

Ellen Euler *

Web-Harvesting versus Urheberrecht

Was Bibliotheken und Archive dürfen und was nicht

Mit dem Einsatz von Web-Harvesting zur Sammlung und Archivierung von Web-Inhalten sind nicht nur technische, sondern auch rechtliche Probleme verbunden. Der Beitrag setzt sich mit der rechtlichen Bewertung auseinander und reagiert damit auf ein dringendes, aktuelles Bedürfnis öffentlicher Einrichtungen (wie z. B. der Deutschen Nationalbibliothek), zu deren Auftrag es gehört, Web-Inhalte zu sammeln und zu archivieren, und die zur Erfüllung dieses Auftrages Web-Harvesting einsetzen.

I. Technologischer Hintergrund

Unter *Web-Harvesting* ist das automatisierte Sammeln (*harvesting*, engl. das Ernten) von Inhalten aus dem World Wide Web (WWW)¹ zum Zwecke der Archivierung in einem digitalen Archiv zu verstehen.² Zentrales Element des Web-Harvesting ist die Software-Komponente „Web-Crawler“. Ein Web-Crawler ist ein Computerprogramm, das automatisch das Web durchsucht, Webseiten analysiert und die Ergebnisse als Archivdatei in eine definierte Zielumgebung schreibt.

1. Formen des Web-Harvesting

Unterscheiden lassen sich das *zielgerichtete Web-Harvesting* (focused crawl) und das *flächige Web-Harvesting* (broad crawl). Während beim zielgerichteten Web-Harvesting der Web-Crawler nur bestimmte (in einer Liste, der sogenannten URL seed list) vorgegebene Web-Adressen aufsucht und in ein digitales Archiv abspeichert, durchsucht der Web-Crawler beim flächigen Web-Harvesting das Web anhand lediglich formaler Regeln, die als Auswahlgrundlage der zu archivierenden Inhalte dienen.³ Grundlage des flächigen Web-Harvesting ist die Verfolgung von Hyperlinks anhand formulierter Regeln. Eine URL seed list dient hier nur als Einstieg. Sie wird automatisch dynamisch erweitert, indem die auf den zunächst aufgesuchten Webseiten enthaltenen Hyperlinks weiterverfolgt und (samt integrierten Inhalten) auf Regelkonformität überprüft werden.

* Ellen Euler, LL.M., Stipendiatin der Heinrich-Böll-Stiftung e.V., ist Mitarbeiterin am Institut für Informationsrecht bei Prof. Dr. jur. Thomas Dreier in Karlsruhe. Im Rahmen einer Promotion beschäftigt sie sich mit dem, von den Gedächtnisinstitutionen bei Bestandsaufbau, Bestandserhaltung und (kommunikativer sowie kommerzieller) Bestandsvermittlung kultureller Äußerungen über digitale und vernetzte Medien zu beachtendem Rechtsrahmen.

¹ Das World Wide Web ist ein über das Internet abrufbares Hypertext-System und wird häufig synonym zum Begriff Internet verwendet, ist jedoch nur einer von vielen Diensten des Internets.

² Hans Liegmann, Web-Harvesting: Aktivitäten von Nationalbibliotheken, in: Barbara Hoen (Hrsg.), Planungen, Projekte, Perspektiven. Zum Stand der Archivierung elektronischer Unterlagen, Veröffentlichungen des Landesarchivs Nordrhein-Westfalen Bd. 10, Düsseldorf 2006, S. 57.

³ Hans Liegmann, Web-Harvesting: Aktivitäten von Nationalbibliotheken, in: Barbara Hoen (Hrsg.), Planungen, Projekte, Perspektiven. Zum Stand der Archivierung elektronischer Unterlagen, Veröffentlichungen des Landesarchivs Nordrhein-Westfalen Bd. 10, Düsseldorf 2006, S. 57 f.

2. Anwender des Web-Harvesting

Web-Harvesting wird beispielsweise von professionellen Werbemailsendern („Spamern“) angewandt, um Webadressen im Web zu sammeln. Suchmaschinen nutzen Web-Harvesting, für den Service des „Content Caching“,⁴ bei dem Momentanaufnahmen von einzelnen Webseiten gemacht und Nutzern für einen kurzen Zeitraum (maximal drei Wochen) über einen „Cache-Link“ zur Verfügung gestellt werden. Auch Bibliotheken und Archive, zu deren Sammelauftrag auch Web-Inhalte zählen, nutzen das Web-Harvesting. So nutzt beispielsweise die *Deutsche Nationalbibliothek (DNB)*, deren Sammelauftrag mit dem Gesetz über die Deutsche Nationalbibliothek (DNBG)⁵ auf die Sammlung von „Netzpublikationen“, also solche Publikationen, die ohne körperlichen Träger kommuniziert werden,⁶ erweitert worden ist, das Web-Harvesting. Ohne Web-Harvesting könnte die *DNB* ihren erweiterten Sammelauftrag aufgrund der Masse an „Netzpublikationen“ nicht erfüllen.⁷ Mithilfe von Web-Harvesting lässt sich automatisiert ein großer Bestand an Inhalten aus dem Web aufbauen, beziehungsweise ein bestimmter Bereich des Web (der Bereich .de) sammeln und als Zeugnis der menschlichen Entwicklung der Gesellschaft im „Informationszeitalter“ dokumentieren. Werden amtliche Dokumente über das Web kommuniziert, dann fallen diese in den Sammelauftrag von Archiven, die zur Erfüllung dieses Auftrages teilweise Web-Harvesting einsetzen.⁸

3. Archivierungstiefe des Web-Harvesting

Die derzeit auf dem Markt erhältlichen Produkte, mit denen sich Web-Harvesting zur Archivierung bestimmter Bereiche des Web einsetzen lässt, sind mit zahlreichen technischen Defiziten behaftet.⁹ So können Inhalte des sogenannten „deep web“¹⁰ beim flächigen Harvesting ebensowenig erfasst werden, wie gut verborgene Hyperlinks (deep-links). Dynamische Elemente als Teile von Webseiten können Endlosschleifen (crawler traps) verursachen, in denen

⁴ Vgl. zur Technik des Cache: *Martin Bahr*, The Wayback Machine und Google Cache – eine Verletzung deutschen Urheberrechts?, in: *JurPC Web-Dok.* 29/2002, Abs. 2, *Stephan Ott*, Der Google Cache – eine milliardenfache Urheberrechtsverletzung? In: *MIR* 2007, Dok. 195, Rn. 2.

