Wie entsteht ein Institutional Repository? – Eine systematische Hinführung in acht Schritten

Susanne Dobratz | dobratz@cms.hu-berlin.de
Uwe Müller | u.mueller@cms.hu-berlin.de

Der Artikel beschreibt in Form eines instruktiven Leitfadens mögliche Herangehensweisen beim Aufbau eines Institutional Repository und stellt Vorüberlegungen und Entscheidungsoptionen dar, die sich für Vorbereitung und Betrieb einer Publikationsplattform ergeben. Die Definition der Anforderungen und die darauf aufbauende Argumentation erfolgen auf der Basis des DINI-Zertifikats für Dokumenten- und Publikationsservices.

Einleitung

Ein digitales Repository gehört heute zur Standardausstattung nahezu jeder wissenschaftlichen Einrichtung. Fast alle Universitäten und viele Fachhochschulen, Forschungseinrichtungen und Informationseinrichtungen bieten Publikationen – zumeist solche eigener Mitarbeiter bzw. Studierender – in elektronischer Form an. Aufgrund der Vielzahl und der Verschiedenheit von Repositories, die nicht immer durch die jeweiligen Einrichtungsbibliotheken betrieben werden und von denen es an vielen Einrichtungen gleich mehrere gibt, und wegen der hohen Dynamik in diesem Bereich ist es kaum möglich, einen umfassenden Überblick über den aktuellen Stand in Deutschland zu erlangen. Das *Registry of Open Access Repositories*[1] (ROAR), das von der University of Southampton betrieben wird, listet derzeit 97 deutsche Dokumentenserver auf. Eine von DINI gepflegte Übersicht[2] enthält derzeit sogar 138 Einträge, von denen insgesamt 22 das DINI-Zertifikat tragen, sechs davon dessen aktuelle Version aus dem Jahr 2007. Größere Einrichtungen wie Universitäten verteilen ihre elektronischen Publikationen zuweilen auf mehrere Repositories – beispielsweise einen Medienserver zur Verwaltung multimedialer Lehr- und Lernmaterialien, einen Hochschulschriftenserver zur Veröffentlichung universitärer Qualifikationsarbeiten und eine Publikationsplattform für Pre- und Postprints wissenschaftlicher Veröffentlichungen. Belastbare exakte Zahlen sind auch deshalb schwer zu ermitteln, weil viele Repositories Teil eines größeren Angebots, z. B. eines Fachportals und in den entsprechenden Verzeichnissen nicht als separater Publikationsserver gemeldet sind.

1 siehe http://roar.eprints.org/ [09.02.2009]
2 siehe http://www.dini.de/wiss-publizieren/repository/ [09.02.2009]

Was ist ein digitales Repository?

Ein *Repository* (engl. für Depot, Lager, Speicher oder Repositorium) wird als System verstanden, mit dem eine Sammlung elektronischer Objekte verwaltet und nach außen online zur Verfügung gestellt werden kann. Dabei handelt es sich in der Regel vor allem um textorientierte Dokumente, was andersartige digitale Objekte jedoch nicht ausschließt. Wissenschaftliche Repositories stellen das Rückgrat des so genannten Grünen Weges zu Open Access dar (siehe [1]).

Gegenstandsbereich eines *Institutional Repository* (IR) sind die elektronischen Publikationen oder andere digitale Objekte, die durch die Mitarbeit von Mitgliedern einer – zumeist wissenschaftlichen – Institution entstanden sind. Das Institutional Repository einer solchen Einrichtung gehört demnach zu deren Informationsinfrastruktur. Es bildet eine Art digitales Schaufenster, das den wissenschaftlichen Output der Institution – zumindest teilweise – nach außen darstellt. Mit dieser Prämisse sollten die Überlegungen zum Aufbau eines eigenen IR beginnen, der im Folgenden schrittweise beschrieben wird.

1. Schritt: Klärung grundlegender Fragestellungen

Zunächst ist es wichtig, Klarheit über folgende Fragestellungen in Bezug auf das aufzubauende Institutional Repository zu gewinnen:

1. Welchem übergreifenden Zweck soll das IR dienen?
2. Wer soll welche Inhalte über das IR publizieren?
3. An welche potentielle Leserschaft soll sich das IR richten?
4. Wie soll sich das IR in andere Aktivitäten im Bereich der Informationsbereitstellung integrieren?

Anhand der Antworten sollte auch versucht werden, das jährlich zu erwartende Publikationsvolumen abzuschätzen und notwendige Maßnahmen zur Unterstützung von Autoren sowie weitere bereitzustellende Dienstleistungen und technische Schnittstellen abzuleiten.

Der Klärung der Frage, welche Publikationen in Form von Open Access veröffentlicht werden sollen, sollte besondere Aufmerksamkeit geschenkt werden. Ihr liegen prinzipielle Erwägungen zugrunde – etwa eine mögliche Open-Access-Strategie der jeweiligen Einrichtung. Sie muss aber im Einzelfall vielfach in Kooperation mit den jeweiligen Autoren und vor allem mit den potentiellen Herausgebern abgewogen werden.

Zu den ersten Aufgaben beim „Neustart" eines IR gehört es außerdem, dieses organisatorisch sinnvoll zu verankern und mit entsprechenden technischen und personellen Kapazitäten auszustatten. Auch hierfür ergeben sich wichtige Hinweise aus der Beantwortung der oben genannten Fragen.

Zum Beispiel: Die HU

Sollen als Autoren – wie beispielsweise an der Humboldt-Universität – grundsätzlich alle Wissenschaftlerinnen und Wissenschaftler bzw. Studierende, die gleichzeitig Angehörige der jeweiligen Hochschule sind, angesprochen werden, ist ein großes Spektrum an unterschiedlichen Publikationsarten und eine Vielzahl von Wissenschaftsdisziplinen zu berücksichtigen. Die Betreuung dieser durch sehr heterogene Anforderungen geprägten Veröffentlichungen muss sich entsprechend auch in den Leitlinien für das Institutional Repository wiederfinden. Einen dergestalt konzipierten Dokumentenservers innerhalb der zentralen Einrichtungen der Hochschule anzusiedeln, liegt nahe und verlangt zentral von der Universität gesteuerte Ressourcen.

