

OPUS - Elektronisches Publizieren an Hochschulen

Frank Scholze, Universitätsbibliothek Stuttgart, scholze@ub.uni-stuttgart.de

Zusammenfassung

Das elektronische Volltextinformationssystem OPUS, das aus einem DFN-Projekt hervorgegangen ist, hat sich zu einem stabilen Produkt im Hochschulbereich entwickelt. Aufgrund der Verbreitung wurden eine Reihe von aktuellen Entwicklungen im Bereich des elektronischen Publizierens von Hochschulschriften in einer technisch vollkommen überarbeiteten Version des Systems funktional aufgegriffen. Neben dem Protokoll der Open Archives Initiative für einen internationalen Datenaustausch wurde die Möglichkeit implementiert, Uniform Resource Names zu vergeben und zu registrieren, um eine langfristige stabile Adressierung von Online-Publikationen zu gewährleisten.

In Forschung und Lehre wird heute eine Vielzahl von Dokumenten mittels elektronischer Werkzeuge in druckreifer Qualität erstellt. Dazu zählen unter anderem wissenschaftliche Prüfungsarbeiten (Diplom-, Magister-, Doktorarbeiten, Habilitationen), Forschungsberichte, Konferenzbeiträge, Vorlesungsmanuskripte und Lehrunterlagen, d.h. Dokumentarten, die aufgrund ihrer schlechten Erschließung und Verfügbarkeit größtenteils als "graue Literatur" bezeichnet werden. Gleichzeitig erfordert die Suche nach relevanten Dokumenten und das Verbreiten von Informationen und Forschungsergebnissen einen immer höheren Zeitaufwand. Hochschulschriften stellen daher ein gutes Versuchsfeld für den Bereich des elektronischen Publizierens insgesamt dar.

Vor diesem Hintergrund entstand das Projekt eines Online Publikationsverbundes für Hochschulschriften (OPUS)¹, das vom DFN-Verein 1997 und 1998 im Rahmen des BWIN4M-Programmes gefördert und an der Universität Stuttgart vom Rechenzentrum und der Universitätsbibliothek durchgeführt wurde. Seit Februar 1998 öffentlich zugänglich, wird das entwickelte Volltextinformationssystem seit September 1998 im Regelbetrieb an der Universität Stuttgart eingesetzt. Jeder Universitätsangehörige, sei es Lehrender, Lernender oder Forschender, hat so die Möglichkeit, vom eigenen Arbeitsplatz aus Dokumente im World Wide Web zu veröffentlichen und gleichzeitig durch unterschiedliche Suchmöglichkeiten gezielt auf elektronische Texte zugreifen zu können.

Das System wurde rasch von anderen Hochschulen übernommen, da es im Produktionseinsatz sehr stabil war und vollständig auf frei verfügbaren Basiskomponenten wie PHP und MySQL aufbaute. Anfang 2003 wird es an 19 Universitäten (Bayreuth, Braunschweig, Freiburg, Giessen, Halle, Hamburg (Universität der Bundeswehr), Heidelberg, Hohenheim, Kaiserslautern, Kassel, Konstanz, Lüneburg, Mannheim, Marburg, Passau, Regensburg, Saarbrücken, Tübingen und Würzburg) an zwei Fachhochschulen (Dortmund und Hochschule der Medien Stuttgart) und an zwei Bibliotheks-Verbänden (Bibliotheksservice-Zentrum Baden-Württemberg und Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen) produktiv eingesetzt.

Ähnlich wie das Ariadne-System² für multimediale Lehr- und Lernmaterialien stellt OPUS ein einfaches Container-System für beliebige Objekte und Formate dar. Diese können über eine Webschnittstelle eingebracht, verwaltet und mit beschreibenden Daten versehen werden. Die Objekte selbst werden außer für eine Volltextsuche serverseitig nicht weiter aufbereitet.