⁵ Gesetz über die Deutsche Nationalbibliothek vom 22.06.2006, BGBl. I 2006, 1338.

⁶ Der Begriff „Netzpublikation“ steht nicht im DNBG, ist aber in der Definition zum Begriff „Medienwerk“ aus § 3 Abs. 1 DNBG enthalten. Netzpublikationen sind diejenigen Medienwerke, die „über Netze unkörperlich der Öffentlichkeit zugänglich gemacht werden“.

⁷ Siehe Pressemitteilung der DNB vom 29.06.2006, „Deutsche Nationalbibliothek mit erweitertem Sammelauftrag“. Durch automatisierte Web-Archivierung kann ein großes Datenvolumen aufgenommen werden, gleichzeitig ist der Personalaufwand wesentlich geringer und wächst nicht proportional zur Größe der Sammlung, wie das bei manuellen Methoden der Fall ist. Doch auch mithilfe des Web-Harvesting, egal ob flächig oder fokussiert, lässt sich nicht jeder Inhalt aus dem Web zu jedem Zeitpunkt sichern und erschließen. Vielmehr wird (beispielsweise zwei mal jährlich) eine Momentanaufnahme der Inhalte von bestimmten Bereichen des Web (zum Beispiel der Domain „.de“) gemacht.

⁸ So z.B. das Landesarchiv von Baden-Württemberg, siehe das Baden-Württembergische Online-Archiv BOA: <http://la.boa-bw.de/menu.do?index>

⁹ Das kommerzielle, Freeware-, und Open-Source- Angebot ist vorrangig auf die Bedürfnisse des *selektiven Web-Harvesting* ausgerichtet und genügt überwiegend den Anforderungen an Archivierungswürdigkeit nicht. Siehe hierzu: *Hans Liegmann*, Web-Harvesting: Aktivitäten von Nationalbibliotheken, in: Barbara Hoen (Hrsg.), *Planungen, Projekte, Perspektiven*. Zum Stand der Archivierung elektronischer Unterlagen, Veröffentlichungen des Landesarchivs Nordrhein-Westfalen Bd. 10, Düsseldorf 2006, S. 58 f.

¹⁰ Unter dem Begriff „deep web“ versteht man denjenigen Teil des World Wide Web, der von Suchmaschinen nicht erfasst wird und deshalb bei einer Recherche nicht über übliche Suchmaschinen auffindbar ist. Dazu zählen einerseits die Inhalte, die erst nach einer Authentifikation zugänglich sind (z. B. passwortgeschützte Daten, Intranet) und andererseits Inhalte, die aus technischen Gründen nicht erfassbar sind, wie zum Beispiel Dokumente, die nicht als statische Seiten existieren, sondern erst auf Anfrage von einem Web-Dienst generiert werden (z.B. aus einer Datenbank).

sich der Web-Crawler verfängt.¹¹ Oft sind Webseiten nicht ihren jeweiligen Spezifikationen entsprechend programmiert,¹² und während moderne Web-Browser kulant über solche HTML-Fehler hinwegsehen, können die gegenwärtigen Web-Crawler fehlerhafte Dokumente nicht angemessen verarbeiten.

II. Urheberrechtliche Zulässigkeit von Web-Harvesting

In Bezug auf urheberrechtlich geschützte Werke im Web stellt die mit dem Abspeichervorgang beim Web-Harvesting verbundene Vervielfältigung eine (in Deutschland nach § 16 UrhG) relevante Handlung dar und bedarf als solche entweder der Einwilligung des Urhebers (beziehungsweise Rechteinhabers),¹³ oder einer gesetzlichen Grundlage.

Wird das Web-Harvesting nicht auf eine nationale Länderkennung (z.B. den Bereich .de) beschränkt, dann stellen sich grenzüberschreitende Urheberrechtsfragen. Ein einheitliches weltweites Urheberrecht gibt es ebenso wenig, wie ein einheitliches Gemeinschaftsurheberrecht der EU. Aus diesem Grund ist im Einzelfall zu klären, welches nationale materielle Urheberrecht zur Anwendung kommt. Nach dem im Urheberrecht geltenden Schutzlandprinzip, ist an das Recht des Schutzlandes anzuknüpfen, also desjenigen Landes, für das Schutz beansprucht wird.¹⁴ Das heißt, deutsches Recht ist dann anzuwenden, wenn die Verletzung deutscher Urheberrechte geltend gemacht wird. Aus Platzgründen muss sich der Beitrag auf die Konfliktsituationen mit dem deutschen Urheberrecht beschränken.

1. Zulässigkeit zielgerichteten Web-Harvestings

Wenn beim *zielgerichteten* Web-Harvesting die Sammlung von Web-Inhalten dahingehend beschränkt wird, dass nur eigene Inhalte aus dem Web geharvestet werden, also solche Inhalte, die von der eigenen Institution im Web kommuniziert werden, ergeben sich rechtlich keine Probleme. Landesarchive,¹⁵ Parteienarchive,¹⁶ oder etwa das Bundestagsarchiv¹⁷ beschränken sich auf solche Inhalte und kommen daher nicht in rechtliche Bedrängnis.

