

DOI: <http://dx.doi.org/10.11588/ip.2015.1.18627>

Alexandra S. LINHART

Eine Untersuchung der Sichtbarkeit von Open-Access Abschlussarbeiten auf deutschen institutionellen Repositorien

Zusammenfassung

Sichtbarkeit ist ein zentrales Argument für Open Access. In dieser Untersuchung wurde die Sichtbarkeit von Abschlussarbeiten unter Dissertationsniveau, die auf institutionellen Repositorien deutscher Hochschulen veröffentlicht wurden, in verschiedenen Suchdiensten betrachtet. Hierzu wurden zufällig ausgewählte Dokumente mit vier verschiedenen Suchanfragen gesucht, die sowohl eine Known-Item-Suche als auch einen thematischen Sucheinstieg simulieren. Die Nachweissituation der Dokumente ist gut, kann aber noch verbessert werden. Bei einem thematischen Sucheinstieg jedoch stellt sich die Sichtbarkeit dieser Publikationen als schlecht heraus.

Schlüsselwörter

Hochschulschrift, Open Access Repository, Sichtbarkeit

Testing the visibility of Open Access theses on German institutional repositories

Abstract

Visibility is a main argument for open access. This paper looks into the visibility of bachelor and master theses published on institutional repositories of German universities. Using four queries simulating known-item search and a topical search, randomly selected publications were searched in four search engines and catalogs. The indexing rates of these publications were acceptable but can still be improved. Using a topical searching approach however, visibility of the published theses declines.

Keywords

Thesis, open access repository, visibility

Dieser Beitrag wurde im Open Peer Review begutachtet. Die Preprint-Version sowie die dazu eingegangenen Kommentare finden sich unter:
<http://informationspraxis.de/2015/04/09/open-peer-review-linhart-sichtbarkeit-von-open-access-hochschulschriften-eine-untersuchung-deutscher-institutioneller-repositorien/>

Inhaltsverzeichnis

1 Einleitung	2
2 Verwandte Arbeiten	3
3 Sichtbarkeitskriterien	4
4 Untersuchungsdesign	5
4.1 Überblick und Stichprobenziehung	5
4.2 Auswahl der Suchdienste	6
4.3 Untersuchung der Sichtbarkeit	6
4.4 Suchanfrageformulierung	7
5 Ergebnisse	8
5.1 Nachweis und Sichtbarkeit nach Ebenen	8
5.2 Known-Item Suche	9
5.3 Thematische Suche	9
6 Diskussion	9
6.1 Diskussion der Ergebnisse	9
6.2 Diskussion der Methode	10
7 Fazit	11
8 Quellen	11
Autorin	12

1 Einleitung

Sichtbarkeit wird als zentraler Vorteil des Open Access angeführt. Für die Umsetzung des Open Access haben sich in den letzten Jahren einige Publikationswege herauskristallisiert. Zwar sprechen sich die Open Access Policies deutscher Universitäten häufig für das Publizieren auf dem „Goldenen Weg“ aus, mindestens jedoch solle man den „Grünen Weg“, das Publizieren auf dem oft hochschuleigenen Open Access Publikationsserver, in Betracht ziehen (Linhart 2013).

Es gibt in ganz Deutschland inzwischen eine Vielzahl solcher von Hochschulen betriebene Repositorien für das elektronische Publizieren von wissenschaftlichen Arbeiten und Hochschulschriften, angefangen von der Bachelor-Thesis, bis hin zu Doktorarbeiten oder Zeitschriftenbeiträgen. Hochschulen und Hochschulbibliotheken betreiben hierfür Server, die sich manchmal auf einen fachlichen Kontext begrenzen oder für die Wissenschaftler und Studierenden der eigenen Institution die Möglichkeit bieten, die ganze Bandbreite ihrer Forschung frei zu veröffentlichen.

Wie jedoch steht es um die Sichtbarkeit der Publikationen auf diesen Servern? Lassen sich diese Arbeiten wirklich so einfach finden, wie Open Access es verspricht? In dieser Untersuchung wird zunächst ein Überblick über die bisherigen Untersuchungen zur Sichtbarkeit von Open Access Publikationen gewonnen, um so dann die Landschaft deutscher Hochschulpublikationsserver darzustellen. Viele Untersuchungen befassen sich bisher mit dem Impact und der Sichtbarkeit von OA-Journals oder Zeitschriftenartikeln auf Repositorien. Qualifikationsarbeiten sind bisher nicht in den Blickpunkt gerückt. Traditionell müssen erst

Dissertationen veröffentlicht werden. Auf institutionellen Repositorien ist es aber auch üblich, unterhalb des Dissertationsniveaus bereits Arbeiten zu erfassen. Auch sie können durch ihre oft sehr spezielle Ausrichtung interessant für die Forscher oder Studierende sein. Anhand bestimmter Kriterien wurde stichprobenhaft die Sichtbarkeit von Hochschulschriften unter Dissertationsniveau auf institutionellen Repositorien untersucht. Das Vorgehen dieser Untersuchung nebst den Ergebnissen wird dargelegt.

2 Verwandte Arbeiten

Obwohl Sichtbarkeit ein häufig genanntes Argument für Open Access ist, gibt es wenige Untersuchungen, die sich genau damit befassen. Oft wird eine erhöhte Sichtbarkeit als gegeben vorausgesetzt, sobald das Dokument nur irgendwie über das Internet erreichbar ist.

