

Repositoriensoftware

Einleitung

Bereits in den 1990er Jahren haben sich an den meisten Hochschulen Schriftenserver etabliert, auf denen anfänglich Dissertationen veröffentlicht wurden. Diese wurden oftmals zu institutionellen Repositorien erweitert, um Parallelveröffentlichungen von Forschenden zu unterstützen.¹ Durch die stetige technologische Weiterentwicklung der internationalen Repositorienlandschaft sind inzwischen diverse technische Lösungen für das Betreiben von Dokumentenservern vorhanden. Ein Überblick über die Anforderungen, Standards, derzeitige Systeme und zukünftige Ideen wird gegeben.

Anforderung an die Software

Für Forschungspublikationen sind spezielle Anforderungen an ein Publikationssystem und damit an die Software zu stellen. Hauptaugenmerk bei wissenschaftlichen Arbeiten ist Zitierbarkeit. Diese benötigt als Grundlagen die Langzeitverfügbarkeit sowie eine eindeutige Referenz der Inhalte. Für die Adressierung der Inhalte können für die Dokumente sog. Persistent Identifier² vergeben werden, welche sich bei einer Änderung der Internetadresse nicht verändern. Beispiele sind der Unified Resource Name (URN)³, welcher von der deutschen Nationalbibliothek vergeben wird, sowie der Digital Object Identifier (doi)⁴, der ursprünglich im Verlagswesen eingesetzt wurde.

Langzeitarchivierung ist ein sehr komplexes Thema und kann sicherlich nicht von Repositoriensoftware abgedeckt werden. Hier ist die Zusammenarbeit mit einem Archivierungssystem, das internationale Normen erfüllt, wünschenswert und sinnvoll.⁵

¹ Vgl. Horstkemper 2016.

² Siehe Schroeder 2010.

³ Ausführliche Information zur URN-Service der DNB findet sich unter http://www.dnb.de/DE/Netzpublikationen/URNService/urnservice_node.html.

⁴ Siehe hierzu die Seite der International DOI Foundation (IDF), siehe <https://www.doi.org/>.

⁵ Vgl. DINI-Arbeitsgruppe „Elektronisches Publizieren“ 2016.

Neben der Verfügbarkeit sind Verfahren zur Sicherstellung der Authentizität (Sicherstellung der Echtheit der Dokumente) sowie der Integrität (Korrektheit der Daten) unerlässlich. Um dies gewährleisten zu können, müssen Vorkehrungen zur IT-Sicherheit getroffen werden. Dies beinhaltet neben der allgemeinen Serversicherheit eine Verschlüsselung der Datenübertragung, die Authentifizierung der Autoren und die Vergabe von speziellen Rechten für Benutzergruppen.

Für die Akzeptanz eines Repositoriums ist die Usability ein entscheidender Faktor. Diese muss für verschiedene Nutzergruppen einzeln umgesetzt werden. Lesern sind Tools zur Verfügung zu stellen, welche eine Nachnutzung der Daten ermöglichen, beispielsweise den Export für Literaturverwaltungsprogramme oder das Teilen in sozialen Netzwerken. Autoren benötigen eine einfache Möglichkeit zu publizieren und für den Metadatenimport aus Literaturverwaltungsprogrammen oder kommerziellen Anbietern. Ein einfaches redaktionelles Bearbeiten von Einträgen ist ebenso notwendig.

Durch die Internationalisierung von Forschung und Lehre wird die Bereitschaft zu Multilingualität unumgänglich. Webseiten und Nutzerschnittstellen müssen neben der lokalen Sprache zudem in englischer Sprache angeboten werden. Die Software soll als Zeichenkodierung Unicode unterstützen.

Interoperabilität

Ein entscheidendes Kriterium für die Sichtbarkeit und damit die Nutzung des Repositoriums ist die Möglichkeit der Interoperabilität mit anderen Systemen. Wichtigste Schnittstelle ist das OAI-Interface. Durch ein Protokoll zum Auslesen von Metadaten (Open Archives Initiatives - Protocol Metadata Harvesting⁶) wird exakt definiert, wie Metadaten zur Verfügung gestellt werden und somit von zentralen Diensten, wie etwa der Bielefeld Academic Search Engine (BASE)⁷ ausgewertet und weiterverarbeitet werden können. Als Austauschformat dient dabei XML. Das XML-Datenformat ist nicht fest vorgegeben, sondern wird mit angefordert. Dieses kann von den Standardformaten Dublin Core⁸ und METS⁹ auch in Richtung spezieller Strukturen abweichen.