Wenn beim *zielgerichteten* Web-Harvesting die Sammlung von Inhalten dahingehend beschränkt wird, dass nur solche Inhalte aus dem Web gesammelt werden, bei denen der Urheber beziehungsweise Rechteinhaber in die Sammlung eingewilligt hat, dann bestehen aufgrund des Einverständnisses in die urheberrechtlich relevante Handlung ebenfalls rechtlich keine Probleme.¹⁸

¹¹ Zu den technischen Problemen siehe ausführlich: *Hans Liegmann*, Web-Harvesting: Aktivitäten von Nationalbibliotheken, in: Barbara Hoen (Hrsg.), Planungen, Projekte, Perspektiven. Zum Stand der Archivierung elektronischer Unterlagen, Veröffentlichungen des Landesarchivs Nordrhein-Westfalen Bd. 10, Düsseldorf 2006, S. 58 f.

¹² Vgl. den HTML/XHTML Prüfdienst unter <http://validator.w3.org>

¹³ Welche vorab (und für das Web-Harvesting bedeutsam ggfs. sogar maschinenlesbar) in einer sogenannten „Copyleft“-Lizenz (Creative Commons u.a.) erklärt werden kann. Die Möglichkeit zur Einräumung unbeschränkter einfacher Nutzungsrechte ergibt sich aus den §§ 29 Abs. 2, 31 Abs. 1, 2 UrhG.

¹⁴ *Thomas Dreier*, in: ders./Gernot Schulze, UrhG Kommentar, 2. Auflage, München 2006, Vor §§ 120ff. RN28.

¹⁵ Wie z.B. das Baden-Württembergische Online-Archiv BOA siehe FN 10.

¹⁶ Vgl. die kooperativen Aktivitäten einiger deutscher Parteienarchive unter www.fes.de/archiv/spiegelungsprojekt.htm (zuletzt aufgerufen am 15.11.2007).

¹⁷ Auch der Deutsche Bundestag hat ein digitales Archiv, in dem eigene digitale archiviert werden.

¹⁸ So verfährt beispielsweise das wohl bekannteste Projekt aus dem Ausland, das von der Nationalbibliothek Australiens betriebene Web-Archiv PANDORA (Preserving and Accessing Networked Documentary Resources of Australia), ein Archiv das aus ausgewählten australischen Online-Publikationen besteht.

Unter das rechtlich als unproblematisch einzustufende „zielgerichtete Web-Harvesting mit Einwilligung“ fällt auch die mithilfe des „Open Archives Initiative (OAI)¹⁹-Harvesting“ betriebene Sammlung. Im Fall des OAI-Harvesting bietet ein Informationsdienstleister einen Dokumentenserver (OAI-Repository) an,²⁰ über den die Metadaten der auf diesem Server befindlichen Inhalte such- und auffindbar sind, sodass sich ein Informationsdienstleister (nicht nur Bibliotheken und Archive) automatisch über neue Inhalte informieren kann. Falls es solche gibt, können über die OAI-Schnittstelle alle für die Verarbeitung notwendigen Metadaten abgeholt werden. Eine manuelle Interaktion ist hierfür nicht mehr erforderlich.²¹ OAI-Harvesting geschieht mit dem ausdrücklichen Einverständnis der die Inhalte auf dem Server hinterlegenden Urheber.

Insbesondere die DNB nutzt das OAI-Harvesting zur Erfüllung ihres Sammelauftrages und kommt damit auch den nach § 15 DNBG Ablieferungspflichtigen entgegen,²² die Ihre sich aus § 14 DNBG ergebende Ablieferungspflicht nach § 16 S. 2 DNBG auch durch Bereitstellung der dem Sammelauftrag der DNB unterfallenden Werke nach Maßgaben der Bibliothek erfüllen können.²³

2. Zulässigkeit flächigen Web-Harvestings

Beim *flächigen* Web-Harvesting wird nicht zwischen rechtlich geschützten und ungeschützten Inhalten unterschieden und es wird diese Unterscheidung auch nicht zuvor in einer URL seed list getroffen. Es werden auch solche Inhalte gesammelt, die nicht zur Sammlung bestimmt sind und bewusst zur Abfrage eines Informationsdienstleisters über OAI-Repositorys bereitgehalten werden. Die geharvesteten Informationen werden nicht, wie beim OAI-Harvesting zur Erfüllung einer Ablieferungsverpflichtung, oder aus sonstigen Motiven, zur Sammlung „angeboten“,²⁴ sondern ohne Zutun (und gegebenenfalls Einwilligung) der Urheber gesammelt.²⁵ Beim flächigen Web-Harvesting lässt sich nicht verhindern, dass fremdes, urheberrechtlich geschütztes Material ohne die erforderliche Einwilligung archiviert wird. Die mit der Archivierung verbundene Vervielfältigung bedarf dann, wenn keine Einwilligung des Urhebers beziehungsweise Rechteinhabers vorliegt, einer gesetzlichen Grundlage.

¹⁹ Die Open Archives Initiative ist eine Initiative von Betreibern von Preprint-Servern und anderen Dokumentenservern, um die auf diesen Servern abgelegten elektronischen Publikationen im Web besser auffindbar und nutzbar zu machen. Dazu werden verschiedene einfache Techniken entwickelt und bereitgestellt, insbesondere das OAI Protocol for Metadata Harvesting (OAI-PMH) zum Einsammeln und Weiterverarbeiten von Metadaten.

²⁰ Z.B. die Digital Peer Publishing NRW: <http://www.dipp.nrw.de/ueberdipp/> (zuletzt aufgerufen am 16.11.2007).