In der Bibliometrie wird in diesem Zusammenhang in erster Linie gefragt, ob Open Access die Zitationsrate oder den Impact erhöht, was zweifellos im Interesse der Autoren von Open Access Artikeln ist. Ein positiver Zusammenhang wäre für die Befürworter des Open Access ein starkes Argument. Die Frage ist daher politisch aufgeladen. Es gibt eine große Anzahl von Studien, die jedoch zu widersprüchlichen Aussagen kommen. Gaulé/Maystre kamen zu dem Schluss, dass es keinen kausalen Zusammenhang zwischen einer Veröffentlichung im Open Access und einer erhöhten Zitationsrate gibt (Gaulé/Maystre 2011). Björk/Solomon attestieren nach ihrer Untersuchung den im Web of Science verzeichneten Journals immerhin eine gleiche Qualität und Chance auf Impact wie nicht Open Access Journals (Björk/Solomon 2012). Davis konnte in seiner Untersuchung feststellen, dass Open Access Artikel zwar häufiger heruntergeladen werden als nicht freie, jedoch nicht häufiger oder schneller zitiert werden. Er schließt daraus, dass Open Access es erlaubt, ein breiteres Publikum zu erreichen. Da Zitate jedoch meist von Forschern an Institutionen kommen, die ihnen Zugriff auf nicht Open Access Veröffentlichungen bieten, ergibt sich kein Unterschied bei der Zitationsrate (Davis 2011). Diese Untersuchungen befassen sich jedoch in erster Linie mit Artikeln in Open Access Journals und nicht mit Publikationen auf Repositorien.

Um überhaupt gefunden werden zu können, müssen Publikationen gut nachgewiesen sein. Mit der Aggregation von Metadaten von Qualifikationsarbeiten für das „Current Research Information System (CRIS)“ haben sich Schöpfel et.al. beschäftigt. Sie untersuchten, wie und welche Metadaten zu Qualifikationsarbeiten in verschiedenen Ländern für Repositorien zusammengetragen werden (Schöpfel et al. 2014).

Offhaus hat sich mit der Qualität von IRs in Nordrhein-Westfalen befasst, richtet das Augenmerk aber auf allgemeinere Anforderungen. Sichtbarkeit ist nur eines von vielen Kriterien und wird als erfüllt betrachtet, wenn das OAI-PMH Protokoll verwendet wird und die Leitlinien der DINI eingehalten wurden (Offhaus 2012).

Das Projekt „Aufbau eines Netzwerks zertifizierter Open-Access-Repositorien“ erkannte mangelnde Vernetzung der IRs als Problem für die Sichtbarkeit der dort zur Verfügung gestellten Publikationen und versucht einen zentralen Datenbestand aufzubauen (DINI 2014b). Insgesamt ist die von IR-Betreibern erstellte Metadatenqualität bisher aber nicht ausreichend, um einen zufriedenstellenden Austausch von Inhalten zwischen IRs und Suchdiensten zu ermöglichen (Kindling/Vierkant 2013). Ein ähnliches Ziel verfolgte DRIVER im europäischen Kontext. Das Projekt mündete im OpenAIRE Portal, das unter anderem Repositorien-Netzwerke verknüpft und zentral durchsuchbar macht (OpenAIRE 2014).

Die Performance von allgemeinen und wissenschaftlichen Suchmaschinen bei der Suche nach wissenschaftlichen Dokumenten, die in Open Access erschienen sind, haben Pieper/Wolf untersucht. Neben Google und Yahoo untersuchten sie auch spezielle wissenschaftliche und OA-Suchmaschinen. Mithilfe von Retrievaltests stellten sie fest, dass die Abdeckung und der

Zugang zum Volltext gerade bei allgemeinen Suchmaschinen noch nicht zufriedenstellend sind. Auch gehen der wissenschaftliche Treffer in der Menge der Ergebnisse unter. Wissenschaftliche Suchmaschinen erfassen die wissenschaftlichen Dokumente meist deutlich besser (Pieper/Wolf 2009).

Henneberger hat in ihrer Dissertation eine Methode für IRs entwickelt, die es erlaube, anhand der Analyse von Nutzungsdaten die Sichtbarkeit von Dokumenten eines IR zu überwachen (Henneberger 2011). Die Methode befasst sich jedoch zur Überwachung von Sichtbarkeit nur mit der Ermittlung von Downloads, die nicht durch Robots erfolgen.

Die Sichtbarkeit von Dokumenten auf IRs haben sich Arlitsch/O'Brien im Speziellen angesehen. Sie stellten fest, dass die Indexierungsrate der Dokumente vieler IRs in den USA in Google Scholar sehr schlecht war und erzielten eine Verbesserung durch die Nutzung der von Google empfohlenen Metadaten schemata (Arlitsch/O'Brien 2012).

Dieser Überblick zeigt, dass die Argumentation der meisten Untersuchungen mit der Datenqualität oder dem bloßen Nachweis einer Publikation erfolgt. Auch eine explizite Untersuchung der Sichtbarkeit speziell der Hochschulschriften unter Dissertationsniveau hat anscheinend bisher niemand geleistet. Im Folgenden wird die hierzu erfolgte Untersuchung beschrieben.