⁶ Die Spezifikation der Schnittstelle findet sich bei der Open Archives Initiative unter <https://www.openarchives.org/OAI/openarchivesprotocol.html>. Eine Überprüfung des Interfaces ist unter <http://validator.oaipmh.com/> möglich.

⁷ Siehe <https://www.base-search.net/>.

⁸ Siehe <http://dublincore.org/>.

⁹ Der Metadata Encoding & Transmission Standard (METS) wird von der Library of Congress gepflegt, siehe <https://www.loc.gov/standards/mets/>.

Die Pflichtablieferung bei der Deutschen Nationalbibliothek und die Registrierung von URNs erfolgt ebenfalls über die OAI-Schnittstelle in den Formaten XMetaDissPlus¹⁰ und xepicur¹¹.

Neben dem OAI-Interface, bei welchem eine Datenlieferung beschrieben wird, ist noch die Simple Web-service Offering Repository Deposit (SWORD)-Schnittstelle¹² zu anzuführen. Damit können Anbieter von Publikationen Dokumente in das System einbringen.

Für die Verbreitung der Information ist es wünschenswert, ein SOAP-Interface¹³ anzubieten. Dadurch kann gezielt im Repository gesucht und die Ergebnisse können in Webseiten, wie zum Beispiel Personen-, Lehrstuhl- oder Projektseiten, eingebunden werden.

Daneben sind externe Metadatenschemata zu berücksichtigen. Ein Austausch von Daten ist oftmals nur über diese möglich. Dies sind beispielsweise Vorgaben von Literaturverwaltungsprogrammen und von Publikationsplattformen, wie zum Beispiel Research Papers in Economics (RePEc)¹⁴.

Für die Benutzerverwaltung sind ebenfalls Verbindungen zu implementieren. Zu unterscheiden ist dabei, ob Daten aus lokalen Systemen angeboten werden sollen, beispielsweise über eine LDAP¹⁵ Authentifizierung oder über einen weltweiten Dienst, wie das Open Researcher and Contributor ID (ORCID)¹⁶ Konsortium. Nur durch derartige Vernetzung können Publikationen eindeutig den jeweiligen Autoren zugeordnet werden.

¹⁰ Siehe Koordinierungsstelle DissOnline 2010.

¹¹ Siehe <http://www.persistent-identifier.de/xepicur/version1.0/xepicur.xsd>.

¹² Siehe <http://swordapp.org/>.

¹³ Siehe <https://www.w3.org/TR/soap/>.

¹⁴ Die Definition des Formats ReDIF findet man unter <https://ideas.repec.org/p/rpc/rdfdoc/redif.html>.

¹⁵ Lightweight Directory Access Protocol. Eine Beschreibung findet sich unter https://de.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol

¹⁶ Siehe <http://orcid.org/>. Für Deutschland ist eine flächendeckende Umsetzung durch das Projekt ORCID DE (<http://www.orcid-de.org/>) und dem damit verbunden ORCID Deutschland Konsortium angestrebt

Die COAR Roadmap Future Directions for Repository Interoperability¹⁷ gibt einen Überblick über den derzeitigen Stand und zukünftige Desiderate im Themengebiet Interoperabilität.

Nicht unerwähnt bleiben soll die Optimierung für Suchmaschinen. Gerade bei dem Rechenweg der einfachen Internetsuche ist die Reihenfolge der Auflistung in den Suchergebnissen enorm wichtig.

Entscheidungshilfe

Neben den oben genannten Standards, welche eine Software für Repositorien anbieten muss, sind weitere Überlegungen zur lokalen Situation sinnvoll. Grundsätzlich muss die Software in der Außenansicht flexibel sein, so dass das Repository in das Corporate Design der Einrichtung eingebettet werden kann. Zusätzlich müssen die modellierten Geschäftsprozesse und eine Darstellung der damit verbundenen Rechteverwaltung für verschiedene Nutzergruppen auf die Software übertragbar sein. Vorhandene Systeme, wie etwa eine zentrale Benutzerverwaltung, müssen mit der Software verknüpfbar sein. Die Datenstruktur, welche die einzelnen Dokumente beschreibt, sollte flexibel und um lokale Daten erweiterbar sein. Dabei ist zu berücksichtigen, wie das Softwaresystem administrierbar und konfigurierbar ist. Wünschenswert ist, dass eine Konfigurationsänderung ohne technische Kenntnisse durchgeführt werden kann, gerade wenn das Repository nicht von einer technischen Einrichtung betreut wird.

Bereits im Vorfeld ist eine Einschätzung über die zukünftige Größe des Systems notwendig. Die Skalierbarkeit der Software spielt eine große Rolle, damit diese der zukünftigen Entwicklung des Repositoriums gewachsen ist.