Sobald diese grundlegenden Fragen geklärt sind und die organisatorische Verortung des IR innerhalb der Einrichtung vollzogen ist, kann mit der praktischen Realisierung begonnen werden. Besonders hilfreich hierfür sind die Hinweise und Empfehlungen der Deutschen Initiative für Netzwerkinformation (DINI, siehe Kasten).

Das DINI-Zertifikat

Die 1998 gegründete Deutsche Initiative für Netzwerkinformation, ein Zusammenschluss von Rechenzentren, wissenschaftlichen Bibliotheken, Medienzentren und Vertretern aus dem Wissenschaftsbereich, vergibt seit 2004 das so genannte DINI-Zertifikat für Dokumenten- und Publikationsservices. In seiner aktuellen Fassung aus dem Jahr 2007 beschreibt das DINI-Zertifikat Qualitätsstandards für die zu verwendenden Technologien, die organisatorische Einbettung und den Betrieb von IRs einschließlich der dazugehörigen Dienstleistungen [2]. Die Lektüre der Anforderungen, die im DINI-Zertifikat als Mindeststandards beschrieben werden, bietet sich als hilfreiche Orientierung für den Aufbau eines IR an. Sie kann darüber hinaus als Entscheidungsgrundlage dafür dienen, ob der Service tatsächlich an der eigenen Institution eingerichtet werden oder im Rahmen einer externen Kooperation durch einen Dritten erbracht werden soll. Gerade für kleinere Institutionen ist das technologische Hosting eines IR häufig überlegenswert (vgl. [3]).

Innerhalb Deutschlands bilden die im DINI-Zertifikat festgelegten Anforderungen unter anderem die Grundlage für die Vernetzung der einzelnen IRs sowie den Aufbau darauf basierender übergreifender Dienste. Beispielsweise arbeitet das durch die Deutsche Forschungsgemeinschaft geförderte Projekt *Open-Access-Netzwerk* auf eine solche Vernetzung deutscher IRs hin und zielt dabei auf eine verbesserte nationale und internationale Wahrnehmung der lokal angebotenen Publikationen und damit des Forschungsbeitrags insgesamt (siehe [4]).

2. Schritt: Integration des IR in die strategische Planung

Im nächsten Schritt ist es wichtig, für das Vorhaben die Unterstützung innerhalb der oberen Entscheidungsebenen der Institution zu gewinnen. Das ist unter anderem erforderlich, um die Nachhaltigkeit des aufzubauenden Repository zu gewährleisten und dessen Betrieb für die nächsten Jahre und Jahrzehnte finanziell abzusichern. Dabei ist zu berücksichtigen, dass vor allem Personalkapazitäten langfristigen Planungen und Verträgen unterliegen. Aber auch die regelmäßige Erneuerung der Hard- und Software für das IR verursacht Folgekosten. Durch die strategische Planung wird die Einbettung des IR in die Informationsinfrastruktur der Einrichtung und darüber hinaus gewährleistet. Es muss an fachliche und bibliothekarische Netzwerke sowie an weitere Aktivitäten zur Unterstützung der digitalen Forschung und Lehre angebunden werden.

3. Schritt: Formulierung von Leitlinien für das IR

Die Leitlinien für ein Institutional Repository dienen der Dokumentation der Zielstellungen, Grundsätze und Rahmenbedingungen. Sie stellen damit öffentlich dar, wie die im ersten Schritt dieses Artikels formulierten Grundsatzfragen für das jeweilige IR beantwortet werden. Im DINI-Zertifikat ist eines der Kriterien dieser Problematik gewidmet und es wird detailliert beschrieben, was in einer solchen *Policy* enthalten sein sollte. Die Leitlinien stellen sowohl für potentielle

Autoren als auch für die Nutzer der sonstigen Dienstleistungen des IR klar, auf welcher Basis es betrieben wird und welche Erwartungen sie in Bezug auf Verfügbarkeit, Zuverlässigkeit und Qualität der Dienste haben dürfen.

Gemäß DINI-Zertifikat muss die Policy für ein IR folgende Punkte beinhalten:

1. eine Beschreibung der Zielgruppe und Aussagen darüber, wer auf dem Repository publizieren darf,
2. eine Definition von Verfahren zum Betrieb des IR und die Festlegung von Leistungen, die der Anbieter des Publikationsservice Autoren und Herausgebern bereitstellt,
3. die Festlegung der inhaltlichen, funktionalen und technischen Qualität der auf dem IR veröffentlichten Dokumente im Sinne einer Beschränkung (beispielsweise auf wissenschaftliche Abhandlungen oder Veröffentlichungen, die einen Peer-Review-Prozess durchlaufen haben),
4. eine Definition von Archivierungszeiträumen und die Garantie, die Dokumente in Abhängigkeit ihrer inhaltlichen, funktionalen und technischen Qualität für einen definierten Zeitraum nutzbar anzubieten,
5. eine Beschreibung der Rechte und Pflichten von Autoren und Herausgebern,
6. eine Erklärung zu Open Access, die das jeweilige Verständnis des Begriffs in der Institution ebenso enthält wie einrichtungsspezifische Handlungsrichtlinien für Autoren (ggf. bezogen auf bestimmte Publikationsarten) und vorhandene Umsetzungsinstrumente.

4. Schritt: Bereitstellung von Personal

Entscheidend für die erfolgreiche Realisierung eines Institutional Repository an einer wissenschaftlichen Einrichtung ist nicht allein die verwendete Technologie, sondern auch das zur Verfügung stehende Personal. Das Ziel muss darin bestehen, ein Team zusammenzustellen, das alle notwendigen Fähigkeiten und Fertigkeiten zum Aufbau, zur Wartung und zum Betrieb des IR und der damit verbundenen Dienstleistungen abdeckt. Zumindest bibliothekarische und Informatik-Kompetenzen sollten im Team vertreten sein. Darüber hinaus sollte der Bereich der Öffentlichkeitsarbeit bzw. des Marketing fachlich abgedeckt sein.