3 Sichtbarkeitskriterien

Es muss aufgrund der Literaturlage davon ausgegangen werden, dass die Sichtbarkeit von Open Access Hochschulschriften auf institutionellen Repositorien (IR) unzureichend ist.

Hypothese 1: Der Nachweis von Hochschulschriften auf IRs in allgemeinen und wissenschaftlichen Suchmaschinen und Katalogen ist schlecht.

Hypothese 2: Die Sichtbarkeit der Hochschulschriften auf IRs bei einem thematischen Sucheinstieg in allgemeinen und wissenschaftlichen Suchmaschinen ist schlecht.

Um die Sichtbarkeit der Publikationen auf IRs zu untersuchen, müssen Kriterien für Sichtbarkeit gefunden werden. Das Stichwort „Sichtbarkeit“ fasst in der Informations- und Bibliothekswissenschaft indes eine ganze Bandbreite an Bedeutungen. In den Anforderungen für das DINI-Zertifikat 2013 wird von den Betreibern eines IR in erster Linie verlangt, dass sie das IR bei diversen Registrierungsstellen eintragen lassen und ansonsten auch innerhalb ihrer Institution dafür sorgen, dass der Dienst bekannter wird (DINI 2013). Als „Auffindbarkeit“ aus der Sicht der Nutzergruppe definiert Thiessen seinen Ansatz bei der Untersuchung der Sichtbarkeit von Monographien im Open Access (Thiessen 2013). Beim Ranking Web of Repositories wird unter dem Punkt „Visibility“ auf die Zahl der Backlinks gesetzt, die ein IR im Internet besitzt (CSIC 2014). In der Suchmaschinenoptimierung gibt es das Maß des Sichtbarkeitsindex für eine Webseite für eine Suchanfrage, der aus der Beliebtheit der Suchanfrage und dem Ranking der Webseite errechnet wird (Onpage 2014).

In dieser Untersuchung wurde von folgenden, daran anschließenden Überlegungen ausgegangen:

1. Ein Dokument muss indexiert sein, um gefunden werden zu können. Dies ist die Ebene des Nachweises, wie von der DINI gefordert.
2. Ein Dokument muss so indexiert sein, dass es auffindbar ist, wenn man recht genau weiß, was man sucht, also für eine sogenannte Known-Item-Suche.
3. Ein Dokument muss so indexiert sein, dass es auch auffindbar ist, ohne genaue Titeldaten zu kennen. Es muss also rein thematisch gefunden werden können und deshalb entsprechend nachgewiesen sein.

4. Ein Dokument muss dort auffindbar sein, wo ein Nutzer üblicherweise danach suchen würde. Die Suchoption muss dem Nutzer bekannt sein, bzw. häufig genutzt werden. Eine Datenbank, die keiner kennt, ist so gut wie keine Datenbank. Je beliebter also ein Suchdienst ist, desto größer ist sein Potential für die Sichtbarkeit eines Dokuments. Für die Betreiber von Dokumentenservern ist daher nicht nur die breite Verteilung der Metadaten von entscheidender Bedeutung, sondern auch die Wahrscheinlichkeit, dass die Suchdienste von Kunden genutzt werden. Dies sollte daher in einem Maß der Sichtbarkeit berücksichtigt werden.

Im Folgenden wird das Forschungsdesign vorgestellt, um die Sichtbarkeit der Qualifikationsarbeiten auf IRs zu untersuchen.

4 Untersuchungsdesign

4.1 Überblick und Stichprobenziehung

Die Deutsche Initiative für Netzwerkinformation (DINI) führt eine Liste der in Deutschland betriebenen Repositorien für die Veröffentlichung von Open Access Publikationen (DINI 2014a). Um einen Überblick über die Landschaft der deutschen institutionellen Repositorien (IR) zu gewinnen, die Hochschulschriften unter Dissertationsniveau verzeichnen, wurde diese Liste herangezogen. Die DINI-Liste hat keinen Anspruch auf Vollständigkeit.

Zum Zeitpunkt der Untersuchung im Juni 2014 verzeichnete die Liste 184 Repositorien (Stand April 2014). Aus dieser Liste wurden für die Untersuchung alle Repositorien ausgeschlossen, die

1. eine fachliche Beschränkung aufweisen, da sie meisten nicht allein institutionell ausgerichtet sind.
2. nicht von einer Hochschule oder Universität bzw. einer dort angeschlossenen Bibliothek betrieben werden, denn hier ist es unwahrscheinlich Qualifikationsarbeiten zu entdecken
3. keine Arbeiten unter Dissertationsniveau (Bachelor-, Master-, Diplomarbeiten oder Studienarbeiten) verzeichnen, bzw. auch nicht deutlich machen, dass solche Arbeiten prinzipiell erwünscht sind.

Nach Beendigung dieser Sortierung blieben noch 91 IRs übrig, die diesen Kriterien entsprechen. Aus diesen 91 IRs wurde eine zufällige Stichprobe von 10 für die nähere Untersuchung gezogen. Eine vollständige Liste der untersuchten IRs findet sich unter 5.

Die Stichprobe enthielt IRs mit einer Menge von 51 bis 33.550 verzeichneten Dokumenten. Der Anteil der Hochschulschriften unter Dissertationsniveau lag daran zwischen 0,05 und 100%. Über alle Dokumente der Stichprobe lag der Anteil der Publikationen lediglich bei 1,5%. Dies zeigt, dass die Stichprobe eine große Vielfalt unterschiedlicher Repositorien abdeckt. Es gibt sowohl sehr kleine, die hauptsächlich Abschlussarbeiten erfassen, als auch große, die nur in geringem Maße Publikationen unter Dissertationsniveau aufgenommen haben. Insgesamt ist der Anteil dieser Art von Publikationen gering.