²¹ Unterstützt werden derzeit das Datenformat ONIX sowie die Ressourcen-Formate text, application/pdf und die Archivformate application/zip, application/gzip, application/gzip-compressed. Zukünftig soll die Übertragung von Metadaten in weiteren Datenformaten möglich sein.

²² Das OAI-Harvesting als „Abholungsverfahren“ ist vor allem hilfreich für Ablieferer, die große Mengen digitaler sammlungsrelevanter Werke anbieten und für die das „Ablieferungsverfahren“ über Anmeldeformular daher nicht realisierbar ist.

²³ Auf Nachfrage bei der DNB hat Nicole Stephan von der DNB zu den Maßgaben der Bibliothek am 22.10.2007 folgende Auskunft gegeben: „Die „Maßgaben der Bibliothek“ sind...(noch) nicht konkret festgelegt; es können daher Einzelabspachen getroffen werden. Wichtig ist, daß eine Schnittstelle vorhanden sein muß, die es der Deutschen Nationalbibliothek ermöglicht, evtl. vorhandene Metadaten sowie die Internetveröffentlichung selbst abzuholen.“

²⁴ Das Anbieten zur Sammlung setzt ein entsprechendes Erklärungsbewusstsein voraus, welches nicht einfach unterstellt werden kann. Die DNB kann sich zur vollständigen Sammlung aller im Web „angebotenen“ Inhalte daher nicht etwa auf eine vorhandene Einwilligung in die mit der Sammlung verbundene Vervielfältigung berufen.

²⁵ In diesem Zusammenhang wird auch vom „Selbstbedienungsverfahren“ gesprochen, vgl. *Hans Liegmann*, Selbstbedienung oder Lieferung frei Haus? in: *Dialog mit Bibliotheken* 14, 2002, 15 ff.

Bei der Suche nach einer gesetzlichen Grundlage ist einzig darauf abzustellen, ob das Urheberrecht in den §§ 44a ff. UrhG (eine) Ausnahmeregelung(en) zugunsten der Bibliotheken und Archive für den Einsatz von flächigem Web-Harvesting enthält.²⁶ Dies gilt uneingeschränkt auch für *Pflichtexemplarbibliotheken*. Zwar korrespondiert im Falle von Pflichtexemplarbibliotheken die Sammlung mit Pflichtabgabevorschriften, diese können jedoch nicht die gesetzliche Grundlage für urheberrechtlich relevante Handlungen sein. Für die Landespflichtexemplarbibliotheken gilt dies schon aufgrund der Normenhierarchie. Das UrhG als Bundesrecht bricht das Landespflichtexemplarrecht. Aber auch das DNBG begründet nur die Sammlungskompetenz der DNB, die sich in Bezug auf die mit ihrer Sammlungstätigkeit verbundenen urheberrechtlich relevanten Handlungen im Übrigen an geltendes Urheberrecht halten muss.

Gegenwärtig existiert keine spezifisch auf den Einsatz von Web-Harvesting zugeschnittene Urheberrechtsschranke.²⁷ Da die Bibliotheken und Archive beim Web-Harvesting digitale Archive anlegen, liegt ein Rückgriff auf das Archivprivileg aus § 53 Abs. 2 S. 1 Nr. 2 i.V.m. Satz 2 Nr. 3 UrhG nahe. Dieses gestattet unter bestimmten Voraussetzungen das Anlegen digitaler Archive.²⁸

a) Das urheberrechtliche Archivprivileg

Im Folgenden ist zu prüfen, ob § 53 Abs. 2 S. 1 Nr. 2 i.V.m. Satz 2 Nr. 3 UrhG tatsächlich als Grundlage für flächiges Web-Harvesting herangezogen werden kann.

aa) Anwendbarkeit

Das Archivprivileg aus § 53 Abs. 2 S. 1 Nr. 2 i.V.m. S. 2 Nr. 3 UrhG, welches grundsätzlich sowohl die bei der Digitalisierung eines analogen Werkstücks als auch die bei der Kopie eines digitalen Werkstücks anfallende Vervielfältigungshandlung erfasst, ist nicht auf alle Werkkategorien anwendbar. Bestimmte Werkkategorien lassen sich nicht gestützt auf § 53 Abs. 2 S. 1 Nr. 2 i.V.m. S. 2 Nr. 3 UrhG vervielfältigen. Dies gilt namentlich für Computerprogramme²⁹ und Datenbanken.³⁰

Der Anwendbarkeit von § 53 Abs. 2 S. 1 Nr. 2 i.V.m. S. 2 Nr. 3 UrhG könnte entgegenstehen, dass die beim Web-Harvesting gesammelten Webseiten, oder in diese integrierten Werke, zu diesen Werkkategorien zählen.

²⁶ Die wiederum nach den Grundsätzen des Regel-/Ausnahmeprinzips restriktiv auszulegen ist, vgl. *Jürgen W. Goebel / Jürgen Scheller / Wolfgang Zimmermann*, Digitale Langzeitarchivierung und Recht, Bad Homburg v.d.H. 2004, S. 40 ff.

²⁷ Eine spezifische das Web-Harvesting erlaubende Schrankenbestimmung wird in der wissenschaftlichen Literatur für das diese Technologie verwendende Content-Caching von Suchmaschinen gefordert. So: Stephan Ott, *Der Google Cache – Eine milliardenfache Urheberrechtsverletzung?*, in: MIR 2007, Dok. 195, Rz. 1-20, Rz. 9. Permanenter Link: http://medien-internet-und-recht.de/volltext.php?mir_dok_id=697

²⁸ Die Gesetzesbegründung zum DNBG beruft sich in § 2 (Aufgaben und Methoden) Nummer 1 letzter Teilsatz auch für den Einsatz von flächigen Web-Harvesting auf § 53 Abs. 2 S. 1 Nr. 2 i.V.m. Satz 2 Nr. 3 UrhG.

