der so genannte Harvester, versendet OAI-konforme Anfragen an Data Provider und wertet die entsprechenden Antworten aus.

³ Im europäischen DIEPER-Projekt wurde das Ziel verfolgt, einen zentralen Zugang zu verteilt gespeicherten retrodigitalisierten Zeitschriften in Europa (und weltweit) zu schaffen. Gleichzeitig wurden Fragen des zentralen Zugangs- und der Lizenzbedingungen diskutiert und untersucht. Der Benutzer erhält mit dieser virtuellen Zeitschriftenbibliothek den Nachweis von Volltexten digitalisierter Zeitschriften. Mithilfe einer DIEPER-Suchmaschine kann man nach Zeitschriftentiteln, Autoren, Aufsatztiteln und Artikelinhalten suchen. Zur inhaltlichen und bibliographischen Beschreibung der elektronischen Dokumente wurde in enger Anlehnung an Dublin Core ein eigenes Metadatenmodell entwickelt, dessen Darstellungsform auf XML basiert. Damit bot dieses Datenformat ideale Voraussetzungen zur Definition des gemeinsamen Metadatenformates.

⁴ Details zum Metadatenatz siehe unter <http://edoc.hu-berlin.de/proprint/>

ProPrint aus Sicht der Benutzer



Der Benutzer des edoc-Servers hat sich schon manchmal gefragt, was der kleine grüne

Button in Form eines Buches hinter jedem Eintrag einer Publikation bedeutet. Mit Hilfe dieses Buttons können u. a. Dissertationen, Zeitschriftenartikel und Tagungsbeiträge im Paperbackformat bestellt werden.

Die Bestellung verläuft in mehreren Schritten und dabei können folgende Optionen ausgewählt werden:

- Umwandlung farbiger Bereiche in Graustufen,
- Formatauswahl (DIN A4 / DIN A5),
- Einbandauswahl (mehrere Publikationen in einem Einband oder separate Einbände).

Der Bestellvorgang beginnt damit, ob die Seiten in Graustufen oder farbig gedruckt werden sollen. Die Auswahl der Graustufen beeinflusst sofort den Preis für die Bestellung. Sie kann in einigen Fällen bis über 50% den Preis reduzieren. Anschließend ist das Format auszuwählen. Dabei kann wiederum der Endpreis bei Wahl DIN A5 verringert werden. Bei der Bestellung von mehr als zwei Publikationen hat der Nutzer die Möglichkeit, diese in einem einzigen Einband zu binden oder separat. Das macht vor allem Sinn, wenn ein individueller Tagungsband oder eine thematische Sammlung als Studienmaterial für ein Seminar zusammengestellt werden soll. Eine separate Bestellung der Publikationen ist auch möglich. Mit der Angabe der Lieferadresse und der Auswahl der Bezahlung, Kreditkartenzahlung ist möglich, wird der Bestellvorgang abgeschlossen. Binnen fünf Tagen wird entweder die Bestellung per Post ausgeliefert oder man kann sie direkt beim Druckdienstleister abholen.

Standardisierung der Anforderungen an das Format der Dokumente

Der Dokumentenaustausch zwischen Dokumenten- und Publikationsservern kann nur gewährleistet werden, wenn Übereinkünfte zu Formaten und deren

technischer Qualität bestehen. ProPrint setzt das PDF-Format von Adobe als Austauschformat ein. Jedoch kann die technische Qualität dieses Formates unterschiedlich sein. Bei der Erstellung dieses Formates können Fehler unterlaufen, die die weitere Verwendung des Dokumentes ausschließen. Außerdem ergeben sich bei Versionswechsel Kompatibilitätsprobleme.

Welche Fehler auftreten können und welche Aspekte unbedingt für einen standardisierten Austausch erfüllt sein müssen, wurde in einer umfangreichen PDF-Testreihe ermittelt. An diesem Test beteiligten sich auch die Dokumentenserver von OPUS-Stuttgart und das Universitätsrechenzentrum der TU Chemnitz.