Die ausgewählten Repositorien wurden zum Zeitpunkt der Untersuchung alle mit der Software OPUS betrieben. Die verwendete Software sollte daher kaum Einfluss auf das Ergebnis genommen haben.

Um zu einer Stichprobe von Dokumenten zu kommen, deren Sichtbarkeit exemplarisch getestet werden soll, wurde wiederum eine zufällige Stichprobe aus den Listen der in dem Server verzeichneten Qualifikationsarbeiten gezogen. Die Liste der ausgewählten Dokumente findet sich unter den angehängten Forschungsdaten. Um den Umfang der Untersuchung zu begrenzen wurden pro IR vier Dokumente ausgewählt.

4.2 Auswahl der Suchdienste

Es konnten im Rahmen dieser Untersuchung nicht alle möglichen Suchdienste für elektronische Dokumente oder wissenschaftliche Dokumente untersucht werden. Es macht jedoch Sinn, die bekanntesten und am häufigsten genutzten zu wählen, da die Sichtbarkeit maßgeblich von der Wahrscheinlichkeit der Nutzung abhängt. Bei einer guten Nachweissituation und Platzierung in Suchdiensten, die kaum Nutzer haben, kann eigentlich nicht von einer guten Sichtbarkeit gesprochen werden.

Es gibt leider keine Marktanalyse über die Nutzung wissenschaftlicher Suchmaschinen und -dienste. Es muss daher von den Ansprüchen ausgegangen werden, die die Anbieter an sich selbst stellen. Außerdem ist eine Art Kategorisierung der Suchdienste sinnvoll. So ist es zunächst wichtig für die IR-Anbieter, auf ihrer institutionellen Ebene Sichtbarkeit zu erlangen: dies geht vor allem über die lokalen Kataloge der Bibliothek. Ein weiterer Aspekt ist die Sichtbarkeit in den auf OA-Dokumente spezialisierten Suchmaschinen. Deutschlandweit am größten ist hier die BASE (Bielefeld Academic Search Engine).

Auf der nächsten Ebene interessiert die weltweite wissenschaftliche Suche. Hier ist Google Scholar zu nennen. In den weiteren Untersuchungen aus der Literatur waren auch Scirus und ScientificCommons genannt (Pieper/Wolf 2009). Beide sind jedoch nicht mehr am Markt und können daher nicht berücksichtigt werden.

Auf der Ebene der allgemeinen Websuche wiederum gibt es gute Marktanalysen. Hier ist auf dem deutschen Markt in erster Linie Google relevant (SEO-United 2014).

Hieraus ergibt sich also die Liste, der vier untersuchten Suchdienste:

Auf lokaler Ebene wurden die jeweiligen Bibliothekskataloge untersucht, auf der OA-spezifischen Suche BASE für den deutschen Kontext, auf der wissenschaftlichen weltweiten Suche Google Scholar und auf der Ebene der allgemeinen Websuche Google.

4.3 Untersuchung der Sichtbarkeit

In dieser Untersuchung geht es nicht um einen Vergleich der Suchmaschinen wie bei Pieper/Wolf. Ihre Vorgehensweise kann jedoch in bestimmten Teilen übernommen werden. So wurde auch in diesem Forschungsdesign ein Punktesystem erstellt, die Punkte jedoch an die Dokumente auf den IRs vergeben.

Um einen Vergleich der Sichtbarkeit der Dokumente auf den IRs zu erhalten, wurde ein Wert gebildet, der neben der Art der Suche, auch die Reichweite des Suchdienstes gewichtet.

Um sicherzustellen, dass nicht zufällig eine unglückliche Suchanfrage gewählt wurde, wurden für die Known-Item-Suche und für die thematische Suche jeweils zwei unterschiedliche Anfragen gewertet. Die Ergebnislisten wurden bis Seite 5 ausgewertet. Das sind die ersten 50 Ergebnisse, was schon weit über dem durchschnittlichen Nutzer von allgemeinen Suchmaschinen liegt, der meist nicht mehr als 5 Ergebnisse betrachtet (Lewandowski 2008). Ferner wurden reine Metadaten Seiten oder dasselbe Dokument von einem anderen Host nicht mit in die Wertung genommen. Auch reine Zitationsnachweise in Google Scholar wurden nicht als Treffer gewertet, lediglich das Volltext PDF.

Wurde nach der Durchführung der vier Suchanfragen das Dokument nicht gefunden, wurde davon ausgegangen, dass es auch nicht in dem entsprechenden Suchdienst nachgewiesen ist. Eine separate Suche erfolgte hierfür nicht mehr. Insbesondere wenn die Known-Item-Suchen 0 Treffer erreichten, war davon auszugehen, dass auch ein genaueres Überprüfen zu keinem Ergebnis geführt hätte.

Natürlich muss für die Durchführung des Tests möglichst alles ausgeschaltet werden, was die

Ausgabe der Suchmaschine beeinflussen könnte (Nutzerprofile, lokale Suche, Cookies, Suchhistorie soweit möglich).