²⁹ *Thomas Dreier*, in: ders./Gernot Schulze, UrhG Kommentar, 2. Auflage, München 2006, § 69d Rn. 3: Computerprogramme dürfen nur unter den Voraussetzungen des § 69 d UrhG vervielfältigt werden, welcher den § 44 a ff. UrhG als *lex specialis* vorgeht. Durch § 69 d UrhG ausgeschlossen ist insbesondere § 53 UrhG

³⁰ *Thomas Dreier*, in: ders./Gernot Schulze, UrhG Kommentar, 2. Auflage, München 2006, § 87c UrhG, Rn. 1: Datenbanken dürfen nur unter den Voraussetzungen von § 87 c UrhG, der als abschließende Sonderregelung den §§ 44 a ff. UrhG vorgeht, vervielfältigt werden. Datenbankwerke werden gemäß § 53 Abs. 5 S. 1 UrhG von der Privilegierung aus § 53 Abs. 1 S. 1 Nr. 2 UrhG ausdrücklich ausgenommen.

Webseiten an sich können einen Schutz als Computerprogramm, Datenbank(werk) oder Werk der bildenden Kunst genießen.³¹ Die einzelnen, in die Webseite integrierten, Werke können einen Schutz als Filmwerk und filmähnliches Werk, Sprachwerk, Darstellung wissenschaftlicher oder technischer Art, oder Lichtbildwerk genießen.³²

Auf der Grundlage von § 53 Abs. 1 S. 1 Nr. 2, S. 2 Nr. 3 UrhG lassen sich nur diejenigen Webseiten, oder in eine Webseite integrierten Werke sammeln, welche keinen zusätzlichen Schutz gem. §§ 4, 69d, 87c UrhG genießen.

bb) Privilegierte Einrichtungen

Privilegiert werden nur *Archive*, die (so § 53 Abs. 2 S. 2 Nr. 3 UrhG) *im öffentlichen Interesse tätig* sind,³³ und die *keinen unmittelbar oder mittelbar wirtschaftlichen oder Erwerbszweck verfolgen*. Unter den Begriff *Archiv* fallen auch *Archivbibliotheken*. Archivbibliotheken sind in der Regel im öffentlichen Interesse tätig und sie verfolgen in der Regel (zumindest bis jetzt, was die zukünftige Entwicklung angeht ist bereits die kommerzielle Verwertung kultureller Äußerungen angedacht)³⁴ keinen unmittelbar oder mittelbar wirtschaftlichen oder Erwerbszweck. Privilegiert sind ferner die Archive von Einrichtungen, die im öffentlichen Interesse der Lehre und Forschung dienen (z.B. Institutsarchive).³⁵ Damit wird der vom europäischen Gesetzgeber in Art. 5 Abs. 2 lit. a und Art. 5 Abs. 3 lit. a der Urheberrechtsharmonisierungsrichtlinie³⁶ gelassene Spielraum voll ausgeschöpft.

cc) Aufnahme in ein „eigenes“ Archiv

Weitere Voraussetzung für die nach § 53 Abs. 2 S. 1 Nr. 2 UrhG privilegierte Nutzung urheberrechtlich geschützter Werke ist, dass die Vervielfältigungsstücke eines Werkes zur Aufnahme in ein *eigenes* Archiv hergestellt werden. Dieses Adjektiv wird nach herrschender Ansicht verstanden als ein lediglich zur internen Nutzung bestimmtes Archiv.³⁷ Diese Auslegung deckt sich mit dem Archivierungszweck, kann allerdings nicht mehr getrennt von diesem betrachtet werden und verliert eine eigenständige prüfungsrelevante Bedeutung. Daher sprechen gute Gründe dafür, das Adjektiv „eigenes“, als „zu Eigentum gehörendes“ aufzufassen. Einer solchen Auslegung steht insbesondere nicht der Archivierungszweck entgegen, der als weitere Voraussetzung gegeben sein muss. Bibliotheken und Archive nutzen Web-Harvesting um ihren „zu Eigentum gehörenden“ Bestand auszubauen.

³¹ Zur Einordnung einer Webseite in die bestehenden Werkkategorien ausführlich: *Stephan Ott*, Urheber- und wettbewerbsrechtliche Probleme von Linking und Framing, Stuttgart 2004, S. 241 ff.

³² *Malte C.G. Marquadt* in: Artur-Axel Wandtke/Winfried Bullinger (Hrsg.) Praxiskommentar zum Urheberrecht, 2. Aufl., München 2006, § 4 Rn. 13.

³³ Diese mit dem zweiten Korb eingeführte Änderung beschränkt den Kreis der nach § 53 Abs. 2 S. 1 Nr. 2 i.V.m. Satz 2 Nr. 3 UrhG Privilegierten zusätzlich.

³⁴ Vgl. die vom Fraunhofer Institut für Intelligente Analyse- und Informationssysteme Karlsruhe angefertigte und vom BKM herausgegebene „Bestandsaufnahme zur Digitalisierung von Kulturgut und Handlungsfelder“ von 2006, abrufbar unter: http://www.imk.fraunhofer.de/BKM-Studie/BKM_End_55.pdf, zuletzt aufgerufen am 12.11.2007, S. 111 ff.

³⁵ *Thomas Dreier*, in: ders./Gernot Schulze, UrhG Kommentar, 2. Auflage, München 2006, § 53 Rn. 27 f.

³⁶ Urheberrechtsrichtlinie 2001/29/EG.