Betrachtet wurden

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

Zum Abschluss wurde eine Auswahl von Dokumenten zum konkreten Druck an einen Druckdienstleister gegeben und auf fehlerhafte Darstellungen überprüft. Die durch den Test ermittelten Fehler führten zu folgenden ProPrint-Empfehlungen:

- **Einbetten von Schriften und Vorkontrolle der Dokumente:** Wenn PDF als Archivierungsformat verwendet werden soll, sind grundsätzlich alle verwendeten Schriften während der Umwandlung in das PDF-Format über den Distiller oder andere Software-Tools einzubetten. Dies erhöht zwar die Dateigröße und ist wegen der längeren Ladezeit weniger für die Webbetachtung geeignet, aber zur Archivierung von wissenschaftlichen Publikationen ist zu gewährleisten, dass die PDF-Dateien auf jeder Plattform gelesen werden können, zumal ohne eingebettete Schriften das Dokument keine feste optische Erscheinung hat und somit nicht plattformunabhängig nutzbar ist.
- **Keine Sicherheitseinstellungen bei PDF-Dokumenten:** Um die Archivierung von Dokumenten im PDF-Format zu ermöglichen, sollten Sicherheitsein-

stellungen und Benutzungsbeschränkung grundsätzlich ausgeschlossen werden. Nutzungseinschränkungen können in jedem Fall umgangen werden, indem die PDF-Datei über Ghostscript in eine Postscript-Datei gewandelt und daraus anschließend durch den Acrobat Distiller oder ein anderes Tool ein neues PDF-Dokument generiert wird (»redestillieren«). Die Nutzungsbeschränkungen sind deshalb weitestgehend nutzlos und behindern den sinnvollen Einsatz zusätzlicher elektronischer Dienstleistungen. Die Sicherheitseinstellungen sind für Dokumente, die durch ProPrint nutzbar sein sollen, 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 unnötigerweise in Farbe erscheint, beispielsweise die blaue Einfärbung von Hyperlinks, die u. a. in MS Word automatisch vorgenommen wird. 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 würde als Farbseite erkannt und als solche bei dem Druckdienstleister beauftragt. Dies würde den Preis der Seite um ein Vielfaches erhöhen, ohne dass ein ersichtlicher Nutzen entstehen würde.
- **Format bei Print-on-Demand:** Den meisten Dokumenten ist gemein, dass die Schriftgröße jeweils so gering ist, dass ein Ausdruck im DIN A5-Format nicht sinnvoll erscheint. Der Text wäre nicht mehr angenehm leserlich. Darum sollte für den Ausdruck grundsätzlich das DIN A4-Format vorgesehen werden. Sollte der Ausdruck in DIN A5 gewünscht sein, ist die Schriftgröße beim Seitenlayout entsprechend zu berücksichtigen.

Wie können Autoren ihre elektronische Publikation auf Print-on-Demand-Tauglichkeit prüfen? Dazu stellt der edoc-Server jedem Autor ein Prüfprogramm in Form eines Preflight-Profiles für Adobe Acrobat zur Verfügung. Dieses Profil kann in jedem Adobe Acrobat importiert

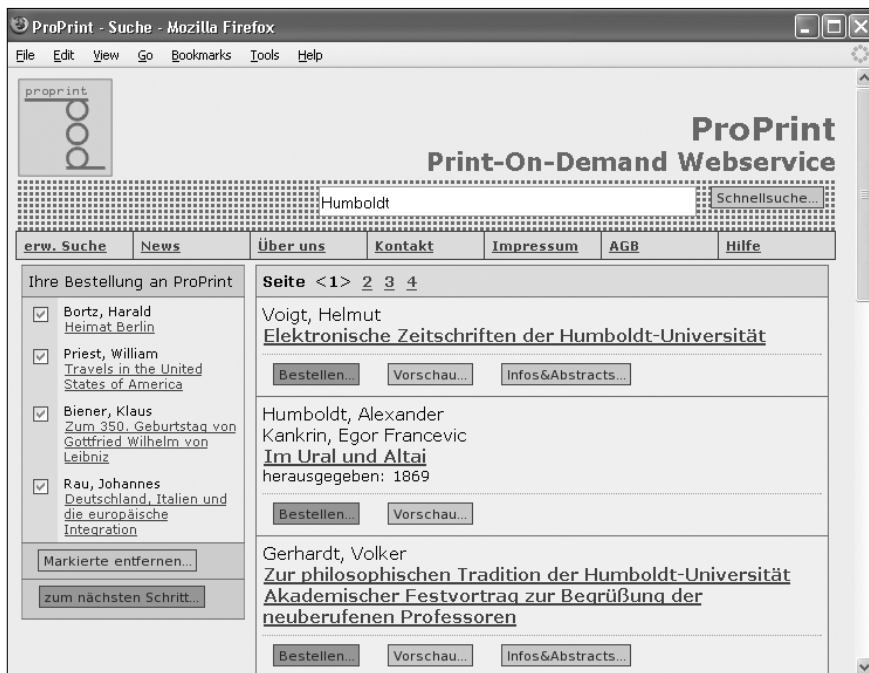


Abb. 2: Suchmaske des ProPrint-Webservice.

werden und die elektronische Publikation kann automatisch auf die oben genannten Schwerpunkte geprüft werden.