Aus diesen Ergebnissen wurde dann ein Sichtbarkeitswert ermittelt. Für den reinen Nachweis gab es 1 Punkt pro Dokument, für Erfolg bei der Known-Item-Suche gab es 2 Punkte und für Erfolg bei der thematischen Suche gab es 3 Punkte.

Diese Punktzahlen pro Suchdienst wurden dann noch gewichtet, je nachdem wie groß die (avisierter) Nutzerschaft des Suchdienstes ist, um ein Maß für die Reichweite zu schaffen. Demnach ist die Sichtbarkeit im eigenen OPAC weniger stark gewichtet, wie die Sichtbarkeit in Google Scholar, da die (avisierter) Reichweite als höher angenommen werden kann. Tab. 1 zeigt die Faktoren für die Gewichtung der Punkte in jedem Suchdienst.

Lokaler Dienst	BASE	G Scholar	Google
1	2	3	4

Tabelle 1: Gewichtungsfaktoren der Suchdienste

Diese Art die Sichtbarkeit von Dokumenten auf den IRs zu vergleichen erlaubt es, sowohl die Art der Suche zu berücksichtigen, als auch die Reichweite des Suchdienstes. Da davon auszugehen ist, dass es bei der thematischen Suche mehr potenziell relevante Treffer gibt, als bei der Known-Item-Suche, wird diese Art der Suche in dem Wert stärker gewichtet, als die Known-Item-Suche. Die maximal mögliche Punktzahl pro Dokument liegt also bei 110 Punkten.

4.4 Suchanfrageformulierung

Um einen möglichst realistischen Test durchzuführen und nicht zu unfair gegenüber den IR-Betreibern zu werden, wurde bei der Formulierung der bei dem Test genutzten Suchanfragen folgendermaßen vorgegangen:

Pro Dokument wurden jeweils 4 Anfragen an jeden Suchdienst gestellt - zwei für jede Suchart. In der ersten Anfrage der Known-Item-Suche wurden zwei Wörter verwendet, bei denen das erste jeweils den Nachnamen des Autors und ein markantes Titelstichwort enthielt. Es wurden bei der Known-Item-Suche nur zwei Wörter gewählt, da die durchschnittliche Anzahl der Suchwörter in der deutschen allgemeinen Websuche bei 1,7 Wörtern liegt (Lewandowski 2008). Die zweite Suchanfrage der Known-Item-Suche enthielt zwei Titelstichwörter.

Die Suchanfragen für die thematische Suche wurden aus vergebenen Schlagworten für die Dokumente und/oder etwas unspezifischeren Titelstichwörtern gebildet. Die erste Anfrage war noch etwas spezifischer formuliert als die zweite Anfrage, die allgemeinere und größere Themenkreise abfragte. In Ausnahmefällen, wenn eine viel zu große Treffermenge erwartet werden konnte (wie z.B. bei „U-Bahn“ und „Verkehr“) wurde ein drittes Stichwort (in diesem Fall „Berlin“) zugelassen, um fair zu bleiben. Die vollständige Liste der verwendeten Suchanfragen findet sich bei den Forschungsdaten zu diesem Artikel.

5 Ergebnisse

Die Nachweissituation der Dokumente war in Ordnung, ist aber noch ausbaufähig. Es ließ sich eine insgesamt eher schlechte Sichtbarkeit nachweisen, jedoch nicht für alle untersuchten Ebenen gleichermaßen.

Akkumuliert man die Punkte der einzelnen Dokumente nach Herkunftserver und errechnet dann den Anteil an der erreichbaren Gesamtpunktzahl (440), zeigt sich, dass es ein breites Mittelfeld gibt, wo die Sichtbarkeit der Dokumente sich kaum unterschied. Tabelle 2 zeigt die erreichten Prozentwerte über alle Dokumente und den Menge verzeichneter Dokumente auf dem IR insgesamt.

Hochschule	Anteil an Punkten	Umfang des IR
Universität Marburg	73 %	7.304 Dok.
Universität Tübingen	46 %	6.374 Dok.
Fachhochschule Dortmund	43,1 %	51 Dok.
Universität Frankfurt	42,7 %	33.550 Dok.
Universität Würzburg	42 %	7.394 Dok.
TU Berlin	41,8 %	4.175 Dok.
Hochschule Hannover	41,8 %	362 Dok.
Fachhochschule Aachen	36,8 %	6.008 Dok.
Hochschule Heilbronn	32,7 %	64 Dok.
Hochschule Esslingen	14,5 %	56 Dok.

Tabelle 2: Ergebnisse nach Herkunftserver

5.1 Nachweis und Sichtbarkeit nach Ebenen

Es gab lediglich zwei Dokumente, die keinen Treffer nach den aufgestellten Kriterien erzielen konnten (Dokumente Nr. 18 und 20), beide vom IR der Hochschule Esslingen.

Auf lokaler Ebene hatte die Fachhochschule Dortmund keines der untersuchten Dokumente über den lokalen Katalog zugänglich gemacht. Gleiches gilt für die Dokumente der Hochschule Esslingen und der Hochschule Heilbronn. Nur 45,6% aller Suchanfragen erzielten den gesuchten Treffer in den lokalen Katalogen. 27 der 40 Dokumente wurden über den lokalen Katalog überhaupt gefunden.

