

Eckhard ARNOLD, Stefan MÜLLER

Wie permanent sind Permalinks?

Zusammenfassung

In diesem Artikel versuchen wir einige Missverständnisse aufzuklären, die das Konzept von „Permalinks“ betreffen. Wir erörtern zunächst, was Permalinks sind und welchem Zweck sie dienen. Unserer Ansicht nach sind Permalinks weniger als eine technische Kategorie zu verstehen, denn als ein Versprechen der bereitstellenden Institution oder Person, die als Permalink ausgewiesenen Links künftig nicht fallen zu lassen oder zu ändern. Sodann grenzen wir den Begriff gegen andere Linkarten ab. Wir diskutieren kurz das Verhältnis von Permalinks zu anderen digitalen Identifikatoren. Anschließend untersuchen wir, worauf beim praktischen Gebrauch von Permalinks zu achten ist. Wir schließen mit einem konkreten Praxis-Beispiel, das den Aufbau des Publikationsservers für die Bayerische Akademie der Wissenschaften betraf.

Schlüsselwörter

Permalink, Permanente Links, Digitale Identifikatoren, Zitierbarkeit elektronischer Quellen

Hinweis:

Der Beitrag ist eine gründlich überarbeitete und stark ergänzte (mindestens doppelt so lang) Fassung eines Blogbeitrags: <http://dhmuc.hypotheses.org/1179>. Einige Textstellen, insbesondere die Kurzzusammenfassung am Anfang, decken sich wortwörtlich.

1 Das Wichtigste in Kürze

- Permalinks sind dauerhafte Netzadressen von digitalen Ressourcen. Dabei sichert der Bereitsteller der digitalen Ressource zu, dass die digitale Ressource unter der als Permalink erklärten Netzadresse in unveränderter Form verfügbar bleibt.
- Permalinks erkennt man daran, dass der Bereitsteller ihre Permanenz durch öffentlich zugängliche Permalink-Leitlinien ausdrücklich zusichert.

Ein Beispiel für Permalink-Leitlinien ist: de.wikipedia.org/w/index.php?title=Hilfe:Permalink&oldid=152277048. Hier heißt es: „Dieser Verweis bleibt statisch auf stets dieselbe Version gerichtet, selbst wenn die Seite anschließend verändert wird und damit neue Versionen erzeugt werden.“

Ein Beispiel für einen darunter fallenden Permalink: de.wikipedia.org/w/index.php?title=Lola_Montez&oldid=154537216 .

Kein Permalink ist dagegen die Adresse: de.wikipedia.org/wiki/Lola_Montez, da der unter dieser Adresse verfügbare Artikel sich mit der Zeit durch Überarbeitungen verändern kann. – Dieses Beispiel zeigt die Bedeutung der ausdrücklichen Erklärung zu den Permanenzeigenschaften. Ohne sie würde man gerade im ersten Link nicht den Permalink vermuten, weil er ein Implementierungsdetail enthält: „.php“. Das ist ungünstig, wenn die Seite einmal nicht mehr mit PHP, sondern einer anderen Programmiersprache läuft.

- Für die Zitierbarkeit genügt ein Permalink. Digitale Identifikatoren, wie z.B. eine DOI-Nummer, sind als Alternative ebenfalls möglich, aber nicht zwingend erforderlich, da sie keine größere Permanenz der Ressource garantieren können als ein Permalink.
- Bei der Angabe einer URL in einem Literaturverweis sollte immer der Permalink angegeben werden. Die Angabe eines Zugriffsdatums ist eher überflüssig.
- Vorsicht: Nicht alles, was als Permalink bezeichnet wird, ist in dem eben beschriebenen Sinne permanent. Entscheidend ist, welche Permanenz-Eigenschaften in den Permalink-Leitlinien ausdrücklich zugesichert werden (und wie vertrauenswürdig die dahinter stehende Institution ist).

Beispiel: Die BV-Nummern im BV-Katalog haben ebenso wie die im BV-Katalog als solche bezeichneten Permalinks nur schwache Permanenz-Eigenschaften. Sie können im Rahmen der Überarbeitung des Katalogs spurlos verschwinden.