Die Entwicklung und Einrichtung des ProPrint-Service

Der ProPrint-Webservice wurde in einfach zu erfassenden Webseiten gestaltet, die komplett dynamisch erstellt werden. Bei der Implementierung wurden folgende Vorgaben umgesetzt:

- **Metadaten:** Bildung eines zentralen, für alle Dokumentenserver einheitlichen Metadatensatzes, der darüber hinaus für viele noch in Zukunft denkbare Dokumententypen und Objekte vorbereitet ist, wie beispielsweise geographische Karten. Dafür wurden in Anlehnung an Dublin Core (DC) ein ergänzender ProPrint-Namespaces und ein ProPrint-Application-Profile gebildet. Die Metadaten müssen dazu vom ProPrint-Webservice regelmäßig und automatisch mit den Dokumentenservern abgeglichen werden.
- **Dokumente:** Dieser Bereich kann in zwei Gruppen eingeteilt werden: die Gruppe der Dokumente, die von den Betreibern der Dokumentenserver verwaltet werden und die Gruppe von Dokumenten, die auf dem ProPrint-

Server für die Druckdienstleister als druckreife Vorlagen vorbereitet werden. Das ProPrint-System muss hierbei (mittels sog. »mergen«, also einem Verbinden der zugrunde liegenden Originaldateien) letztgenannte Dokumente aus den erstgenannten erzeugen können.

- **Rechteverwaltung:** Anwender, Administratoren und Druckdienstleister werden nach ihrem Nutzungsprofil unterschieden und haben unterschiedliche Zugriffsrechte auf Bereiche und Inhalte des ProPrint-Dienstes. Diese Rechte werden durch Nutzer mit entsprechenden Vergaberechten zugewiesen (Rechte-Administratoren).
- **Druck und Vertrieb:** Die dezentralen Druckdienstleister vertreiben die Dokumente in gedruckter und gebundener Form an den ProPrint-Nutzer. Ergänzt wird der direkte Weg über den Versand bei gleichzeitiger Rechnungsstellung.
- **Rechnungsverfahren:** Alle für die Rechnungsstellung relevanten Daten werden im ProPrint-System zentral erfasst und gespeichert. Die Dokumente werden auf ihre Seitenzahl und ihre Farbigkeit ausgelesen. Dementsprechend wird der Preis pro Seite mittels eines Preisschemas für s/w- und farbige Seiten ermittelt. Das ProPrint-System generiert eine Rechnungsvorlage, die druckbar ist und bereits alle Preise enthält.

Die Administration des ProPrint-Webservice erfolgt auf dem Server und über eine Webschnittstelle. Durch diese Schnittstelle sind folgende Funktionen administrierbar:

- Test/Sprache (u. a. die Hilfetexte),
- Nutzerverwaltung,
- OAI-Server (Einbindung von OAI-Servern in den ProPrint-Webservice),
- Zeitplan/Tasks,
- Druckdienste (Einbindung von Druckdienstleistern),
- Rechnungsnummern (Verwaltung der Rechnungsnummer der einzelnen Transaktionen).

Was bietet ProPrint in der Zukunft?

Zurzeit sind 4000 verschiedene Dokumente aus dem Bestand der Dokumentenserver der Humboldt-Universität zu Berlin, der Staats- und Universitätsbibliothek Göttingen und der Universitätsbibliothek Tübingen mit ProPrint verfügbar. Darunter befinden sich Monographien, Zeitschriften, digitalisierte Dokumente aus historischen Beständen, Habilitationsschriften, Dissertationen und Konferenzbeiträge von drei nationalen und internationalen Konferenzen. ProPrint steht für ein offenes System. Das heißt, dass jeder Betreiber eines Dokumentenservers diesen Dienst nutzen kann. ProPrint beschränkt sich dabei nicht nur auf öffentliche Einrichtungen, sondern richtet sich auch an kommerzielle Content Provider.

Literatur

- [1] MÜLLER, U.: OAI-Protokoll: Metadaten für alle. *cms-journal* 27, August 2005, S. 55–58.