Abgesehen davon ist der Aspekt der Nachhaltigkeit zu berücksichtigen. Befristete und fremdfinanzierte Projektmitarbeiter sind für die Konzeption und den unmittelbaren Aufbau eines IR und den damit verbundenen Service durchaus hilfreich und oftmals alternativlos. Allerdings sollte bereits zu Beginn eines solchen Vorhabens bedacht werden, wie Betrieb und Wartung des IR nach Ablauf einer Startphase mit unbefristet tätigem Personal abgedeckt werden können und das entsprechend notwendige Know-how weitergegeben werden kann, um die Stabilität des entstandenen Dienstes auch langfristig sicherzustellen.

5. Schritt: Auswahl der Technologie

Für die Auswahl der geeigneten Technologie ist entscheidend, die gängigen Anforderungen, die an ein IR und die darauf basierenden Funktionen und Dienstleistungen gestellt werden, zu berücksichtigen und in bestmöglicher Weise zu erfüllen. Auf der Basis internationaler Entwicklungen und anerkannter Standards hat DINI für die technologische Basis von Repositories Mindeststandards erarbeitet, die im bereits erwähnten DINI-Zertifikat festgeschrieben sind (siehe Kasten auf Seite 50).

Zu den allgemeinen Aspekten, die zur Bewertung von IR-Software herangezogen werden (siehe z. B. [5], [6]), gehören unter anderem die Einfachheit der Installation und der Wartung des Systems, die Möglichkeit, mehrere Installationen gleichzeitig innerhalb einer Hard- und Softwareumgebung laufen zu lassen (relevant beispielsweise für Hosting-Services, siehe [3]), die Interoperabilität mit anderen Systemen und das Vorhandensein entsprechender Schnittstellen, um das IR in die Informationsinfrastruktur der Einrichtung und darüber hinaus zu integrieren. Beispielsweise mag es aus Sicht einer konkreten Einrichtung von Bedeutung sein, das IR in den OPAC zu integrieren, in Lehr- und Lernumgebungen einzubinden oder bei der Authentifizierung auf das lokale Nutzermanagement (LDAP) zurückzugreifen. Weitere Aspekte sind die Skalierbarkeit und die Sicherheit des Systems, die Größe der Entwickler- und Nutzergemeinde der Software, die jeweiligen Lizenzbedingungen für die Nachnutzung sowie die Möglichkeiten, eigene Veränderungen und Erweiterungen des Quellcodes vorzunehmen.

Die meisten der heute verwendeten Lösungen sind ursprünglich aufgrund einer am konkreten Bedarf einer einzelnen Institution ausgerichteten Entwicklung entstanden. In einigen Fällen ist daraus später ein nachnutzbares und an lokale Erfordernisse anpassbares Produkt geworden, dessen Weiterentwicklung inzwischen zumeist in den Händen einer größeren Entwicklergemeinde liegt und das die üblichen Zyklen einer Softwareentwicklung durchläuft. Inzwischen existiert eine ganze Reihe technischer Lösungen für Institutional Repositories, die sofort einsetzbar sind und die Mehrzahl der technischen Anforderungen des DINI-Zertifikats bereits erfüllen. Im nichtkommerziellen Bereich sind dabei international vor allem EPrints, DSpace und Fedora sowie innerhalb Deutschlands zusätzlich OPUS und MyCoRe zu nennen.

Spätestens hier stellt sich die Frage, welches der zur Verfügung stehenden Produkte ausgewählt werden sollte. Dies hängt nicht nur von den bereits genannten allgemeinen bzw. durch das DINI-Zertifikat gegebenen Anforderungen ab, sondern muss mit Berücksichtigung der lokalen Gegebenheiten und Erfordernisse entschieden werden, beispielsweise in Abhängigkeit der Art des jeweiligen Betreibers und der bereitzustellenden Publikationen. Beispielsweise werden Repositories nicht nur von Universitätsbibliotheken oder Rechenzentren betrieben, sondern als fachspezifische Publikationsplattformen zuweilen auch durch Fachbereiche, Fachgesellschaften oder außeruniversitäre Forschungseinrichtungen. Daher sind auch die Ansprüche, die hinsichtlich organisatorischer und technischer Fragestellungen an die Software gestellt werden, sehr unterschied-

Anforderungen an ein IR gemäß DINI-Zertifikat

(Die numerischen Referenzen beziehen sich auf die Strukturierung des Textes im DINI-Zertifikat [2].)

- **WWW:** Es muss eine Webschnittstelle für Nutzer vorhanden sein (Punkt 2.6.3).
- **Upload:** Die Bereitstellung von Online-Formularen für Autoren zum Upload ihrer Publikationen (Document upload tools), ist erforderlich (Punkt 2.3).
- **Rechte:** Der Autor kann eigene Publikationen über entsprechende Vereinbarungen mit Lizenzen versehen, z. B. CC, DPPL (Punkt 2.4).
- **Nachweis:** Die Dokumente müssen aus technischer Sicht kontrolliert und nachweisbar aufgenommen werden (Punkt 2.5.1).
- **Persistent Identifiers:** Die Verwendung von Persistent Identifiers ist erforderlich, z. B. URN, DOI (Punkt 2.5.2).
- **Keywords:** Jede Publikation sollte mit aussagekräftigen Schlagwörtern zumindest in deutscher und englischer Sprache versehen sein. Auf technologischer Ebene muss daher eine verbale Sacherschließung durch freie Schlagwörter oder ein Klassifikationssystem unterstützt werden (Punkt 2.6.1).
- **DDC:** In Ergänzung dazu ist die Anwendung der Dewey-Dezimalklassifikation (DDC) gemäß der verwendeten Sachgruppen in der Deutschen Nationalbibliografie erforderlich (Punkt 2.6.1).
- **Metadaten:** Die Metadaten müssen frei zugänglich angeboten werden und zumindest gemäß Dublin Core (ISO 15836:2003) strukturiert sein (Punkt 2.6.2).
- **OAI:** Das IR muss über eine OAI-Schnittstelle verfügen, die entsprechend den OAI-Empfehlungen von DINI realisiert ist (Punkt 2.6.2).
- **Statistik:** Jedes IR muss im Rahmen der rechtlichen Bestimmungen eine eigene konsistente Zugriffsstatistik führen (Webserver-Log) (Punkt 2.7).
- **Langzeitarchivierung:** Die Dokumente und ihre Metadaten müssen dauerhaft miteinander verbunden sein. Für die Langzeitarchivierung der Publikationen sollte die Kooperation mit einer anerkannten Archivierungsinstitution, wie z. B. der Deutschen Nationalbibliothek (DNB), gesucht werden (Punkt 2.8).
- **Dokumentation:** Es muss eine Dokumentation des technischen Systems mit relevanten Versionsangaben und technischen Parametern zu allen Komponenten existieren (Punkt 2.5.1).