³⁷ *Haimo Schack*, Dürfen öffentliche Einrichtungen elektronische Archive anlegen?, in: AfP 2003, 1 ff., S. 2.

dd) Zweck der Archivierung

Außerdem setzt § 53 Abs. 2 S. 1 Nr. 2 UrhG voraus, dass die Aufnahme eines Werkes in ein eigenes Archiv *geboten* sein muss und die Aufnahme selbst *keinen gewerblichen Zwecken dient*. Mit dem Ausschluss, dass die Aufnahme in ein digitales Archiv keinen gewerblichen Zwecken dienen darf, soll klargestellt werden, dass die zunehmend als attraktiv erkannte Vermarktung der Archivierungsleistung als solche³⁸ nicht vom Privileg des § 53 Abs. 2 S. 1 Nr. 2 UrhG gedeckt wird.³⁹ Mit dem „*Archivierungszweck*“ vereinbar ist nur die „Bestandssicherung“, nicht mit dem Archivierungsprivileg vereinbar ist dagegen der „Bestandsaufbau“.⁴⁰ Grund für die Privilegierung der Aufnahme von Vervielfältigungen in ein eigenes Archiv ist, dass die Archivierung insbesondere in Bibliotheken erleichtert werden soll.⁴¹ Der Archivierungszweck besteht in der *Sicherung* und *internen Nutzung* des *vorhandenen Bestandes* an geschützten Werken; bei Bibliotheken etwa darin, Raum zu sparen, oder den Bestand sicher zu lagern, ohne dass dadurch eine zusätzliche Verwertung des Werkes vorliegt, durch welche der Urheber betroffen wird. Die Anfertigung eines digitalen Archivs nach § 53 Abs. 2 S. 1 Nr. 2, S. 2 Nr. 3 UrhG darf also nur darauf abzielen, durch digitale „Sicherungskopien“ bereits gegebene, mit bestehenden (Leistungsschutz- und Urheber-) Rechten in Einklang stehende Nutzungsmöglichkeiten zu sichern und zu konservieren.

Ungeachtet von § 53 Abs. 6 S. 1 UrhG, nach dem privilegiert hergestellte Vervielfältigungsstücke weder verbreitet, noch zu öffentlichen Wiedergaben benutzt werden dürfen,⁴² ist es den Bibliotheken und Archiven schon aufgrund des von § 53 Abs. 2 S. 1 Nr. 2 UrhG geforderten *Archivierungszwecks* verwehrt, ihre Archive für externe Dritte zu öffnen.⁴³ § 53 Abs. 2 S. 1 Nr. 2 UrhG privilegiert nur zur Erfüllung des Archivzwecks, keineswegs jedoch dazu (leistungsschutz- oder urheber-) rechtlich geschützte Werke durch möglichst leichte und kostengünstige Vervielfältigung einer breiten Öffentlichkeit zu vermitteln.⁴⁴

Die Grenzen des privilegierten internen Gebrauchs sind überschritten, wenn das Archiv auch zur Verwendung durch außenstehende Dritte bestimmt ist.⁴⁵ Ein Archiv im Sinne von § 53 Abs. 2 S. 1 Nr. 2 UrhG dient allein der Bestandssicherung und Bewahrung von kulturellen Äußerungen, sein Zweck ist es nicht, archivierte Vervielfältigungsstücke einer Nutzung durch Dritte zugänglich zu machen.⁴⁶ Externen Dritten dürfen aufgrund von § 53 Abs. 2 S. 1 Nr. 2, S. 2 Nr. 3 UrhG nur diejenigen „Sicherungskopien“ zugänglich gemacht werden, durch die ein ihnen im Einklang mit bestehenden Urheberrechten bereits zugänglicher, aber beschädigter Bestand wiederhergestellt wird.

³⁸ Vgl. *Haimo Schack*, Dürfen öffentliche Einrichtungen elektronische Archive anlegen?, in: AfP 2003, 1 ff., S. 2.

³⁹ Der Hinweis darauf, dass mit der Archivierung keine gewerblichen Zwecke verfolgt werden dürfen, dient lediglich der Klarstellung, versteht sich im Übrigen von selbst, so der RegE eines Zweiten Gesetzes zur Regelung des Urheberrechts in der Informationsgesellschaft vom 15.06.2006, BT-Drucksache 16/1828, S. 26.

⁴⁰ Vgl. die amtliche Begründung der BT-Drucks. IV/270, S. 73

⁴¹ *Thomas Dreier*, in: ders./Gernot Schulze, UrhG Kommentar, 2. Auflage, München 2006, § 53 Rn. 27.

⁴² BGHZ 134, 250 ff., S. 259 f.

⁴³ *Haimo Schack*, Dürfen öffentliche Einrichtungen elektronische Archive anlegen?, in: AfP 2003, 1 ff., S. 2.

⁴⁴ *Haimo Schack*, Dürfen öffentliche Einrichtungen elektronische Archive anlegen?, in: AfP 2003, 1 ff., S. 2.

⁴⁵ BGHZ 134, 250 ff., S. 258, *Ulrich Loewenheim*, Urheberrechtliche Grenzen der Verwendung geschützter Dokumente in Datenbanken, in: AfP 1993, 613 ff., S. 618.

⁴⁶ *Norbert P. Flechsig / Martin Fischer*, Speicherung von Printmedien in betriebseigene Datenbankarchive und die Grenze ihrer betrieblichen Nutzung, in: ZUM 1996, 833 ff., S. 846.

Zwar wird beim Einsatz von Web-Harvesting durch Archive und Bibliotheken kein gewerblicher Zweck, im Sinne der Vermarktung einer Archivleistung verfolgt, jedoch ist nicht die Bestandssicherung oder Bestandswiederherstellung, sondern der Bestandsaufbau im Hinblick auf Netzpublikationen intendiert. Dieser soll weitergehend auch der breiten Öffentlichkeit zugänglich gemacht werden. Digitale Web-Inhalte sollen nicht im Hinblick auf ihre möglicherweise irgendwann bestehende Verwertbarkeit archiviert werden, sondern mit der Intention sie sofort vermitteln zu können. Beim flächigen Web-Harvesting werden Inhalte aus dem WWW zu dem Zweck vervielfältigt und archiviert, die archivierten Vervielfältigungen zumindest den Bibliotheksbenutzern, also externen Dritten, dauerhaft zugänglich zu machen.