2 Einleitung

Die Zitierbarkeit digitaler Publikationen hängt wesentlich davon ab, dass sie über dauerhafte Netzadressen, sogenannte „Permalinks“ verfügen. Dauerhafte Netzadressen sind erforderlich, um die Nachvollziehbarkeit einer wissenschaftlichen Argumentation sicherzustellen, die sich auf digital vorliegende Referenzen stützt. Nachvollziehbar ist eine sich auf Zitate und Referenzen stützende Argumentation nur dann, wenn der Rezipient oder die Rezipientin die referierte Sekundärliteratur (oder auch Quellen und Forschungsdaten) wiederfinden und einsehen kann. Bei Druckwerken dienen dazu seit jeher die bibliographischen Angaben, die es im Zusammenhang mit einem ausgebauten Bibliothekssystem ermöglichen, die referierte Literatur zu beschaffen. Bei elektronischen Publikationen ermöglicht dies die dauerhafte Netzadresse, sprich der „Permalink“. Sind Permalinks vorhanden, dann können digitale Referenzen alle ihre Vorteile ausspielen, insbesondere den, dass sie -- zumindest bei Open Access Publikationen -- unmittelbar verfügbar sind („nur einen Mausklick weit entfernt“). Das kann die Nachvollziehbarkeit und Überprüfbarkeit einer Argumentation in der wissenschaftlichen Praxis erheblich vereinfachen und beschleunigen.

Aber woher bekommen Wissenschaftlerinnen und Wissenschaftler, die digitale Publikationen zitieren möchten, einen Permalink auf die digitale Publikation? Und welchen Praxis-Regeln sollten Informationsdienste wie Bibliotheken, Archive, Repositorien und Digital-Humanities-

Datenzentren folgen, die Permalinks auf die von ihnen verwalteten Publikationen oder Daten anbieten möchten? Wie das bei einem noch recht jungen Begriff wie dem des Permalinks, noch dazu aus dem sich rasch entwickelnden und ständigem Wandel unterworfenen Bereich der Digitaltechnik, vielleicht nicht anders zu erwarten ist, wird der Begriff uneinheitlich verwendet. Daraus können sich Unsicherheiten und im schlimmsten Fall auch eine falsche Handhabung von Permalinks ergeben -- ganz zu schweigen von den Vorbehalten, die besonders in den Geisteswissenschaften bezüglich der Zitierbarkeit von Internetquellen zuweilen immer noch anzutreffen sind.

Im Folgenden möchten wir daher einige Überlegungen dazu anstellen, was Permalinks sind, d.h. insbesondere woher sie ihre Permanenzeigenschaft beziehen und über welche weiteren Eigenschaften sie verfügen müssen, um ihre Zitierbarkeit in wissenschaftlichen Kontexten zu ermöglichen. Den Begriff Permalink verstehen wir dabei so, dass es sich immer um eine URL handelt, auch wenn wir gelegentlich Querverbindungen zu anderen Arten von permanenten Identifikatoren ziehen. Wir grenzen den Begriff gegen andere Linktypen und atypische, d.h. für den wissenschaftlichen Kontext untaugliche, Verwendungsweisen ab. Weiterhin berichten wir über eigene Erfahrungen, die wir an der Bayerischen Akademie der Wissenschaften mit der Einführung eines Permalinksystems beim Aufbau unseres Publikationsservers gemacht haben.

Bei diesem Artikel handelt es sich um eine überarbeitete und erweiterte Fassung eines Blogbeitrags auf dhmuc.hypotheses.org/, der die Anregungen und die Kritik aus der Diskussion aufgreift. Wir möchten allen, die sich an der Diskussion auf dem Blog beteiligt haben, an dieser Stelle für ihre konstruktiven und hilfreichen Beiträge danken.

3 Was sind Permalinks?

3.1 Worauf die Permanenzeigenschaft von Permalinks beruht

Die Permanenz eines Permalinks beruht einzig und allein auf der erkennbaren und im besten Falle öffentlich erklärten Selbstverpflichtung der Institution oder Person, die die Links ausgibt, unter ein- und demselben Permalink auch künftig stets dasselbe digitale Dokument bereit zu stellen, und der konsequenten Einhaltung dieser Selbstverpflichtung.

Es ist wichtig, sich klar zu machen, dass die Permanenzeigenschaft eine Frage der Leitlinien der bereitstellenden Institution ist und kein technisches Merkmal eines Links, so wie etwa die einer URL zugeordnete IP-Adresse als ein technisches Merkmal aufgefasst werden könnte. Anders ist es aber auch gar nicht möglich, denn es existiert kein technisches Verfahren, durch das man einen Internetlink dauerhaft machen könnte, so wie man z.B. die Echtheit eines digitalen Dokuments durch eine elektronische Signatur garantieren kann. Technische Systeme, wie z.B. „Handle-Server“ können die Bereitstellung von Permalinks nur unterstützen, aber nicht garantieren, dass deren Betreiber die Links beibehält. Das bedeutet aber auch, dass die Nutzerinnen und Nutzer eines Permalinks der bereitstellenden Institution vertrauen müssen, und diese Institution ihrerseits hinreichend vertrauenswürdig sein muss.