lich. Die Wahl einer geeigneten Software hängt insbesondere auch davon ab, welche Anforderungen aus Sicht der jeweiligen Betreibereinrichtung tatsächlich Priorität besitzen. Zudem haben sich speziell in Deutschland Lösungen etabliert, die auf die spezifischen Belange hiesiger Universitätsbibliotheken zugeschnitten sind und somit im internationalen Vergleich wenig Berücksichtigung finden.

Der Rahmen des vorliegenden Artikels erlaubt zwar eine detaillierte Betrachtung der einzelnen Systeme und eine Analyse der jeweiligen Vor- und Nachteile nicht (siehe dazu beispielsweise [7] und [8]). Einige Anhaltspunkte zu den genannten Produkten für Institutional Repositories finden sich jedoch in den folgenden Abschnitten sowie im nebenstehenden Kasten (Seite 51).

OPUS

Die ursprünglich als *Online-Publikationssystem der Universität Stuttgart* entwickelte Software existiert seit 1998 und diente zunächst ausschließlich der Bereitstellung elektronischer Hochschulschriften an dieser Universität. Mit seinen bundesweit mehr als 60 Installationen ist *OPUS* heute die am weitesten verbreitete IR-Software in Deutschland und Grundlage für Hosting-Angebote mehrerer Bibliotheksverbünde. Durch die Ausrichtung auf deutsche Hochschulen unterstützt OPUS die wichtigsten hier erforderlichen Schnittstellen und Standards – etwa die Generierung von URNs, deren maschinenlesbare Bereitstellung im Format EPICUR zur Abfrage durch die Deutsche Nationalbibliothek (DNB) und eine den Anforderungen des DINI-Zertifikats entsprechende OAI-Schnittstelle, die neben Dublin Core auch das für die Meldung an die DNB erforderliche Metadatenformat XMetaDiss beherrscht. Darüber hinaus stellt OPUS ein Modul zur Definition eigener Nutzungslizenzen (u. a. Creative Commons) bereit, aus denen beim Upload von Publikationen ausgewählt werden kann.

OPUS unterstützt vor allem den Publikationsprozess in der Bibliothek: Nach der Einstellung eines Dokuments durch dessen Autor wird es durch einen Mitarbeiter bezüglich technischer und formaler Kriterien – etwa Vollständigkeit der Publikation und Korrektheit der Metadaten – überprüft, ggf. korrigiert und anschließend freigeschaltet. Die Publikationen können durch Browsing nach unterschiedlichen Kategorien sowie mit einer Suche in den Metadaten und den Volltexten (über Google) aufgefunden werden. Abgesehen von der Open-Access-Einstellung erlaubt es die Definition differenzierter Zugriffsbeschränkungen für Dokumente und ermöglicht darüber hinaus die Verwaltung von Schriftenreihen und Sammlungen mit beliebig gestuften Hierarchien.

OPUS ist ein robustes und vielfach erprobtes System, dessen Entwicklergemeinde jedoch sehr abgeschlossen ist und dessen Anpassungsmöglichkeiten – beispielsweise in Bezug auf das Datenmodell und die grafische Oberfläche – sehr beschränkt sind.

DSpace

DSpace, am Massachusetts Institute of Technology (MIT) in Kooperation mit der Firma HP entwickelt, befindet sich beim MIT seit April 2004 im Produktionsbetrieb. Die stetige Weiterentwicklung erfolgt inzwischen über die *DSpace Foundation*, in der sich weitere große Forschungseinrichtungen zusammengeschlossen haben, sowie parallel durch Einzelinstitutionen, da der Quellcode – als Open Source zur Verfügung steht. So hat beispielsweise die Universitätsbibliothek Kassel das System so erweitert, dass es auch das URN-Format der DNB unterstützt. In der Standardversion unterstützt DSpace bereits die gängigen Schnittstellen und Standards wie OAI,

IR-Software im Überblick

OPUS

Entwickler: Universität Stuttgart
Verbreitung: ca. 60 Installationen in Deutschland
Lizenz: Universität Stuttgart, 250 € einmalige Gebühr
Programmiersprache: PHP
Systemumgebung: Linux/UNIX, Apache, MySQL
Download: Per Mail über opus@uni-stuttgart.de
Support: Wiki, Bug-Tracker

DSpace

Entwickler: DSpace Foundation
Verbreitung: ca. 550 Installationen weltweit, davon 7 in Deutschland
Lizenz: Open Source (BSD)
Programmiersprache: Java, JSP
Systemumgebung: Linux/UNIX bzw. Windows, Apache, Tomcat, PostgreSQL bzw. Oracle
Download: http://sourceforge.net/project/showfiles.php?group_id=19984
Support: DSpace Federation, Bug-Tracker, Wiki, FAQ, User Group Meetings