Für die Einsatzmöglichkeit eines Systems ist die Unterstützung der Entwickler bzw. der Anwender entscheidend. In erster Linie ist das Lizenzmodell, unter dem die Software nutzbar ist, zu nennen. Eine ausführliche Dokumentation sowie die Bereitstellung von Kommunikationsplattformen (Wiki, Mailinglisten etc.) sind für die Einrichtung eines Repositoriums sicherlich ein immenser zeitlicher Vorteil.

¹⁷ Siehe https://www.coar-repositories.org/files/Roadmap_final_formatted_20150203.pdf

Hier ist darauf zu achten, wie viele Installationen der Software vorhanden sind und inwieweit die Benutzer untereinander kommunizieren.

DINI-Zertifikat

Die Deutsche Initiative für Netzwerkinformation e. V. (DINI) formuliert Mindestanforderungen und Empfehlungen für den Betrieb von Repositorien.

Bei Einhaltung eines Kriterienkatalogs kann eine Einrichtung das sog. DINI-Zertifikat¹⁸ erwerben. Dabei entwickeln sich die Anforderungen stetig weiter.

Seit 2013 können sich Hostingservices eine Bestätigung für die DINI-Kompatibilität der technischen Realisierung ausstellen lassen. Diese Services werden durch „DINI ready“¹⁹ markiert.

Die technischen Richtlinien im DINI-Zertifikat, Version 5.0, Oktober 2016 sind

- Kriterium 5: Informationssicherheit. Hierzu zählen Ausfallsicherheit, Betriebssicherheit und Vertrauenswürdigkeit sowie Verfügbarkeit, Integrität und Authentizität der veröffentlichten Dokumente. Spezifiziert sind diese in den sog. Common Criteria, die als internationaler Standard ISO/IEC 15408 veröffentlicht worden sind.
- Kriterium 6: Erschließung und Schnittstellen. Dies beinhaltet die beschreibenden Metadaten sowie deren Bereitstellung zur maschinellen Weiterverarbeitung. So wird beispielsweise die Beschreibung des Dokumenttyps nach dem gemeinsamen Vokabular für Publikations- und Dokumenttypen sowie die Bereitstellung der Metadaten mittels einer OAI-Schnittstelle gefordert.
- Kriterium 7: Zugriffsstatistik. Hierbei wird gefordert, dass Aufrufe von Dokumenten im Rahmen von gesetzlichen Vorschriften des Datenschutzes gespeichert und bestenfalls sogar angezeigt werden.
- Kriterium 8: Langzeitverfügbarkeit. Das DINI-Zertifikat grenzt sich von technischen Langzeitdiensten ab. Empfohlen wird, mit einem zertifizierten Dienst zusammen zu arbeiten und langzeitarchivierbare Dateiformate zu verwenden.

Open Access Repository Ranking

¹⁸ Siehe DINI-Arbeitsgruppe „Elektronisches Publizieren“ 2016.

¹⁹ Siehe <https://dini.de/dini-zertifikat/dini-ready/>.

Im Jahre 2014 startete das Open Access Repository Ranking²⁰. Die zugrunde liegende Metrik bewertet die Benutzbarkeit und technische Implementierung in den vier Kategorien Bedienbarkeit, Dienstleistungen, Metadaten und Interoperabilität. Sie wird in einem öffentlichen Evaluierungsverfahren stetig verbessert.

Eigenentwicklung vs. Softwarepaket

Bei dem Betrieb eines Repositoriums stellt sich grundsätzlich die Frage, ob man die Software selbst entwickelt oder bereits vorhandene Software nach nutzt. Es folgt ein kurzer Überblick, welche Chancen, aber auch welche Risiken sich für die jeweilige Entscheidung ergeben.

Bei der Eigenentwicklung entsteht eine maßgeschneiderte Software, welche sich ideal an die lokalen Gegebenheiten anpasst. Ferner hat man die Kontrolle über die Entwicklung des Systems. Zu berücksichtigen ist aber, dass die Entwicklung eines Systems gerade personell sehr aufwändig ist und dauerhafte Stellen für die Langzeitbetreuung und den Support zur Verfügung gestellt werden müssen. Bei der Software MediaTUM²¹, welche an der Technischen Universität München entwickelt wurde, handelt es sich bspw. um eine Eigenentwicklung.