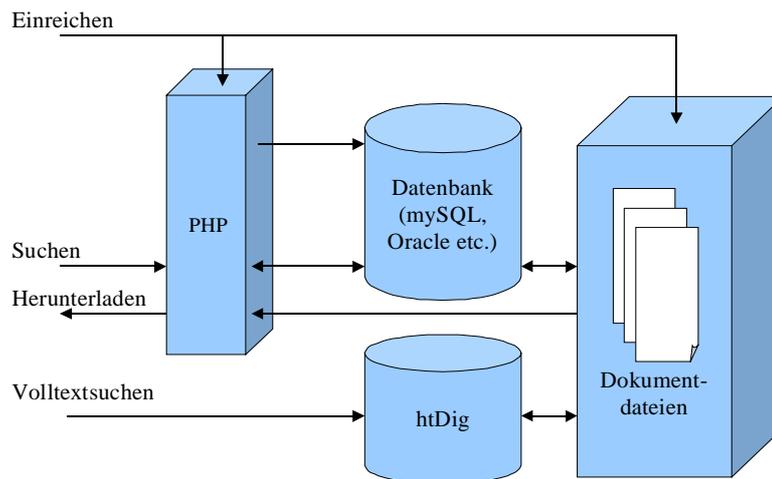
¹ <http://elib.uni-stuttgart.de/opus/> Vgl. auch Werner Stephan und Frank Scholze: Online Publikationsverbund: Erfassung und Organisation elektronischer Hochschulschriften, in: *Bibliotheksdienst* Heft 1, 1999, S. 92-102, http://www.dbi-berlin.de/dbi_pub/bd_art/bd_99/99_01_10.htm

² <http://www.ariadne-eu.org/>

Sie sind direkt im Dateisystem abgelegt. Die Metadaten können wiederum an unterschiedliche Suchdienste (Deutsche Bibliothek, Südwestdeutscher Bibliotheksverbund, Physnet, Mathnet, Suchmaschinen etc.) übergeben werden. Dies deckt die Suchgewohnheiten und –interessen unterschiedlicher Nutzergruppen ab.

Aufgrund der Verbreitung des Systems und der Förderung aus Mitteln der Zukunftsoffensive Junge Generation des Landes Baden-Württemberg wurde im Jahr 2001 eine vollständige technische Überarbeitung vorgenommen. Alle Programme sind nun in PHP 4 geschrieben und können über eine konfigurierbare Schnittstelle mit beliebigen SQL-Datenbanken betrieben werden. Als Volltextsuchkomponente wird das ursprünglich an der San Diego State University entwickelte htDig³ verwendet. Es könnten mit geringem Aufwand andere auch kommerzielle Suchmaschinen an dieser Stelle eingesetzt oder ganz auf eine lokale Volltextindizierung verzichtet werden und z.B. die site-Suche von Google in einer eigenen Suchoberfläche angeboten werden. Da hier jedoch kein Einfluss auf die Aktualisierungsintervalle genommen werden kann, wurde diese Möglichkeit nicht weiter verfolgt.

Funktional wurde das System an einige regionale, nationale und internationale Entwicklungen im Bereich elektronisches Publizieren angepasst.



(Abb1: OPUS Systemarchitektur)

Autorenschnittstelle

Das in der Praxis bewährte Webformular zur Abgabe von Dokumenten wurde gestrafft und erweitert. Gemeinsam mit dem Bibliotheks-Servicezentrum Baden-Württemberg (BSZ) wurde eine Online-Unterstützung für die verbale inhaltliche Beschreibung von Dokumenten erarbeitet. Dabei steht dem Benutzer bzw. Autor ein kontrolliertes interdisziplinäres Vokabular⁴ zur Verfügung, in dem er suchen und blättern kann, um dann passende Begriffe in einen Warenkorb zu legen. Diesen kann er der übrigen Beschreibung seines Dokuments (Name, Titel, Zusammenfassung etc.) hinzufügen. Der Online-Abgabevorgang umfasst damit für den Autor nur noch drei Schritte

- Beschreibende Daten des Dokuments eingeben

³ <http://www.htdig.org/>

⁴ Schlagwortnormdatei (SWD)

- Daten kontrollieren und ggf. korrigieren
- Dokument-Datei(en) auf den Server laden

Die Schlagwörter werden beim Export aus OPUS in den Südwestdeutschen Bibliotheksverbund und lokale Online-Kataloge mit dort vorhandenen Datensätzen verknüpft, so dass hierüber eine Suche auch nach verwandten Begriffen und Synonymen gegeben ist.