Bei der Suche in der Open Access Suchmaschine BASE konnten keines der Dokumente aus den IRs der Fachhochschule Aachen, und der Hochschule Esslingen gefunden werden. 51,3% der Suchanfragen waren in BASE erfolgreich. 30 von 40 Dokumenten konnten überhaupt gefunden werden.

Die Suche in Google Scholar verlief insgesamt eher enttäuschend. Nur 17 der 40 Dokumente wurden überhaupt aufgefunden. Die Erfolgsquote bei den Suchanfragen lag bei 27,5%. Gar kein Dokument (lediglich einige Zitationen) konnte von den IRs der TU Berlin, der Hochschule Heilbronn und der Hochschule Esslingen gefunden werden.

Auf der Ebene der allgemeinen Suche, erzielten 44,37% der Suchanfragen auf den ersten 5 Seiten der Ergebnisliste den gesuchten Treffer. 30 von 40 Dokumenten wurden überhaupt aufgefunden.

5.2 Known-Item Suche

	Lokal	BASE	G Scholar	Google
Erfolgsquote	51,3%	61,3%	38,8%	52,5%
Gefundene Dok.	26	28	17	27

Tabelle 3: Ergebnisse der Known-Item Suche

Die Known-Item-Suche war erfolgreicher, als die thematische Suche. Über alle Suchanfragen gesehen war jedoch gerade die Hälfte (50,9 %) überhaupt erfolgreich. Es fällt auf, dass die Erfolgsquote in Google Scholar deutlich geringer ist, als in den anderen Suchdiensten (Tab. 3).

5.3 Thematische Suche

In der thematischen Suche war ebenfalls die Erfolgsquote in Google Scholar die geringste (Tab. 4). Insgesamt war nicht einmal die Hälfte der Anfragen erfolgreich (42,2%). Meist wurde nur die Hälfte der Dokumente überhaupt gefunden.

	Lokal	BASE	G Scholar	Google
Erfolgsquote	36,3%	42,5%	16,3%	37,5%
Gefundene Dok.	20	22	11	22

Tabelle 4: Ergebnisse der thematischen Suche

6 Diskussion

6.1 Diskussion der Ergebnisse

Der durchgeführte Test der Sichtbarkeit ist nur eine Stichprobe aus einer Fülle von Dokumenten, kann jedoch als Indikator für die Situation der Sichtbarkeit der Qualifikationsarbeiten auf IRs in Deutschland gelten.

Die Untersuchung hat gezeigt, dass immerhin bis zu 75% der Dokumente, in den meisten untersuchten Diensten aufgefunden werden könnten. Zu einem gewissen Grad, muss Hypothese 1 hiermit also abgelehnt werden. Jedoch muss negativ angemerkt werden, dass nicht einmal die Hälfte über Google Scholar auffindbar war. Da Google Scholar als eine bekannte wissenschaftliche Suchmaschine für eine hohe Sichtbarkeit der Dokumente sorgen könnte, fällt dies doch negativ ins Gewicht. An dieser Stelle besteht noch großer Optimierungsbedarf für den Nachweis in der allgemeinen wissenschaftlichen Suche.

Hypothese 2 kann bestätigt werden. Sowohl auf den unterschiedlichen Ebenen, auf denen sich Sichtbarkeit erlangen ließe, als auch in den unterschiedlichen Suchtypen, ist der Anteil an erfolgreichen Suchanfragen gering. In Anbetracht der Tatsache, dass die Suchanfragen aus den Schlagworten und Stichworten der von den jeweiligen Institutionen erstellten IR-Metadaten stammten, ist gerade die geringe Ausbeute auf der lokalen Ebene (36,3%) sehr schade. Nimmt man dazu an, dass die ersten 50 Treffer schon großzügig ausgelegt sind, ist das Ergebnis bei der Sichtbarkeit nicht befriedigend.

Weiterhin waren die Titel der Suchtreffer in der allgemeinen Suche bei Google oft nicht der Titel der Arbeit, sondern enthielten den Dateinamen (z.B. „Dokument1.pdf“). Daneben zeigt sich, dass Google dazu tendierte, die Namen der Uploader als Autoren in der allgemeinen Suche anzugeben, statt die Namen der Autoren der Arbeiten. Auch dies könnte durch eine Optimierung seitens der IR-Betreiber geändert werden.

Im Falle von Google Scholar muss dahingehend relativiert werden, dass Treffer nur dann als solche gezählt wurden, wenn das PDF direkt verlinkt war. In Google Scholar fanden sich in einigen Fällen auch Zitationshinweise von der DNB oder anderer Stelle, die eventuell einen Hinweis auf die Arbeit geben könnten, jedoch nicht in diese Wertung eingegangen sind. Dennoch sind die Ergebnisse in dieser Untersuchung ein Hinweis darauf, dass die IRs der deutschen Hochschulen ihre Dienste nicht auf eine Indexierung in Google Scholar optimiert haben. Da eine Reihe von Arbeiten in Google Scholar gefunden werden konnte, ist nicht davon

auszugehen, dass Google Scholar Qualifikationsarbeiten unter Dissertationsniveau prinzipiell ausschließt. Hier gibt es das größte Optimierungspotential für die IR-Betreiber.