Mit einem wesentlich geringeren Personalbedarf ist beim Einsatz von Softwarepaketen für Repositorien zu planen. Eine fertige Lösung wird meist mit einer freien Lizenz angeboten. Viele Standardfunktionen sind darin bereits integriert. Somit ist ein schnelles Setup möglich, allerdings bleibt zu berücksichtigen, dass lokale Anpassungen durchzuführen sind. Diese sind zudem bei den regulären Updates zu leisten. Zu beachten ist, dass gerade lokale Einstellungen, beispielsweise die Ablieferung an die Deutsche Nationalbibliothek im Format XmetaDissPlus, nicht in der Standardinstallation implementiert sind. Oftmals gibt es aber dafür schon Lösungen, welche in der jeweiligen Nutzergemeinschaft bereitwillig zur Verfügung gestellt werden. Auf die zukünftige Entwicklung hat man allerdings wenig Einfluss, oftmals ist es aber möglich, eigene Erweiterungen in zukünftige Versionen des Quellcodes einfließen zu lassen.

Softwarepakete

²⁰ Siehe <http://repositoryranking.org/>. Ausgangspunkt war der Census on Open Access Repositories in Germany, Austria and Switzerland, siehe Vierkant 2014.

²¹ Siehe <https://mediatum.github.io/>.

Die am häufigsten eingesetzten Softwarepakete sollen kurz vorgestellt werden. Eine ausführliche Liste findet sich im Open Access Directory.²² Für weiterführende Informationen wird aus Aktualitätsgründen auf die Webseiten der Projekte verwiesen. Einen Überblick über den Einsatz der verschiedenen Softwarepakete findet man unter The Directory of Open Access Repositories – OpenDOAR.²³

DSpace: <http://www.dspace.org/>

DSpace wurde am Massachusetts Institute of Technology (MIT) und den HP Labs, der Forschungsabteilung von Hewlett-Packard, entwickelt. Durch die DSpace Foundation wird die Weiterentwicklung in einem kooperativen Modell gesteuert. DSpace wird von der Non-Profit-Organisation Duraspace verwaltet. Diverse Dienstanbieter bieten eine kommerzielle Weiterentwicklung des Codes an.²⁴

EPrints: <http://www.eprints.org>

Das Softwarepaket EPrints wird an der Universität Southampton entwickelt. Derzeit bietet EPrints spezielle Lösungen für Dokumenten- und Datenrepositorien sowie offene Lernplattformen an. Erweiterbar ist die Software mit speziellen Plugins, welche von der Nutzergemeinde zur Verfügung gestellt werden.²⁵

Fedora: <http://fedorarepository.org/>

Ursprünglich ist Fedora an der University of Virginia und der Cornell University entstanden. Die Entwicklung leistet die Fedora Leadership Group unter der Betreuung von Duraspace. Fedora zeichnet sich gerade durch eine sehr hohe Flexibilität aus. Auf Fedora basieren weitere Repositoriensysteme. Die bekanntesten davon sind Hydra²⁶ und Islandora²⁷.

²² Siehe http://oad.simmons.edu/oadwiki/Free_and_open-source_repository_software. Vgl. hierzu auch Castagné 2013 und Dobratz 2007.

²³ Siehe <http://www.opendoar.org/>.

²⁴ Eine Liste mit registrierten Dienstleistern findet sich unter <http://www.dspace.org/service-providers>.

²⁵ Dies geschieht im sog. EPrints Bazaar, siehe <http://bazaar.eprints.org/>.

²⁶ Die Projektseite findet sich unter <https://projecthydra.org/>.

²⁷ Weitere Information findet man unter <http://islandora.ca/>.

Invenio: <http://invenio-software.org/>

Invenio wurde ursprünglich am CERN für den CERN Document Server entwickelt. Inzwischen wird es an zahlreichen weiteren Institutionen verwendet und weiterentwickelt. Invenio unterstützt unterschiedlichste Datenformate und zeichnet sich dadurch aus, Repositorien mit einer sehr großen Anzahl an Dokumenten betreiben zu können. Als Metadatenschema wird MARC21 verwendet.

Mediatum: <https://mediatum.github.io/>

Mediatum wird an der Technischen Universität (TU) München entwickelt. Es eignet sich neben Textdateien zur Veröffentlichung von Bild-, Multimedia und Forschungsdaten. Über die TU München hinaus hat es nur eine geringe Verbreitung.

MyCore: <http://www.mycore.de>

MyCore wurde an der Universität Duisburg-Essen entwickelt. Die Geschäftsstelle befindet sich derzeit am Regionalen Rechenzentrum der Universität Hamburg. MyCore besitzt ein frei konfigurierbares Metadatenschema und ist damit sehr anpassungsfähig und über ein klassisches Repository hinaus verwendbar.