Da die Existenz und Dauerhaftigkeit von Permalinks von der bereitstellenden Institution abhängt, können auch Permalinks „verweisen“, wenn die Institution aufhört zu existieren und

das Weiterbestehen der Permalinks nicht im Abwicklungsprozess der Institution sichergestellt wird. Das sollte aber kein Hinderungsgrund für die wissenschaftliche Nutzung von Permalinks und insbesondere das Zitieren elektronischer Dokumente unter Angabe des Permalinks sein. Durch die richtige Handhabung von Permalinks seitens der bereitstellenden Institutionen lässt sich dieses Restrisiko nämlich hinreichend eingrenzen.

3.2 Wesentliche Eigenschaften von Permalinks

Aus dem Erfordernis, für die Referenzierung wissenschaftlicher Quellen und Sekundärliteratur verwendbar zu sein, ergibt sich eine Reihe von Eigenschaften, über die Permalinks mindestens verfügen sollten, um ihren Zweck zu erfüllen. Diese Eigenschaften können zugleich als Kriterien verstanden werden, an denen sich die Permalink-Leitlinien eines Informationsdienste-Anbieters messen lassen müssen.

3.2.1 Permalinks sollen als Permalinks öffentlich gekennzeichnet sein

Damit diejenigen, die elektronische Dokumente zitieren möchten, wissen, dass es sich um einen Permalink handelt, muss die bereitstellende Institution in irgendeiner Form deutlich machen, dass und welche Links sie laut Selbstverpflichtung nicht mehr ändern wird. Zitierbar ist ein Permalink nur dann, wenn man einigermaßen sicher sein kann, dass die als permanent erklärte URL auch in Zukunft erreichbar ist und auf denselben Inhalt verweist.

In der Praxis existieren dabei Grauzonen, weil die Permalink-Leitlinien nicht immer hinreichend deutlich erkennbar gemacht werden oder weil der Ausdruck „Permalink“ mehrdeutig und manchmal nur in einem eingeschränkteren technischen Sinn verwendet wird, der wissenschaftlichen Ansprüchen nicht genügt. So haben wir z.B. bei archive.org keine deutlichen Permalink-Leitlinien gefunden,¹ obwohl wir davon ausgehen, dass die Links, die archive.org für archivierte Websites vergibt, hinreichend vertrauenswürdig sind, um zitiert zu werden, da diese Institution sonst ihren erklärten Zweck verfehlen würde.

Auch ein Link, der in einer Zitierempfehlung angegeben ist, sollte unserer Ansicht nach ein Permalink sein, wie es z.B. beim Deutschen Textarchiv der Fall ist. Da man sich bislang allerdings nicht darauf verlassen kann, ist es besser, wenn Permalinks bis auf weiteres ausdrücklich als solche gekennzeichnet werden.

3.2.2 Ein Permalink muss immer zu demselben Ziel führen

Im Sinne der Zitierbarkeit ist zu verhindern, dass jemand, der einem Permalink in einer wissenschaftlichen Arbeit folgt, eine (für die vorliegende Frage entscheidend) andere Fassung des Dokumentes sieht als jene, die vom Autor der Arbeit verwendet wurde. Im Falle von Ergänzungen oder Berichtigungen eines Dokumentes sind folgende (einander nicht ausschließende) Lösungen denkbar:

- Neben der Angabe des Permalinks wird deutlich festgehalten, dass und in welchem Umfang sich das Verweisziel ändern kann: Welche Änderungen können stillschweigend

¹ Nur etwas versteckte Hinweise darauf wie „Whenever you refer to files at archive.org, use our permalink-style form for the file within the item. It is of the form <http://archive.org/download/IDENTIFIER/FILE>“ auf archive.org/help/video.php.

geschehen? Bei welchen wird dagegen eine der folgenden Maßnahmen ergriffen?

