

OPUS und andere digitale Dokumentspeichersysteme

Vortrag beim CMS-Workshop in der UB Düsseldorf am 14.10.2004

Friedrich Summann
Universitätsbibliothek Bielefeld

Seit 1996 befasst sich die Universitätsbibliothek Bielefeld mit der Speicherung und Bereitstellung digitaler Dokumente. Zunächst wurden diese Produktionen für von der Hochschule gelieferte Dokumente und selbst aufbereitete Werke (ältere Drucke aus dem Bestand der Bibliothek) mit eigen-entwickelten Tools erstellt und für die web-basierte Nutzung bereitgestellt. Für die Bereitstellung von Hochschulschriften wurden zunächst einfache HTML-Listen verwendet, bis für diesen Zweck im vorherigen Jahr eine OPUS-Installation eingeführt wurde. Für das DFG-Projekt Retrospektive Digitalisierung Zeitschriften der Aufklärung wurden seit 2001 verschiedene Zusatzentwicklungen durchgeführt. Parallel zu diesen technischen Entwicklungen wurde schon seit längerem die Evaluierung geeigneter Systeme durchgeführt, deren Ergebnisse hier vorgestellt werden sollen. Insbesondere standen dabei weiterführende Aspekte wie die Volltextindexierung, die Behandlung von Strukturdaten und die Einbeziehung des Publikationsprozesses im Mittelpunkt der Diskussion.

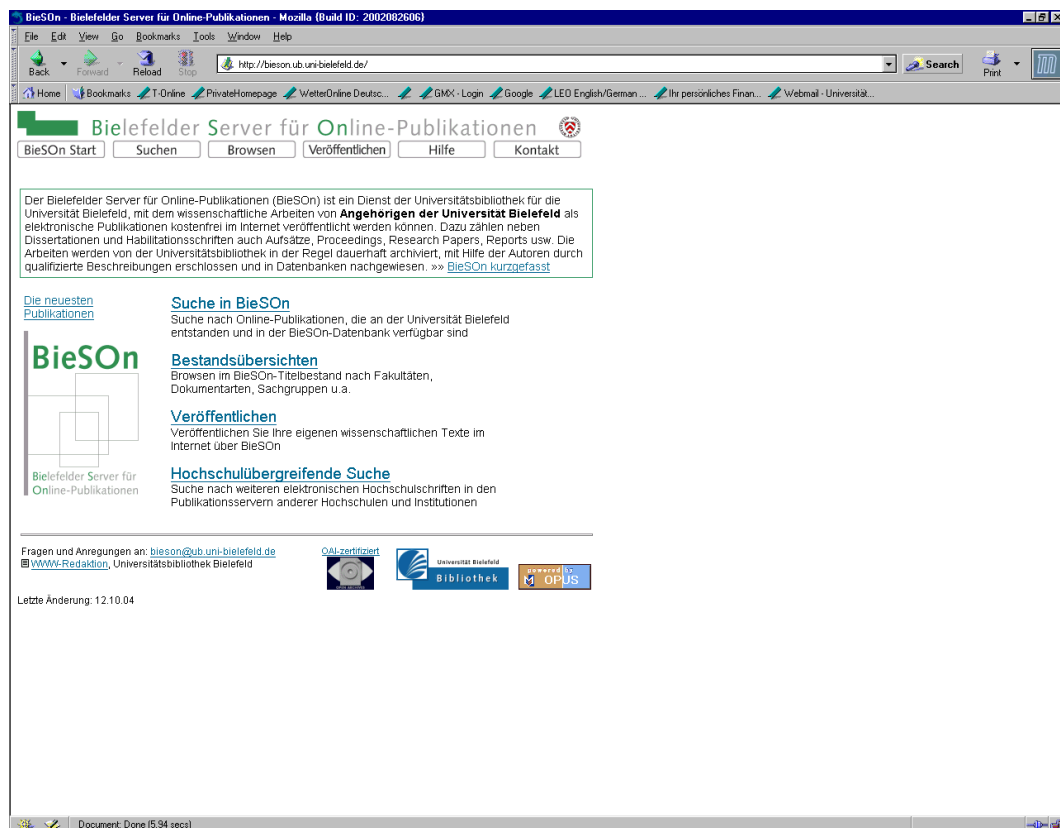


Abbildung 1: Startseite des Bielefelder BIESON-Servers

Das OPUS-System der UB Stuttgart gehörte zu den ersten Systemen, die für einen Einsatz betrachtet wurden. Nach ausführlicher Untersuchung und Anpassung an die lokalen Vorstellungen wurde OPUS 2003 als BieSON (Bielefelder Server für Online-Publikationen) (s. auch Abb. 1) zur Bereitstellung von Hochschulschriften und eigenen UB-Publikationen zum Einsatz gebracht. Zur Zeit stehen 453 Dokumente auf dieser Plattform zur Nutzung bereit. Mit Hilfe des HBZ Köln konnte die OAI-Schnittstelle OAI-2.0-kompatibel seit Sommer 2004 eingesetzt werden. Der Server ist daraufhin bei den bekanntesten OAI-Registries