Der Umfang des Publikationsserver scheint keinen direkten Einfluss auf die Sichtbarkeit der Dokumente zu nehmen. Die Publikationen des Servers der FH Dortmund haben zusammen eine gute Sichtbarkeit erreicht, obwohl der Gesamtumfang des Servers eher gering ist. Gehostet wird das IR vom HBZ Nordrhein-Westfalen. Man könnte daher vermuten, dass die zentrale Betreuung dies erklärt. Jedoch sind auch die Server der Hochschulen Heilbronn, Esslingen und Hannover zentral gehostet vom Bibliotheksservicezentrum Baden-Württemberg. Zwischen den Dokumenten der Universitäten und der Fachhochschulen lag kein großer Abstand. Dennoch erreichten die Dokumente der Universitäten alle die obere Hälfte in diesem Test. Dies wäre einerseits zu erklären durch eine bessere Organisation und Betreuung der IRs, und andererseits evtl. durch einen höheren Trust dieser Institutionen bei Google.

Im Gegensatz zum Fazit von Pieper/Wolf, dass wissenschaftliche Suchmaschinen wissenschaftliche Dokumente besser sichtbar machen, als allgemeine Suchdienste, ist in diesem Fall die Sichtbarkeit der wissenschaftlichen Qualifikationsarbeiten in der untersuchten wissenschaftlichen Suchmaschinen (Google Scholar) am schlechtesten. Dies bestätigt die Ergebnisse von Arlitsch/O'Brien von 2012 auch für deutsche IRs. Am besten ist die Sichtbarkeit in der OA-Suchmaschine BASE. Hier gehen die Treffer auch noch nicht in einer unendlichen Anzahl von Dokumenten unter. Aber auch hier ist eine Nachweisquote von 75% noch ausbaufähig.

6.2 Diskussion der Methode

Zugestanden werden muss, dass diese Untersuchung an einigen Stellen Ungenauigkeiten enthielt, die sich erst im Verlauf der Durchführung zeigten. So werden z.B. alle Suchdienste gleich behandelt. In einigen OPACs wird jedoch nicht automatisch trunkiert, was fehlende Treffer zur Folge haben könnte. Auch könnte es insbesondere bei der Suche in den Google Diensten durch mangelnde Kontrolle über Anfrageinterpretation und den Suchverlauf während der Suche zu verfälschten Ergebnissen gekommen sein. Es wurde versucht dem entgegen zu wirken, indem neben den bereits vor der Untersuchung gespeicherten Daten wie Cookies etc., eine neue IP-Adresse gewählt wurde. Um Verfälschung während der Durchführung vorzubeugen, wurde wenn möglich nicht auf die gesuchten Treffer geklickt (um Google keine Hinweis zu geben, was man genau sucht) und mit der thematischen, allgemeiner formulierten Suche begonnen, statt mit der genaueren Known-Item-Suche.

Die Schwankungen der Treffererfolge unter den Suchdiensten waren bei einigen Suchanfragen gering. Dies ist ein Hinweis darauf, dass die Suchanfragen für den Test schlecht gewählt waren. Andere hatten sehr gute Ergebnisse, vielleicht, weil die Suchanfragen glücklich gewählt wurden. Für weitere Untersuchungen sollte also besser eine noch größere Anzahl unterschiedlicher Anfragen pro Dokument gewählt werden, um solche Glückstreffer besser auszugleichen und ein realistischeres Bild zu erzielen.

Die Methode an sich ist sehr aufwendig, weshalb nur wenige Suchdienste und wenige Dokumente ausgewählt wurden. Dennoch erforderte sie insgesamt 640 Suchanfragen. Im Vergleich mit automatischen Methoden, wie etwa der Backlink-Variante, lässt sie nur eine begrenzte Aussagefähigkeit zu. Für die tatsächliche Sichtbarkeit von Dokumenten ist sie dennoch besser geeignet, als Backlinks zu zählen.

7 Fazit

Der Nachweis der Abschlussarbeiten auf IRs deutscher Hochschulen in Suchdiensten ist nicht sehr gut, aber auch nicht katastrophal, soweit diese Stichprobenuntersuchung es zeigen kann.

Insgesamt ist der Anteil der Abschlussarbeiten an den Dokumenten auf den IRs eher gering, was einen dringenden Handlungsbedarf nicht unterstreicht.

Es gab keinen Hinweis darauf, dass Abschlussarbeiten von Suchdiensten besonders stiefmütterlich behandelt werden. Geht man also davon aus, dass sie Dokumente auf IRs deutscher Hochschulen repräsentieren können, so muss man sagen, dass es um die Sichtbarkeit der Dokumente auf IRs nicht so gut bestellt ist, wie man es nach 10 Jahren Open-Access-Erklärungen in Deutschland erhoffen würde.

Weitere Untersuchungen müssen sich noch eingehender mit der Situation der Dokumente auf IRs befassen. Hierbei sollte jedoch mit deutlich mehr Suchanfragen gearbeitet werden, als in diesem Test, sofern dies durchführbar ist. So kann auch die Aussagefähigkeit der Untersuchung erhöht werden.

8 Quellen

Arlitsch, Kenning; O'Brien, Patrick S. (2012): *Invisible institutional repositories. Addressing the low indexing ratios of IRs in Google Scholar*. In: *Library Hi Tech* Vol. 3 (2012) Nr. 1. S. 60-81.

Björk, Bo-Christer; Solomon, David (2012): *Open Access versus subscription journals: a comparison of scientific impact* [online]. Zugriff am: 24. Juli 2014.