EPrints

Entwickler: University of Southampton
Verbreitung: ca. 280 Installationen weltweit, davon ca. 10 in Deutschland
Lizenz: Open Source (GPL)
Programmiersprache: Perl
Systemumgebung: Linux/UNIX bzw. Windows, Apache, MySQL
Download: http://www.eprints.org/software/
Support: Wiki, Dateiarchiv, Mailingliste

MyCoRe

Entwickler: Hochschulübergreifendes Entwicklerteam
Verbreitung: ca. 20 Installationen in Deutschland
Lizenz: Open Source (GPL)
Programmiersprache: Java
Systemumgebung: Linux/Unix bzw. Windows oder Mac OS, Jetty bzw. Tomcat oder Websphere, relationale Datenbank (hsql, MySQL, Oracle, DB2)
Download: http://sourceforge.net/project/showfiles.php?group_id=92005
Support: Mailingliste, Bug-Tracker, Wiki, IRC-Kanal

Fedora

Entwickler: FedoraCommons
Verbreitung: 150 Installationen weltweit, davon ca. 5 in Deutschland
Lizenz: Open Source (Mozilla Public Licence)
Programmiersprache: Java
Systemumgebung: Linux/Unix, Apache,
Download: http://sourceforge.net/project/showfiles.php?group_id=177054
Support: Bug-Tracker, Wiki, Mailinglisten, Forum

Handle und OpenURL. Außerdem können Creative-Commons-Lizenzen vergeben werden, und es erlaubt nutzerseitig das Browsing über einzelne Sammlungen sowie eine Suche in den Metadaten.

Bei Installation, Anpassung und Wartung des Systems werden Java-Kenntnisse benötigt. Die Einrichtung gestaltet sich etwas komplexer als beispielsweise im Falle von OPUS. Aus technologischer Sicht ist das System wegen der zugrunde liegenden Java-Technologien dafür zukunftsträchtiger, da bei der Programmierung und Erweiterung strukturierter vorgegangen werden kann. DSpace stellt ein ausgereiftes Nutzermanagement bereit, bei dem der Publikationsworkflow rollenbasiert gesteuert wird. Das System arbeitet bezüglich Autorisierung, Urheber- und Verwertungsrechten und Zugriff mit nutzergruppenspezifischen Anpassungen. Es ist skalierbar, gut für größere Forschungseinrichtungen nutzbar und natürlich einsetzbar, um den Grünen Weg zu Open Access zu unterstützen.

EPrints

EPrints ist eine Software der University of Southampton aus dem Jahr 2000 und wird durch das dortige Entwicklerteam auch weiterhin gewartet und herausgegeben. Weltweit existieren fast 300 produktive Installationen des Systems. Im Gegensatz zu den anderen Produkten ist EPrints für die durch Wissenschaftler vollzogenen Eigenpublikationen – das so genannte Self-Archiving – optimiert, wird jedoch auch zum Betrieb von IRs im hier verwendeten Sinne eingesetzt. EPrints unterstützt den Publikationsprozess für Autoren, Herausgeber und Institutionen. Es verfügt über ein dafür ausgelegtes Rechtemanagement. Das System stellt eine OAI-Schnittstelle bereit und erlaubt standardmäßig den Export von Metadaten, unter anderem in den Formaten Dublin Core, MODS, METS, BiBTeX und OpenURL. Eine URN-Unterstützung und eine Alerting-Funktion, die über die Bereitstellung neuer Dokumente im Repository informiert, sind ebenfalls implementiert. Die History-Funktion ermöglicht Administratoren das Verfolgen von Änderungen an Dokumenten und Metadaten. Um EPrints installieren zu können,

werden keine Programmierkenntnisse benötigt. Sie werden erst für Erweiterungen oder tief greifende Anpassungen erforderlich. Grundsätzlich sind mehrere Archive unter einer Installation lauffähig, das heißt, ein Hosting ist prinzipiell möglich. Die Standarddistribution ist für den englischsprachigen Betrieb optimiert und muss daher bei der Installation für deutschsprachige Nutzer entsprechend angepasst werden. Anpassungen sind auch bei der OAI-Schnittstelle notwendig, um den Export von Metadaten in den Formaten EPICUR und XMetaDiss zu ermöglichen.

MyCoRe

MyCoRe, kurz für *My Content Repository*, wurde im Rahmen des Projekts MILESS an der Universität Essen (heute Universität Duisburg-Essen) entwickelt und steht inzwischen als Open-Source-Software zur Verfügung, die durch ein hochschulübergreifendes Konsortium erweitert und gepflegt wird. MyCoRe bildet dabei einen Rahmen, der mehrere Softwaremodule und ein übergreifendes Management-System integriert, mit dem digitale Sammlungen unterschiedlichster Art verwaltet werden können. Das System stellt alle wesentlichen Grundfunktionen eines digitalen Repository und zur Unterstützung des Publikationsprozesses bereit – angefangen von einer Uploadschnittstelle, über die Sacherschließung einschließlich Klassifikation bis hin zur individuellen Anpassung der Oberfläche. Anpassungen für einfache Anwendungen können auf der Konfigurationsebene vorgenommen werden und bedingen keinen Eingriff in den Programmcode. Besonders hervorzuheben ist die Möglichkeit, das Datenmodell vollständig konfigurieren zu können, hierbei kommen XML-basierte Konfigurationsdateien zum Einsatz. Dadurch ist das System nicht auf die Verwaltung bestimmter Publikationsarten beschränkt. Außerdem stellt MyCoRe wichtige Schnittstellen und Standards wie URN, OAI, XMetaDiss, epicur, Z39.50, MAB2 und SOAP bereit. Es unterstützt Video-Streaming ebenso wie eine Suche in den Metadaten und den Volltexten – unter anderem auch für verteilte MyCoRe-Installationen. Das System ist für den deutschen Sprachraum konzipiert und kann damit ohne sprachliche Anpassungen unmittelbar eingesetzt werden. Für den produktiven Einsatz von MyCoRe mit einer langfristigen Perspektive sollten personelle Ressourcen mit Java-Kompetenzen vorhanden sein.