(OpenArchives Initiative, Univ. of Illinois Libraries) eingetragen worden und wird zur Zeit auch bei OALster (Univ. of Michigan Libraries) als Neuzugang gemeldet. Die in der UB Bielefeld seit 1996 aufbereiteten digitalen Drucke (ca. 30) aus dem Bibliotheksbestand wurden gescannt und die dabei erzeugten Bilddateien werden über eine Benutzeroberfläche angeboten, bei dem ein Perl-Skript die Anzeige und Navigation steuert. Bei den Zeitschriften der Aufklärung aus dem 18. und 19. Jahrhundert (Abb. 2) liegen neben TIFF-Dateien auch Metadaten (geliefert im SGML-Format) vor. Für die Bereitstellung dieser Anwendung ist eine Mischung aus Perlskripten und einer Volltextdatenbank zum Retrieval in den Metadaten entwickelt worden. Die Skripte übernehmen die notwendigen Formatttransformationen, die Erzeugung der HTML-Übersichtsseiten für die Darstellung der Zeitschriftenstruktur (Titel-, Band-, Heftübersichten) und die on-the-fly-Konvertierung (TIFF nach GIF) und Anzeige der Bildseiten.

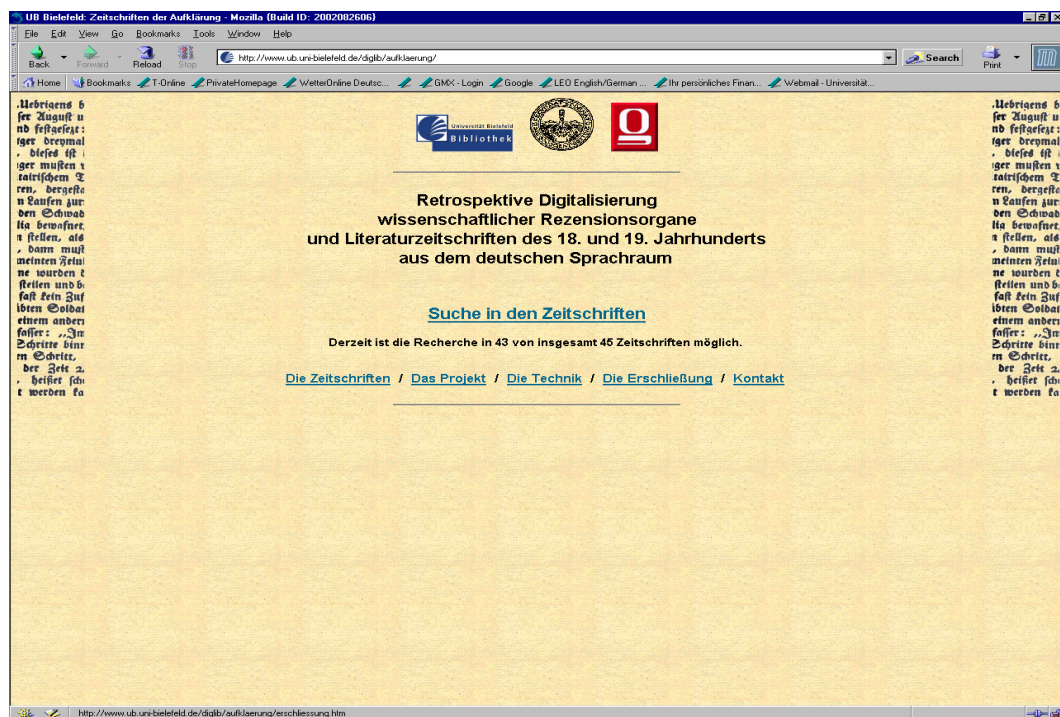


Abbildung 2: Startseite des Projektes Retrospektive Digitalisierung Zeitschriften der Aufklärung

Um derartige komplexen Strukturen adäquat umsetzen zu können, kann man sich die weitgehende Unterstützung durch ein Softwaresystem vorstellen. Dies war ein wesentlicher Grund für die UB Bielefeld, sich nach geeigneten Systemen im Bereich digitaler Dokumentspeicher umzusehen und diese näher im Hinblick auf ihre Einsatzfähigkeit zu untersuchen.



DSpace (<http://www.dspace.org/>) ist eine gemeinsame Entwicklung von Hewlett-Packard und den MIT Libraries.

Folgende grundlegende Eigenschaften gelten für DSpace:

- • Open Source
- • Unix/Linux
- • Java-basiert (Apache, Tomcat, Ant)
- • Nutzung von PostgreSQL, Lucene
- • OAI-Schnittstelle

DSpace wird seit der Bereitstellung 2002 in einigen Installationen eingesetzt, auch in Europa z.B. bei der Erasmus Universität in Rotterdam oder der Univ. Roskilde in Dänemark. Auffällig ist, dass die DSpace-Installationen bisher nur im Bereich Hochschulschriften zum Einsatz kommen und dabei diese eher unauffällig (auch die bei den MIT Libraries selbst) erscheinen. Im November 2003 wurde in einem Gespräch von den MIT Libraries dargelegt, dass die weitere Entwicklung der DSpace-Community überlassen werden soll. Schwerpunktmässig sollen die Bereiche Langzeitarchivierung und METS-Support bearbeitet werden.