- Änderungen am Ziel werden so durchgeführt, dass man eine Änderung als solche erkennen, den Zeitpunkt der Änderung erfahren und möglichst auch den Ausgangszustand erschließen kann. Zum Beispiel, indem Ergänzungen oder Berichtigungen eines Regests unter dem ursprünglichen Text erscheinen, datiert und als Zusätze ausgezeichnet sind. Bestenfalls ist diese Datierung und Auszeichnung auch maschinenlesbar hinterlegt. Dieser Maßnahme entspricht in der Druckwelt die Beigabe von Addenda und Corrigenda zu einem sonst bereits gesetzten Werk.
- Änderungen am Ziel führen zur Vergabe eines neuen Permalinks. Das heißt, alle Bearbeitungszustände werden vorgehalten und mit einem je eigenen Permalink versehen. Um die Permalinks verschiedener Auflagen zu unterscheiden, bietet sich eine Angabe des Ausgabedatums im Format ISO 8061 an, z.B. „.../regesten/342/2016-02-03“. Dieser Maßnahme entspricht in der Druckwelt das Veröffentlichen mehrerer Auflagen.

Es versteht sich von selbst, dass die älteren Versionen und die darauf verweisenden Permalinks verfügbar bleiben müssen, will man nicht das Vertrauen in die eigenen Permalink-Leitlinien untergraben.

Besonders für Dokumente, die sich häufiger ändern, kann es sinnvoll sein, auch Links anzubieten, die immer auf die jeweils neueste Version eines Dokuments verweisen (sogenannte „kanonische Links“, siehe unten). Aber diese Links sind dann keine Permalinks im strengen Sinne. Ein gutes Beispiel für dieses zweigleisige Verfahren bietet Wikipedia <de.wikipedia.org>. Unter dem Artikel-Link findet man immer die aktuellste Fassung eines Artikels. Zum Zitieren verwendet man aber den Permalink, den man über eine gesonderte Funktion im Werkzeugmenü aufrufen muss.

Es ist diskutabel, ob auch minimale Veränderungen (z.B. die Korrektur einzelner Tippfehler) eines Dokuments immer einen neuen Permalink erfordern. Ebenfalls diskussionswürdig ist, ob das Dokument unter ein- und demselben Permanentlink Veränderungen des Erscheinungsbildes erfahren darf oder nicht. Solange sich keine zitierrelevanten Informationen (z.B. Seitenzahlen oder Absatznummerierungen) ändern, könnte man das als unproblematisch ansehen.

Eine ähnliche Grauzone bilden Permalinks, die auf eine Vorschaltseite, z.B. mit weiteren bibliographischen Informationen, verweisen, von der man dann zu dem gemeinten digitalen Zielobjekt gelangt. Zumindest wenn die -- am besten öffentlich zugänglichen -- Permalink-Leitlinien deutlich werden lassen, worauf genau sich die Permanenz bezieht, könnten Änderungen an der Vorschaltseite als vertretbar angesehen werden. Die Vorschaltseite wäre dann -- anders als der von ihr beschriebene Gegenstand -- strenggenommen nicht unter dem Permalink zitierbar. Allerdings sollte sich die digitale Ressource, auf die die Vorschaltseite verweist, nicht ändern.

Die Frage der Identität eines digitalen Dokuments ist insgesamt eine einigermaßen komplexe Frage, da die Identitätsrelation sowohl von der Art des Dokuments als auch dem (antizipierten) Verwendungszweck abhängen kann. Bit-Identität als vermeintlich neutrales Identitätsmaß dürfte in den meisten Fällen zu einschränkend sein, da sich die Bit-Identität z.B. schon bei der Umwandlung von einer älteren auf eine neuere PDF-Version ändern kann. Darüber hinaus gibt es bestimmte Eigenschaften digitaler Dokumente, die selbst jenseits der Bit-Identität flüchtige Eigenschaften sind, wie z.B. der Zeilenumbruch eines HTML-Dokuments.

Es ist auch denkbar, dass hier von unterschiedlichen Institutionen mit jeweils voller subjektiver Berechtigung unterschiedliche Maßstäbe angelegt werden: Eine Bibliothek, die wissenschaftliche Sekundärliteratur wie z.B. Fachartikel in digitaler Form bereitstellt, wird sich vermutlich keiner Kritik aussetzen, wenn sie das Erscheinungsbild der im Internet bereit gestellten Artikel stillschweigend an neue technische Gegebenheiten, etwa neue Geräteklassen oder Bildschirmauflösungen, anpasst, ohne deshalb einen neuen Permalink zu vergeben. Bei einem Archiv dagegen, das dieselben Werke nun nicht mehr als Sekundärliteratur, sondern als Primärquellen für die künftige historische Forschung speichert, könnte es schon eher erwünscht sein, dass es auch das Erscheinungsbild der digitalen Archivalien bewahrt. Explizite und öffentlich zugängliche Permalink-Leitlinien können hier auf jeden Fall Klarheit schaffen.