Daran ändert im übrigen auch der mit dem zweiten Gesetz zur Änderung des Urheberrechts neu geschaffene § 52b UrhG nichts, selbst wenn er dann weitgehend leerläuft. § 52 b UrhG erlaubt bestimmten Einrichtungen unter bestimmten Voraussetzungen die Zugänglichmachung ihres analogen Bestandes an elektronischen Leseplätzen, enthält aber kein Recht zur Annex-Vervielfältigung.⁴⁷ Damit fehlt eine Regelung, wie die privilegierten Einrichtungen zu einem für die öffentliche Zugänglichmachung auf den Leseplätzen erforderlichen digitalen Werkexemplar kommen können.⁴⁸ Dieses Manko lässt sich nicht durch eine nunmehr weite Auslegung des Archivprivilegs beheben sondern nur, indem noch eine dem § 52 a Abs. 3 UrhG vergleichbare Annex-Vervielfältigungskompetenz geschaffen wird.

ee)Vervielfältigung eigener Archivstücke

Zuletzt setzt § 53 Abs. 2 S. 1 Nr. 2 UrhG voraus, dass als Vorlage für die Vervielfältigung ein *eigenes* Archivstück verwendet wird. Damit wird die gegebene Auslegung des Archivzwecks und die Absicht des Gesetzgebers zu verhindern, dass durch die Archivschränke zusätzliche Verwertungen urheberrechtlich geschützter Werke möglich werden, bestätigt. Weder dürfen die privilegierten Einrichtungen gestützt auf § 53 Abs. 2 S. 1 Nr. 2 UrhG ihren Bestand durch Vervielfältigungen/Kopien eigener Werkstücke erweitern, noch dürfen sie fremde Werkstücke vervielfältigen, um den eigenen Bestand auszubauen.

Beim Web-Harvesting werden „fremde“ Werke eingesammelt (geharvestet). Auch dann wenn Pflichtexemplarbibliotheken solche Netzpublikationen, bei denen eine Ablieferungspflicht besteht, harvesten, sind diese zunächst „fremd“. Das Web-Harvesting stellt eine vorweggenommene unrechtmäßige Ersatz-Beschaffungsmaßnahme dar, die jeglicher (rechtlichen) Grundlage entbehrt.⁴⁹

⁴⁷ Zu diesem Begriff: *Thomas Dreier*, in: ders./Gernot Schulze, *UrhG Kommentar*, 2. Auflage, München 2006, § 52a Rn. 16.

⁴⁸ *Christian Berger*, Die öffentliche Wiedergabe von urheberrechtlichen Werken an elektronischen Leseplätzen in Bibliotheken, Museen und Archiven - Urheberrechtliche, verfassungsrechtliche und europarechtliche Aspekte des geplanten § 52b UrhG, in: *GRUR* 2007, 754 ff., S. 756.

⁴⁹ Das wird in der Gesetzesbegründung zum DNBG (Begründung DNBG zu § 2 (Aufgaben und Methoden) Nummer 1 letzter Teilsatz) wohl übersehen, wenn es dort heißt: „Die Bibliothek muss zur Erfüllung ihrer Aufgaben Kopien fertigen, insbesondere im Rahmen der Sammlung elektronischer Daten im Internet oder deren Überführung in andere Formate zum Erhalt einer zeitgemäßen Zugriffsmöglichkeit. Die urheberrechtliche Befugnis der Bibliothek zum Erstellen dieser Kopien ergibt sich aus § 53 Abs. 2 S. 1 Nr. 2 i.V.m. Satz 2 Nr. 3 UrhG und bedarf keiner erneuten Regelung in diesem Gesetz. Große Datenmengen in elektronischer Form – etwa aus dem Internet, aber auch aus internen Datenbanken – können nur in automatisierten Verfahren kopiert werden.“ Richtig ist lediglich, die aus dieser Aussage zu gewinnende Feststellung, dass die Schranken für die Vervielfältigung urheberrechtlich geschützter kultureller Äußerungen (Werke) sich lediglich aus dem UrhG ergeben können und nicht etwa aus dem DNBG. Insoweit sich Erlaubnistatbestände nur aus dem UrhG ergeben können, ist ihre Formulierung im DNBG eine bloße Wiederholung und als solche überflüssig (kann aber der Verständlichkeit dienen).

b) Konkludente Einwilligung

Im nächsten Schritt ist zu prüfen, ob die mit Web-Harvesting verbundenen Eingriffe in Urheberrechte sich gegebenenfalls durch eine konkludente Einwilligung rechtfertigen lassen. Die Annahme einer konkludenten Einwilligung in eine Verwertungshandlung birgt die Gefahr in sich, die Balance der Interessen zwischen Urheberrechtsinhaber und Werkverwender zu verschieben und ist daher streng zu handhaben.⁵⁰

Anders als beim „Content-Caching“, bei dem Suchmaschinen Momentanaufnahmen von Webseiten herstellen und Nutzern über einen Cache-Link zur Verfügung stellen und bei dem eine konkludente Einwilligung des Urhebers auch für den Cache-Link für möglich gehalten wird,⁵¹ scheitert die Annahme einer konkludenten Einwilligung beim Web-Harvesting schon an der von den Bibliotheken und Archiven angestrebten *langfristigen* Verfügbarkeit der gewonnen digitalen Archive. Denn beim Content-Caching ist die nur vorübergehende *kurzfristige* Verfügbarkeit der betroffenen Seiten entscheidend für die Unterstellung, dass es sich hierbei in Parallele zur Proxy-Server-Problematik, um eine dem Web inhärente Technologie handele,⁵² während beim Web-Harvesting eine derartige technologische Inhärenz nicht gegeben ist.