Verfügbar unter: <http://www.biomedcentral.com/1741-7015/10/73>

Consejo Superior de Investigaciones Científicas (CSIC) (2014): *Ranking Web of Repositories. Methodology*. [online] Zugriff am: 27. Juni 2014.

Verfügbar unter: <http://repositories.webometrics.info/en/Methodology>

Davis, Philip M. (2011): *Open access, readership, citations: a randomized controlled trial of scientific journal publishing*. In: *The FASEB Journal Life Sciences Forum* Vol. 25 (2011) Nr. 7, S. 2129-2134.

Deutsche Initiative für Netzwerkinformation (DINI) e.V. (April 2014): *Liste der Repositorien*. [online] Zugriff am: 9. Juni 2014.

Verfügbar unter: <http://dini.de/dini-zertifikat/liste-der-repositorien/>

Deutsche Initiative für Netzwerkinformation (DINI) e.V. (2014): *Netzwerk von Open-Access Repositorien*. [online] Zugriff am: 24. Juli 2014.

Verfügbar unter: <http://www.dini.de/projekte/oa-netzwerk/>

Deutsche Initiative für Netzwerkinformation (DINI) e.V. (2013): *DINI-Zertifikat für Open-Access-Repositorien und Publikationsdienste 2013*. [online] Zugriff am: 9. Juni 2014.

Verfügbar unter: <http://edoc.hu-berlin.de/series/dini-schriften/2013-3/PDF/3.pdf>

Gaulé, Patrick; Maystre, Nicolas (2011): *Getting cited: Does open access help?* In: *Research Policy* Vol. 40 (2011) Nr. 10, S. 1332-1338. [online] Zugriff am: 24. Juli 2014

Verfügbar unter: <http://www.sciencedirect.com/science/article/pii/S0048733311001065>

Henneberger, Sabine (2011) : *Entwicklung einer Analysemethode für Institutional Repositories unter Verwendung von Nutzungsdaten. Dissertation*. Berlin : Humboldt Universität zu Berlin.

Kindling, Maxi; Vierkant, Paul (2013): *Was ist sichtbar? Status Quo und Zukunft der Erschließung von wissenschaftlichen Inhalten in deutschen Open Access Repositorien*. Vortrag auf der 12. InetBib-Tagung am 04.03.2013. [online] Zugriff am: 24. Juli 2014.

Verfügbar unter: <http://de.slideshare.net/MaxiKindling/census-oanvortrag-16948674>

Lewandowski, Dirk (2008): *Search engine user behaviour. How can users be guided to quality content?* In: *Information Services & Use* Vol. 28 (2008), S. 261-268. [online] Zugriff am: 19. August 2014

Verfügbar unter: http://search-studies.org/tl_files/Publikationen_PDFs/2008/ISU2008.pdf

Linhart, Alexandra S. (2013): *Das Reden über freien Zugang. Eine Analyse des Open-Access-Diskurses deutscher Bibliotheken*. [online] Zugriff am: 24. Juli 2014.

Verfügbar unter: <http://edoc.sub.uni-hamburg.de/haw/volltexte/2014/2412/>

Offhaus, Nicole (2012): *Institutionelle Repositorien und Universitätsbibliotheken – Entwicklungsstand und Perspektiven*. In: Institut für Informationswissenschaft: *Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft*. Bd. 63. Köln.

OnPage.org GmbH (2014): *Sichtbarkeitsindex*. In: Onpage Wiki – Digitales Marketing Lexikon. [online] Zugriff am 9. Juni 2014.

Verfügbar unter: <http://de.onpage.org/wiki/Sichtbarkeitsindex>

Open AIRE (2014): *The OpenAIRE project*. [online] Zugriff am: 24. Juli 2014.

Verfügbar unter: <https://www.openaire.eu/general-information/openairefactsheet-40>

Pieper, Dirk; Wolf, Sebastian (2009): *Wissenschaftliche Dokumente in Suchmaschinen*. In: Lewandowski, Dirk (Hrsg.): *Handbuch Internet-Suchmaschinen. Nutzerorientierung in Wissenschaft und Praxis*. Heidelberg: Akademische Verlagsgesellschaft, S. 356-374.

Schöpfel, Joachim; Zendulkova, Danica; Fatemi, Omid (2014): *Electronic theses and dissertations in CRIS*. In: *Procedia Computer Science* Vol. 33 (2014), S. 110-117.

SEO-United (2014): *Suchmaschinenverteilung in Deutschland*. [online] Zugriff am: 8. August 2014.

Verfügbar unter: <http://www.seo-united.de/suchmaschinen.html>

Thiessen, Peter (2013): *Sichtbarkeit von Open-Access- Monographien als Herausforderung – Zur Rolle und Aufgabe von Bibliotheken*. In: *Perspektive Bibliothek* Bd.2 (2013) Nr. 2, S. 4-35.

University of Nottingham (9. Juni 2014): *OpenDOAR: Charts – Usage of Open Access Repository Software – Worldwide*. [online] Zugriff am: 9. Juni 2014.

Verfügbar unter: <http://www.opendoar.org/find.php?format=charts>

Autorin

Alexandra Svantje LINHART, Hochschule für Angewandte Wissenschaften Hamburg,
Finkenau 35, 22081 Hamburg
alexandrasvantje.linhart@haw-hamburg.de