Fedora

Die ursprünglich an der University of Virginia und der Cornell University entwickelte Repository-Software *Fedora* (kurz für *Flexible Extensible Digital Object and Repository Architecture*) stellt im Gegensatz zu den anderen vorgestellten Systemen kein unmittelbar einsetzbares und in sich abgeschlossenes Komplettsystem bereit. Fedora bildet eher eine Basisarchitektur, die über ein flexibles Framework Dienste bereitstellt, auf deren Grundlage ein Repository implementiert werden kann. Das Softwarekonzept folgt dem Ansatz der Serviceorientierten Architektur (SOA), die sich durch gute Modularisierbarkeit und hohe Flexibilität auszeichnet. Das System ist in Java implementiert, die einzelnen digitalen Objekte werden intern auf der Basis von RDF (Resource Description Framework) abgebildet und beschrieben. Für Import und Export von Objekten stehen das Metadatenframework METS sowie das selbst entwickelte Format *Fedora Object XML* (FOXML) zur Verfügung. Im Sinne der bereits erwähnten SOA verfügt Fedora lediglich über einige zentrale Webservices, die über das Protokoll SOAP angesprochen und genutzt werden können.

Mit diesen Schnittstellen stellt die Kernsoftware damit nur die Basisdienste zur Verfügung, die notwendig sind, um auf die digitalen Objekte zuzugreifen und wenige Funktionen daran auszuführen. Alle anderen Komponenten, insbesondere auch die eigentliche Benutzerschnittstelle, müssen auf dieser Basis selbst implementiert werden. Andererseits lässt sich Fedora dafür wie keines der anderen Systeme in bestehende Umgebungen und Anwendungen integrieren.

Im Rahmen des Projekts *eSciDoc* arbeitet die *Max Planck Digital Library* (MPDL) daran, ein vollständiges IR-System auf der Basis von Fedora zu entwickeln und es anschließend als nachnutzbares Produkt zur Verfügung zu stellen.

Wer die Wahl hat ...

Alle hier vorgestellten Systeme genügen – ggf. nach entsprechender Konfiguration – den im DINI-Zertifikat formulierten technologischen Anforderungen. Sie kommen also potentiell alle als Grundlage für ein Institutional Repository in Betracht. Entscheidend für die Auswahl eines der Systeme sollten neben den funktionalen Unterschieden der einzelnen Lösungen vor allem die lokalen Gegebenheiten und Anforderungen sein. Dies betrifft nicht nur die Flexibilität und Anpassbarkeit der einzelnen Produkte – beispielsweise in Bezug auf das zugrunde liegende Datenmodell und das Layout der Benutzeroberflächen – sondern insbesondere auch die an der betreffenden Einrichtung vorhandenen Kompetenzen und Systemumgebungen.

6. Schritt: Etablieren von Workflows

Auf den ersten Blick scheint es vielleicht so, dass Institutional Repositories, einmal vorbereitet und eingerichtet, sozusagen ohne weiteren Aufwand arbeiten bzw. einfach durch Autoren und Herausgeber gefüllt werden, die Veröffentlichungen in der entsprechenden inhaltlichen und technischen Qualität über die Uploadschnittstelle einreichen. Doch in der Regel sind Institutional Repositories keine reinen Selbstbedienungssysteme. Die einzelnen Veröffentlichungsvorgänge werden zwar sozusagen von außen angestoßen, müssen aber in die bibliothekarischen Geschäftsgänge integriert werden. Genau hierin besteht eine der zentralen Herausforderungen beim Aufbau eines IR. Dazu sollten bestehende Arbeitsabläufe vergegenwärtigt und explizit beschrieben, neue Anforderungen spezifiziert und daraus resultierende Geschäftsprozesse definiert und dokumentiert werden.

Dabei ist die Online-Bereitstellung von Zweitveröffentlichungen im Sinne des Grünen Weges zu Open Access noch der einfachste Fall: Hier erfolgt seitens des Betreibers für hochgeladene Publikationen allenfalls eine Plausibilitätsprüfung und ggf. die Korrektur von Metadaten, bevor sie freigeschaltet und somit

öffentlich verfügbar gemacht werden. Mehrteilige Publikationsvorhaben, bei denen als Gegenüber des IR ein Herausgeber in Erscheinung tritt, bedingen oftmals weit komplexere Geschäftsgänge – beispielsweise durch die Unterstützung des gesamten Dokumenterstellungsprozesses einschließlich der Erzeugung eines formatierten Satzspiegels bzw. eines XML-basierten Dateiformats als Grundlage für einen medienneutralen Publikationsansatz. Und bei der Veröffentlichung von Hochschulschriften spielen die gültigen Prüfungsordnungen mit ihren Publikationsbestimmungen in die Workflows mit hinein, sodass möglicherweise die Ausstellung entsprechender Bescheinigungen und die Herstellung gedruckter Exemplare mit in den Aufgabenbereich des Publikationsservice fallen.

Um eine möglichst hohe Qualität der bibliografischen Daten zu erreichen, ist die Erfassung der Metadaten oder zumindest deren Nachbearbeitung durch bibliothekarisches Personal empfehlenswert. Dies gilt insbesondere dann, wenn die Daten in andere Systeme exportiert werden sollen – beispielsweise in den lokalen OPAC oder in einen Verbundkatalog. Für die Sacherschließung müssen – so sieht es auch das DINI-Zertifikat vor – entsprechende Richtlinien vorhanden sein, in denen beschrieben ist, welche Erschließungsinstrumente verwendet werden und welche Regelungen dabei gelten.

Grundsätzlich sollten Autorenverträge geschlossen werden – je nach erforderter Rechtsverbindlichkeit in schriftlicher Form, wie etwa bei Dissertationen, oder zumindest durch entsprechend zu bestätigende Klauseln im Online-Formular, beispielsweise bei Pre- und Postprints.