Selbst wenn man eine konkludente Einwilligung für das Speichern (und Nutzen) von Webseiten im Sinne eines dem Urheberrechtssystem fremden „*Opt-Out-Systems*“⁵³ gleichwohl für möglich hält,⁵⁴ liegen die dafür erforderlichen Voraussetzungen jedenfalls nicht vor, denn (1.) gibt es beim Web-Harvesting formal keine Möglichkeit, der Werkverwendung ausdrücklich zu widersprechen,⁵⁵ weil es beim *flächigen* Web-Harvesting gerade um eine vollständige und lückenlose Sammlung bestimmter Bereiche des Web geht. Dabei beschränkt sich (2.) die Archivierung von Webseiten durch Bibliotheken und Archive gerade nicht auf den absolut notwendigen Umfang. Zudem kann (3.) nicht generell unterstellt werden, dass jeder Anbieter urheberrechtlich geschützter Inhalte im Web ein Interesse am dauerhaften Vorhandensein seines Angebots hat.

IV. Fazit

Nach geltender Rechtslage gibt es keine sichere Rechtsgrundlage für den Einsatz von flächigem Web-Harvesting. Beim *flächigen* Web-Harvesting besteht die technisch nicht auszuschließende *Möglichkeit* einer Urheberrechtsverletzung. Denn bei der Sammlung und Archi-

⁵⁰ *Stephan Ott*, Der Google Cache - Eine milliardenfache Urheberrechtsverletzung? In: MIR 2007, Dok. 195, Rn. 5.

⁵¹ *Martin Bahr*, The Wayback Machine und Google Cache – eine Verletzung deutschen Urheberrechts?, in: JurPC Web-Dok. 29/2002, Abs. 9.

⁵² *Martin Bahr*, The Wayback Machine und Google Cache – eine Verletzung deutschen Urheberrechts?, in: JurPC Web-Dok. 29/2002, Abs. 8 f.

⁵³ Der Annahme der Zulässigkeit einer urheberrechtlich relevanten Handlung, solange der Urheberrechtsinhaber ihr nicht widerspricht. Den Gegensatz bildet das geltende „*Opt-In-System*“, also die Annahme, dass eine urheberrechtlich relevante Handlung nur mit Einwilligung des Urhebers, beziehungsweise bei Vorliegen einer Ausnahmebestimmung, zulässig ist.

⁵⁴ *Stephan Ott*, Der Google Cache - Eine milliardenfache Urheberrechtsverletzung? In: MIR 2007, Dok. 195, Rn. 5 ff.; *Stephan Ott*, Zulässigkeit der Erstellung von Thumbnails durch Bilder- und Nachrichtensuchmaschinen?, in: ZUM 2007, 119 ff., S. 126f.

⁵⁵ Der Abspeicherung von Webseiten lässt sich bspw. durch META-Tags oder Robots.txt im Code einer Webseite widersprechen.

vierung mittels Web-Harvesting werden auch solche urheberrechtlich geschützten Webseiten oder in diese integrierten Teile vervielfältigt, die sich nicht auf der Grundlage von § 53 Abs. 1 S. 1 Nr. 2 i.V.m. S. 2 Nr. 3 UrhG vervielfältigen lassen und in deren Vervielfältigung nicht eingewilligt wurde.

Zwar ist ein Schaden beim Urheber schwer vorstellbar, wenn Bibliotheken und Archive, welche „keinen unmittelbar oder mittelbar wirtschaftlichen oder Erwerbszweck verfolgen und im öffentlichen Interesse tätig sind,“ urheberrechtlich geschützte Webseiten oder Teile von Webseiten mittels Web-Harvesting archivieren,⁵⁶ jedoch können sich die Bibliotheken und Archive im europäischen Raum hierauf - anders als die Bibliotheken und Archive im angloamerikanischen Raum, wo immer ein Rückgriff auf die generalklauselartige Fair-Use-Schranke möglich ist - nicht berufen.⁵⁷ Unabhängig von einem Schaden drohen ihnen urheberrechtliche Unterlassungs-, und Vernichtungsansprüche oder gar die strafrechtliche Verfolgung.⁵⁸

Angesichts der Bedeutung der Arbeit der Bibliotheken und Archive und der Wichtigkeit der dauerhaften Kommunikation kultureller Äußerungen über alle (auch digitale und vernetzte) Medien, ist eine Rechtssicherheit schaffende Ergänzung des UrhG vorzunehmen. Um dies zu gewährleisten sollten sich die Bibliotheken und Archive stärker als bisher in die urheberrechtlichen Diskussionen miteinbringen.⁵⁹

⁵⁶ Bestimmungen aus § 53 Abs. 1 S. 1 Nr. 2 i.V.m. S. 2 Nr. 3 UrhG, dessen Voraussetzungen mindestens erfüllt sein müssen.

⁵⁷ Zum Vorliegen der Voraussetzungen der generalklauselartigen Fair-Use-Schranke beim Content-Caching siehe: *Stephan Ott*, Der Google Cache - Eine milliardenfache Urheberrechtsverletzung? In: *MIR* 2007, Dok. 195, Rn. 11 ff.

⁵⁸ §§ 97 ff. und 106 ff. UrhG.

⁵⁹ Bisherige Bemühungen im zweiten Korb scheinen ungehört verhallt zu sein, was sicher mit dem schon zuvor gesteckten Rahmen des zweiten Korbes zu tun hatte.