Uniform Resource Name

Durch die in den letzten Jahren stark gestiegene Zahl von Online-Publikationen ist für eine effiziente Nutzung die Entwicklung eindeutiger und dauerhafter Adressierungsverfahren unumgänglich. Zur Zeit wird zur Identifikation von Online-Dokumenten der zugehörige Uniform Resource Locator (URL) zitiert. URLs sind jedoch als Standortangabe im Internet oft instabil, weshalb Online-Publikationen bei einer sich verändernden URL nicht mehr auffindbar sind. Eine Lösung dieses Problems ist die Einführung sogenannter Persistent Identifier (PI). Für ein Dokument, das einen Persistent Identifier besitzt, kann dauerhaft der Zugriff und damit die Referenzierbarkeit gewährleistet werden. Gleichzeitig wird mit der Verwendung von Persistent Identifier der Schutz von Rechten an elektronischen Publikationen organisiert und verwaltet. Gegenwärtig existieren verschiedene Systeme zur eindeutigen Identifizierung von elektronischen Ressourcen. Beispiele dafür sind der Digital-Object-Identifier (DOI) der DOI-Foundation⁵, ein internationales Identifikationssystem von Verlagspublikationen, und der Uniform Resource Name (URN) der IETF⁶, ein nichtkommerzielles Identifikationssystem.

Im Rahmen des BMBF-Projekts Carmen AP4⁷ und seinem Nachfolger EPICUR⁸ wird bei der Deutschen Bibliothek (DDB) ein nationaler Resolving-Dienst für einen definierten Namensraum von Uniform Resource Names, die sogenannte National-Bibliography-Number (NBN), eingerichtet. Der Dienst bei der Deutschen Bibliothek ist zunächst nur auf wissenschaftliche Online-Publikationen wie Dissertationen und Habilitationen ausgerichtet.

Im Rahmen von OPUS werden daher zunächst für Dissertationen und Habilitationen Uniform Resource Names erzeugt und bei der Deutschen Bibliothek registriert. Dies geschieht über eine automatisch generierte E-Mail aus OPUS über die dazu erweiterte Metadiss-Schnittstelle⁹ der Deutschen Bibliothek. Für andere Dokumente (Aufsätze, Diplomarbeiten etc.) ist geplant, URNs innerhalb des gleichen Namensraumes beim Bibliotheks-Servicezentrum Baden-Württemberg bzw. je nach regionaler Zuständigkeit bei weiteren Verbundzentralen zu registrieren. Die Registrierungsinstanzen in der DDB und den Bibliotheksverbänden sollen die Daten gegenseitig replizieren. Offen ist derzeit noch die Frage, wann ein globaler, dem DNS vergleichbarer Auflösungsdienst für URNs soweit normiert und etabliert ist, dass er auch von Firmen unterstützt wird.

Open Archives Initiative¹⁰

Die Open Archives Initiative entwickelte sich aus Diskussionen im Kontext von Preprint-Servern und digitalen Bibliotheken in den USA mit zwei eindeutigen Zielen:

- ein gemeinsamer Satz von Metadaten zu bestimmen, der relativ einfach und handhabbar ist, und
- ein Navigationsprotokoll zu entwickeln, das auf der Basis eines Dienstprotokolls mittels dieses Metadatensatzes die wichtigsten Informationen aus den Servern einsammelt und für eine Recherche in einen Suchindex zusammenstellt.

⁵ <http://www.doi.org>

⁶ Internet Engineering Task Force (IETF); <http://www.ietf.org/html.charters/urn-charter.html>

⁷ http://www.bis.uni-oldenburg.de/carmen_ap4/

⁸ <http://www.ddb.de/professionell/projekte.htm#epicur>

⁹ <http://deposit.ddb.de/metadiss.htm>

¹⁰ <http://www.openarchives.org/>

Open bedeutet in diesem Zusammenhang, dass die Metadaten der Dokumente (nicht notwendigerweise die Dokumente selbst) frei zugänglich sind. Archiv bedeutet inzwischen jede Institution, die Online-Publikationen oder Objekte anbietet. Entstanden ist der Terminus aus dem Gedanken, Preprint-Server als Archiv bis zur Veröffentlichung in anderen Medien zu benutzen.¹¹

Sehr schnell wurden die beschriebenen Ziele umgesetzt und in Europa aufgegriffen. In Deutschland gab die Deutsche Initiative für Netzwerkinformation (DINI) auf ihrer ersten Jahrestagung im September 2000 bereits erste Empfehlungen zur Implementierung der Ergebnisse der Open Archives Initiative.¹² Es folgten Vorträge und Workshops, um Interesse zu wecken und Hilfestellungen zu geben, an denen sich auch Mitarbeiter von OPUS beteiligten.