Überaus sinnvoll erscheint die Integration von Geschäftsprozessen rund um das IR in die Erstellung einer Institutsbibliografie, wie sie an vielen wissenschaftlichen Einrichtungen gepflegt wird. IR und Jahresbibliografie haben grundsätzlich denselben Gegenstandsbereich – Publikationen von Angehörigen der jeweiligen Institutionen (siehe z. B. [9]). Durch die Integration beider Systeme können nicht nur Synergieeffekte genutzt werden, sie kann auch die Motivation von Wissenschaftlern erhöhen, ihre Veröffentlichungen in das IR einzustellen – beispielsweise dann, wenn die Anzahl eigener Veröffentlichungen für die leistungsbezogene Mittelvergabe herangezogen und diese Kennzahl anhand der Institutsbibliografie ermittelt wird.

Für komplexe Geschäftsgänge, die unterschiedliche Benutzerrollen und Aktivitäten beinhalten – wie beispielsweise inhaltliche Begutachtungsprozesse – ist die Verwendung von Workflowsystemen, die die Prozesse elektronisch unterstützen, empfehlenswert.

7. Schritt: Kundenbetreuung

Neben dem technischen Betrieb des Institutional Repository muss auch die Betreuung der Autoren und Herausgeber – also der Kunden des IR – bewerkstelligt werden. Dazu zählt beispielsweise, Fragen zu Dokumentformaten zu beantworten und bei rechtlichen Fragen – vor allem im Zusammenhang mit dem Urheberrecht und bestehenden Autorenverträgen mit Dritten – auf entsprechende Informationsangebote zu verweisen bzw. die Informationen selbst zu beschaffen – beispielsweise mithilfe der SHERPA/RoMEO-Liste[3], in der Open-Access-Regelungen zahlreicher Wissenschaftsverlage hinterlegt sind (siehe dazu [10]).

Das DINI-Zertifikat erwartet als Mindestanforderung im Bereich der Autorenbetreuung öffentlich verfügbare Kontaktinformationen und ein Beratungsangebot – beispielsweise per Online-Formular, FAQ-System, E-Mail- oder Telefon-Hotline. Es liegt auf der Hand, dass der Betreuungsbedarf mit komplexer werdenden Dienstleistungen gegenüber Herausgebern und Autoren steigt. Werden beispielsweise Dokumentvorlagen für Texterstellungssysteme zur Verfügung gestellt, um bei komplexeren Publikationsvorhaben die Erstellung eines einheitlichen Satzspiegels bzw. eines XML-basierten Dokumentworkflows vorzubereiten, sind auch dafür entsprechende Beschreibungen bzw. Unterstützungsangebote vorzuhalten, z. B. in Form regelmäßig durchgeführter Kurse.

Besondere Beachtung gebührt Herausgebern von Schriftenreihen, Sammelbänden, Zeitschriften und dergleichen, die ihre Publikationsvorhaben über ein IR veröffentlichen wollen. Hierbei gestaltet sich die Betreuung als kooperativer Beratungsprozess, der unter anderem die Berücksichtigung individueller Anforderungen – beispielsweise an Funktionalität oder Layout – zum Inhalt hat (siehe dazu auch [11]).

8. Schritt: Öffentlichkeitsarbeit und Vernetzung

Wenn die Grundsatzfragen geklärt, Leitlinien erarbeitet, Ressourcen und Personal organisiert sind, ein System installiert ist, Workflows definiert und Betreuungsangebote geschaffen sind, fehlen nur noch – die Autoren. Dass institutionelle Veröffentlichungsplattformen keine Selbstläufer sind, zeigen die relativ geringen Publikationszahlen in nahezu allen vorhandenen IRs. Gründe dafür sind der zusätzliche – wenn auch geringe – Aufwand, den Autoren für die Einstellung von Zweitpublikationen in Repositories aufbringen müssen und der demgegenüber für sie nicht erkennbare Mehrwert sowie andererseits die geringe Akzeptanz gegenüber Open-Access-Publikationen unter Wissenschaftlern (siehe [12]).

Neben der Bereitstellung brauchbarer Mehrwertdienste für die potentielle Leserschaft einerseits und der Unterstützung von Autoren und der Minimierung des erforderlichen Aufwands andererseits (siehe 7. Schritt) lässt sich diesem weit verbreiteten Problem nur mit der Schaffung externer (finanzieller) Anreize und wirkungsvollem Marketing begegnen. Während finanzielle Anreize für die Nutzung von IRs nicht ohne verbindliche und durch die jeweiligen Führungsebenen zu erlassenden Regelungen möglich sind – vor allem Rahmen leistungsbezogener Mittelvergabe – liegt die Öffentlichkeitsarbeit eher in der Hand der Betreiber von IRs. Dabei sollte keine Möglichkeit ausgelassen werden, potentielle Autoren nicht nur auf das bestehende Publikationsangebot, sondern auch auf die Vorzüge dieser Veröffentlichungsform hinzuweisen – angefangen von einer möglichst prominenten Verlinkung des IR von den

3 siehe http://www.sherpa.ac.uk/romeo/ [09.02.2009]

zentralen Internetseiten der Einrichtung, über Aushänge, Informationen per E-Mail oder Hauspost, Artikel und Anzeigen in internen Publikationen bis hin zu Informationsveranstaltungen. Als sinnvoll erweist es sich dabei, auf bestehende Kommunikations- und Organisationsstrukturen zurückzugreifen und potentielle Autoren und Herausgeber darüber anzusprechen – etwa über die Fachreferenten der Bibliotheken oder durch die gezielte Ansprache von Wissenschaftlern durch den Besuch von deren Arbeitsgruppenbesprechungen und Kolloquien. Wirkungsvolle Öffentlichkeitsarbeit ist ein personalintensives Arbeitsfeld und idealerweise eng mit der – sich im Erfolgsfall anschließenden – Kundenbetreuung verzahnt.