In Bielefeld existiert eine eigene DSpace-Installation unter Linux, die für die Evaluation betrachtet wurde. Installation und Handhabung ergaben keine ernstlichen Probleme; die Funktionalität schien aber nicht über die von OPUS hinauszugehen.



Fedora (<http://www.fedora.info/>) ist eine Entwicklung der University of Virginia und der Cornell University, wobei hier die Informatiker und nicht die Bibliotheken die Entwicklung betreiben. Für das System gelten die folgenden Grundlagen:

- • Open Source
- • Linux/Unix/Windows – MySQL/Oracle
- • Java, objekt-orientiert
- • WebServices-Schnittstellen
- • Volltextretrieval (XPAT, eXist)
- • OAI-Schnittstelle

Fedora ist das jüngste der untersuchten Systeme. Es gibt bisher nur vergleichsweise wenige Installationen und Projekte. Auch bei den University of Virginia Libraries, die mit ihrem E-Text-Centre seit 1992 sehr erfolgreich eine digitale Textsammlung betreiben ist eine Fedora-Nutzung bisher nicht über Umsetzungsbemühungen hinausgekommen. Im Gespräch mit den Fedora-Entwicklern und bei der praktischen Demonstration der Software drängte sich der Eindruck auf, dass die Software sehr ambitioniert, theoretisch fundiert und hochkomplex ist, beim praktischen Einsatz aber einige Probleme aufwirft. Als weiteres Entwicklungsziel wird für Fedora die "digital library in a box" ausgegeben, wobei mit Dokumentation und Tools die praktische Einsetzbarkeit entscheidend verbessert werden soll.

greenstone digital library software

Greenstone (www.greenstone.org) ist eine Software, die von der University of Waikato, Neuseeland mit Unterstützung der UNESCO entwickelt und vertrieben wird. Die Software hat die folgenden Basiseigenschaften:

- •Open Source (Windows, Linux, Unix)
- •Perl, C
- •Umfangreiche benutzerorientierte Werkzeuge zur Produktion von Sammlungen
- •Eigene Volltextengine (mit linguistischen Funktionen)
- •CD-ROM-Export

Von der Software existieren weltweit zahlreiche Installationen in verschiedenen Sprachen mit sehr unterschiedlichen Inhalten. Es existieren umfangreiche Dokumentationen - zum Teil von nutzenden Institutionen – und eine rege Community mit Mailing-Listen (jeweils Entwickler und Anwender). Die Software hat den Schwerpunkt und ihre Stärke in der Erstellung von digitalen Kollektionen unter Einbeziehung unterschiedlicher Materialien. Dazu existieren für unterschiedliche Formate (HTML, PDF, Word, Z39.50, OAI, Emails, TIF usw.) Einlesetools und für die Benutzeroberfläche anpassbare Konfigurationsdateien und Perlskripte. Bei eigenen Installationen sowohl unter Windows als auch unter Linux konnten wir uns von der einfachen Installation und praxis-orientierten Handhabung überzeugen. Schwächen hat die aktuelle Version (2.51) von Greenstone mit der Bereitstellung von Standard-Schnittstellen, es existiert insbesondere keine OAI-Schnittstelle, die allerdings seit Sommer 2003 angekündigt wird.

Neben den oben aufgeführten Repository-Systemen gibt es noch weitere auf dem Markt, mit denen sich die UB Bielefeld weniger intensiv beschäftigt hat. Wegen der **MILESS/Mycore Software** gab es Gespräche mit IBM, die den Eindruck, die Software sei hochkomplex, eher verstärkt haben.

Mit **EPrints** hat sich die UB Bielefeld erst in jüngster Zeit intensiver befasst. Das System der University of Southampton ist schon länger verfügbar und hat zahlreiche Installationen in Europa (ca. 130) und stellt eine ähnliche Funktionalität wie OPUS und DSpace bereit. Gelungene Installationen finden sich bei der Universität Glasgow und der Bibliothek der Ludwig-Maximilians-Universität in München.

Interessant für die UB Bielefeld sind außerdem die Systeme, die in den University of Michigan Libraries und den Cornell University Libraries verwendet werden, da beide Bibliotheken umfangreiche digitale Produktionen erstellt haben und zu beiden Institutionen gute Kontakte bestehen. In Michigan wird erfolgreich das **DLXS-System** benutzt, das eine Eigenentwicklung auf Basis der lizenzierten XPAT-Software von OpenText darstellt.

An den Cornell University Libraries wird **DPubS** verwendet, eine Eigenentwicklung, die ursprünglich von den Informatikern der Cornell University stammt und später von der Bibliothek übernommen und weiterentwickelt wurde. Neben einigen digitalen Sammlungen wird insbesondere das Projekt Euclid mit DPubS betrieben, das mathematische Verlagszeitschriften anbietet. Da die Software in Zukunft im Rahmen eines Mellon-Projektes zur Open-Source-Software entwickelt wird, kann eine Nutzung auch für die UB Bielefeld durchaus interessante Aspekte bieten.