In der neuen OPUS-Version 2.0 wird das Protokoll der Open Archives Initiative auf zweifache Weise genutzt. Einmal wie vorgesehen zum Austausch der einfachen Dublin Core Metadaten zwischen Archiven (Dataprovider) und Suchdiensten (Serviceprovider). Zum anderen wird jedoch auch der wesentlich komplexere Datenaustausch zwischen den Dokumentservern in der Region des Südwestdeutschen Bibliotheksverbundes und der Verbundzentrale (BSZ) hierüber abgewickelt. Diesen Daten liegt das gemeinsame Datenmodell DImeta¹³ zur Beschreibung und Verwaltung unterschiedlicher Online-Ressourcen für den Einsatz in wissenschaftlichen Einrichtungen in Baden-Württemberg zugrunde. Die OPUS-Archive im Einzugsbereich des Südwestdeutschen Bibliotheksverbundes treten dadurch gegenüber globalen OAI-Suchdiensten nicht mehr einzeln als kleine Speicher einiger hundert oder tausend Dokumente auf, sondern werden über das Bibliotheks-Servicezentrum Baden-Württemberg gebündelt. Für den Suchenden besteht der Vorteil darin, dass er nicht eine unübersichtliche Zahl von getrennten Ergebnismengen aus kleinen und kleinsten Archiven angezeigt bekommt.

Ausblick

Aufgrund des relativ großen Anwenderkreises sind auch nach Einführung der Version 2.0 neue Anforderungen entstanden. Drei Bereiche sollen zum Abschluss skizziert werden, um einen Ausblick auf die weitere Entwicklung von OPUS zu geben.

Zum ersten sollen die administrativen Funktionen ausgebaut werden, damit z.B. Zugriffsbeschränkungen für einzelne Dokumente über die Administrationsoberfläche vergeben werden oder nachträglich Dateien eingespielt werden können. Zweitens ist eine Anbindung an die Workflow Software geplant, die im Rahmen des DFG-Projekts German Academic Publishers (GAP)¹⁴ entwickelt wird, um auch begutachtete Publikationen mit einem entsprechenden Hinweis in OPUS übernehmen zu können. Um einen automatischen Datenaustausch beider Systeme zu ermöglichen, müssen die verwendeten relevanten Teile der jeweiligen Metadaten schemata aufeinander abgebildet werden. Im Bereich elektronischer Zeitschriften müssen zudem ggf. hierarchische Browsingfunktionalitäten von OPUS erweitert werden.

Ähnlich wie in Ariadne sollen drittens auch multimediale Dokumente, die z.B. im Rahmen des DFG-Projekts Multimedia Versatile Document Architecture (MAVA)¹⁵ entstehen, in OPUS angeboten werden können. Bei MAVA handelt es sich um ein multimediales Autoren- und

¹¹ Vgl. Diann Rusch-Feja: Die Open Archives Initiative (OAI). In: *Bibliothek. Forschung und Praxis* 25 (2001) Nr. 3, S. 291-300.

¹² <http://www.dini.de/dokumente/oai-appell-0912.pdf>

¹³ <http://www.dlmeta.de/>

¹⁴ <http://www.gap-c.de>

¹⁵ <http://www.ub.uni-stuttgart.de/ub/mava/devs/mava.html> Vgl. Jürgen Hauser und Frank Scholze "Integration und Verarbeitung multimedialer Dokumente in Digitalen Bibliotheken - der MAVA-Ansatz" In: *BIT Online*, 2001, Heft 1, S. 35 - 41.

Präsentationssystem, das durch seine Plattformunabhängigkeit und Erweiterbarkeit ideal für die Erstellung kleinerer und mittelgroßer Präsentationen und Lernmodule an Hochschulen geeignet ist. Je nach Größe und lokalen Voraussetzungen können derartige Dokumente als Applet oder lokal abgespielt werden. Dies bedeutet nicht, dass OPUS Funktionalitäten einer Lernplattform übernehmen wird, sondern lediglich als Speicher für einzelne stabile und längerfristig vorzuhaltende multimediale Module dienen kann.