OPUS 4: <http://www.kobv.de/entwicklung/software/opus-4/>

OPUS 4 wird am Kooperativen Bibliotheksverbund Berlin-Brandenburg (KOBV) entwickelt und ist Nachfolger von OPUS, das ursprünglich an der Universitätsbibliothek Stuttgart entwickelt wurde. Es ist das Paket, bei dem die spezifischen deutschen Anforderungen URN-Vergabe und Meldung an die DNB nativ implementiert sind.

Hostingservices

Statt eine Publikationsplattform bzw. ein Repository selbst zu betreiben, bieten Dienstleister an, Repositorien zur Verfügung zu stellen. Man kommt direkt eine funktionierende Plattform zur Verfügung gestellt. Oftmals ist eine große Funktionalität gegeben. Probleme bei den Upgrades der Software werden vom zentralen Provider behoben. Der personelle Aufwand für das Repository wird deutlich verringert.

Nachteil dieser Lösung sind die Kosten, welche hier anfallen, sowie der geringe Einfluss auf Veränderungen und eine erschwerte Anpassung an die lokalen Verhältnisse.

Ausblick

Repositorien haben sich inzwischen für das wissenschaftliche Publizieren etabliert. Allerdings ist das wissenschaftliche Publizieren im Wandel, und die Anforderungen an derartige Systeme und damit an die Software werden in nächster Zukunft steigen. In der Confederation of Open Access Repositories (COAR)²⁸ wurde zu diesem Zweck eine Arbeitsgruppe Next Generation Repositories gegründet, welche auf den Gebieten Interoperabilität, Bewertung, Workflows und Auswirkung Desiderate postuliert, beispielsweise die Realisierung eines Open Peer Review, um eine Möglichkeit zu etablieren, Beiträge qualitativ zu bewerten.

Durch die zunehmende Verbreitung von mobilen Geräten wird eine entsprechende optimierte Darstellung zukünftig eine wichtige Rolle spielen. Eine Unterstützung der Repositoriensoftware, sowohl ein mobiles Layout zu integrieren als auch mobile Dokumenttypen wie EPUB²⁹ zu unterstützen, wird deshalb unabdingbar. Verstärkt werden in der Wissenschaft multimediale Inhalte veröffentlicht. Die Software muss diese Formate unterstützen. Ebenso wird man sich über die Bereitstellung und Zitierbarkeit von veränderbaren Dokumenten Gedanken machen müssen.

Ein wichtiger Punkt wird aber die Integration der Repositorien in die Infrastruktur der Einrichtung und des Publikationsmarktes sein. So sind sowohl Schnittstellen zu Verlagen für das automatische Publizieren als auch für das Anreichern von Information im Repository unabdingbar, und es muss sichergestellt sein, dass Forschungsinformationssysteme diese Information nachnutzen können.

Literatur

- Castagné, Michel: Institutional repository software comparison. DSpace, EPrints, Digital Commons, Islandora and Hydra. 2013. <http://hdl.handle.net/2429/44812> (17.10.2016).
- DINI-Arbeitsgruppe „Elektronisches Publizieren“: DINI-Zertifikat für Open-Access-Repositorien und -Publikationsdienste 2016. Göttingen 2016. urn:nbn:de:kobv:11-100239432
- Dobratz, Susanne: Open-Source-Software zur Realisierung von Institutionellen Repositorien - Überblick. In: Zeitschrift für Bibliothekswesen und Bibliographie 54 (2007), H. 4/5, S. 199–206.
- Horstkemper G. (2015). Elektronisches Publizieren I: Publikationsserver. In R. Griebel, H. Schäffler & K. Söllner (Hrsg.) Praxishandbuch Bibliotheksmanagement (S. 619-629). Berlin: De Gruyter. Doi: 10.1515/9783110303261.631
- Koordinierungsstelle DissOnline: XMetaDissPlus Format des Metadatensatzes der Deutschen Nationalbibliothek für Online-Hochschulschriften inklusive Angaben zum Autor

²⁸ Siehe <https://www.coar-repositories.org/>.

²⁹ Siehe <http://idpf.org/epub>.

(XMetaPers). Version 2.0. Deutsche Nationalbibliothek, Leipzig 2010. urn:nbn:de:101-2010052704.

Schroeder K. (2010). Persistent Identifier (PI) – ein Überblick. In: H. Neuroth, A. Oswald, S. Strathmann & K. Huth (Hrsg.) nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.3 (S. Kap9.22 – Kap9.45). urn:nbn:de:0008-2010071949

Vierkant, Paul: 2012 Census of Open Access Repositories in Germany: Turning Perceived Knowledge Into Sound Understanding. In: D-Lib Magazine 19 (2013), H. 11/12. DOI:10.1045/november2013-vierkant.