Letztlich geht es darum, mit dem IR für die Institution und deren Mitglieder identitätsstiftend zu wirken und zu erreichen, dass Wissenschaftler und Studierende das Repository als ihre Publikationsplattform begreifen – und zwar zum einen in der Rolle potentieller Autoren, aber auch als lesende Nutzer. Insofern ist die Erhöhung der Bekanntheit nicht nur zur Gewinnung von Autoren, sondern auch für die allgemeine Akzeptanz des IR als zentraler Bestandteil institutioneller Informationsversorgung zu verstehen.

Wichtig sind parallel dazu die Vernetzung des IR innerhalb der nationalen und internationalen Community und die Einspeisung der Metadaten in übergreifende – möglicherweise fachbezogene – Recherchedatenbanken. Wenn das IR die durch das DINI-Zertifikat beschriebenen Anforderungen erfüllt, sind dafür die technischen Voraussetzungen im Sinne der Interoperabilität mit entsprechenden Serviceanbietern bereits gegeben. Damit die bibliografischen Daten tatsächlich Eingang in die einschlägigen Nachweissysteme finden und die auf dem IR vorgehaltenen Publikationen über das eigene Webangebot hinaus sichtbar und wahrgenommen werden, ist die Registrierung des IR und der OAI-Schnittstelle bei zentralen Diensten und in der offiziellen Registratur der *Open Archives Initiative* erforderlich[4]. Die meisten OAI-basierten Service Provider übernehmen regelmäßig die dort vorhandenen Informationen über vorhandene IRs, um dann von ihnen die verfügbaren Metadaten einzusammeln. Darüber hinaus ist aber auch die Anmeldung des IR bei weiteren Verzeichnissen wie dem bereits genannten ROAR und OpenDOAR[5] sinnvoll.

Nicht zuletzt sollte nach Abschluss aller Arbeiten das DINI-Zertifikat auch formal beantragt werden. Einmal erlangt, kann es als positiver Marketingeffekt auf potentielle Autoren und Herausgeber wirken. Es ist beispielsweise auch die Voraussetzung, um mit voller Funktionalität in das nationale Portal von OA-Netzwerk eingebunden zu werden (siehe [4]) und damit auch die Tür zu dessen internationalem Pendant – zu DRIVER[6].

Fazit

Institutional Repositories gehören zur Informationsversorgung in wissenschaftlichen Einrichtungen längst zum Alltag. Damit sie durch deren wesentliche Zielgruppe – die jeweiligen Wissenschaftler in der Rolle als Autor, Herausgeber und Leser – auch tatsächlich akzeptiert und selbstverständlich genutzt werden, müssen noch viele Hürden genommen werden. Eine wesentliche Voraussetzung ist ein überlegtes und strukturiertes Vorgehen bei Konzeption und Aufbau von IRs und der Planung der Dienstleistungen und Prozesse. Dabei muss das sprichwörtliche Rad heute nirgendwo mehr neu erfunden werden – weder in Bezug auf die benötigte Technologie noch hinsichtlich der strategischen Überlegungen und der notwendigen organisatorischen Maßnahmen. Dennoch sollten die Besonderheiten der jeweiligen Einrichtungen und die daraus resultierenden spezifischen Anforderungen bei der konkreten Umsetzung nicht unberücksichtigt bleiben. Die Orientierung an den Bedürfnissen der potentiellen Nutzer kann ein wesentlicher Erfolgsfaktor für das gesamte IR sein.

4 siehe http://www.openarchives.org/Register/BrowseSites [09.02.2009].

5 siehe http://www.opendoar.org/ [09.02.2009]

6 siehe http://www.driver-repository.eu/ [09.02.2009]

Literatur

[1] Schirmbacher, P.; Müller, U.: *Der „Grüne Weg zu Open Access" – Aktuelle Entwicklungen in Deutschland*. cms-journal 32, 2009, S. 41–46.

[2] DINI e.V.: *DINI-Zertifikat für Dokumenten- und Publikationsservices 2007*, DINI-Schriften 3, 2007, urn:nbn:de:kobv: *11-10079197* [09.02.2009].

[3] Dobratz, S.: *Hostingservice für andere Dokumentenserver*. cms-journal 32, 2009, S. 86–89.

[4] Malitz, R.: *Open Access –Verfügbar ist nicht präsent*. cms-journal 32, 2009, S. 70–73.

[5] Open Society Institute: *OSI Guide to Institutional Repository Software, Version 3*, 2004, http://www. soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf [Stand 02.07.2007].

[6] Wyles, R.: *Technical Evaluation of Research Repositories*. Open Access Repositories in New Zealand. 2006, https://eduforge.org/docman/view.php/131/1062/Repository%20Evaluation%20Document.pdf [09.02.2009].

[7] Dobratz, S.: *Open-Source-Software zur Realisierung von Institutionellen Repositorien – Überblick*. Zeitschrift für Bibliothekswesen und Bibliografie 54 (4-5). 2007, S. 199-206.

[8] Roggenbuck, R.; Schmitz,: *Vortrag auf dem DINI-WS am 19.02.2009* in Göttingen.

[9] Abel, G.; Dressler, B.; Herb, U.; Müller, M.: *Verzahnung von Jahresbibliographie und Institutional Repository der Universität des Saarlandes*. Bibliotheksdienst 40 (6), 2006, S. 726-733.

[10] Zielke, D.: *Rechtliche Probleme bei Postpublikationen und die Ergebnisse des DFG-Projekts „Open Access Policies"*. cms-journal 32, 2009, S. 60–62.

[11] Müller, U.: *Auf dem goldenen Weg zum edoc-Server – Open-Access-Zeitschriften, ...* cms-journal 32, 2009, S. 29–32.

[12] Over, A.; Maiworm, F.; Schelewsky, A.: *Ergebnisse einer Umfrage zum Publikations- und Rezeptionsverhalten unter besonderer Berücksichtigung von Open Access*. Studie der Deutschen Forschungsgemeinschaft. Weinheim: Wiley-VCH, 2005.