3.2.2 Permalinks müssen mehr-eindeutig, aber nicht unbedingt ein-eindeutig sein

Ein- und derselbe Permalink sollte immer auf ein- und dasselbe Dokument verweisen. Aber umgekehrt schadet es nicht, wenn zu ein- und demselben Dokument unterschiedliche Permalinks führen, denn die Wiederauffindbarkeit referierter digitaler Dokumente wird dadurch nicht eingeschränkt. Permalinks sind also nicht notwendigerweise ein-eindeutig.

Mehrfache Permalinks auf dieselbe Ressource können auf ganz natürliche Weise zustande kommen, wenn ein- und dasselbe Dokument bei unterschiedlichen Institutionen gespeichert wird, die jeweils eigene Permalinks ausgeben. Es kann auch dazu kommen, wenn Permalinks „geändert“ werden, z.B. um sie an ein neues Namensschema anzupassen. Denn, wenn überhaupt, dann können Permalinks nur dadurch „geändert“ werden, dass ein neuer Permalink hinzukommt, da der alte Permalink zum Erhalt der Permanenzeigenschaft unbedingt beibehalten werden sollte. Man kann höchstens in Zukunft auf die weitere Verwendung und Bekanntgabe des älteren Permalinks verzichten.

Für die Nutzerinnen und Nutzer von Permanent-URLs bedeutet das, dass sie sich nur anhand der Permalinks nicht sicher sein können, ob zwei verschiedene Permalinks auf unterschiedliche Dokumente oder auf ein- und dasselbe Dokument verweisen. Für die ein-eindeutige Identifizierung einer Publikation sind vielmehr die vollständigen bibliographischen Angaben erforderlich (die man ggf. durch Aufrufen eines der Permalinks auf das Dokument in Erfahrung bringen kann). Dass auf die vollständigen bibliographischen Angaben (nicht nur aus diesem Grund) in wissenschaftlichen Publikationen ohnehin nicht verzichtet werden kann, bedeutet aber auch, dass es nicht erforderlich ist, dass Permlinks in ihrer URL Hinweise auf den Urheber oder Herausgeber codieren müssten, was manchmal als kritischer Einwand gegen die Zitierbarkeit von Kurz-URLs angeführt wird. Die Verwendbarkeit von Kurz-URLs wiederum

hängt lediglich davon ab, ob auch der verwendete URL-Kürzungsdienst ebenfalls die Permanenz der Kurz-URLs garantiert (siehe unten).

Ob durch die Zwischenschaltung von Linkauflösungssystemen (wie z.B. DOI oder PURL) und die konsequente Verwendung von -- sofern verfügbar -- linkauflöser-basierten URLs eine relativ größere Ein-eindeutigkeit erreicht werden kann, ist fraglich. Da es mehrere solcher Systeme gibt, vervielfältigt sich die Mehrdeutigkeit nur auf einer anderen Ebene. Zudem müssten die Linkauflöser Sorge tragen, dass keine Dubletten vorkommen. Schon allein um die Fehlerkorrektur zu ermöglichen, wird sich die Vergabe mehr-eindeutiger Permalinks nicht hundertprozentig vermeiden lassen.

4 Zur Vermeidung von Missverständnissen: Was sind keine Permalinks?

Nachdem wir eben erläutert haben, durch welche wesentlichen Eigenschaften sich Permalinks auszeichnen müssen, wollen wir im Folgenden Permalinks gegen andere Arten von Links und andere Verwendungsweisen des Ausdrucks „Permalink“ abgrenzen, mit denen sie leicht verwechselt werden können.

4.1 Direktlinks

Als Direktlinks verstehen wir Netzadressen (URLs), die unmittelbar auf eine bestimmte Ressource, etwa einen Blogbeitrag oder Datensatz, innerhalb eines Blog oder Content-Management-Systems oder einer Webapplikation führen. „Unmittelbar“ bedeutet hier ohne den nochmaligen Umweg über die Einstiegsseite oder eine Suchmaske. Leider werden Direktlinks manchmal auch als „Permalinks“ bezeichnet, obwohl sie das höchstens in einem sehr eingeschränkten technischen Sinne sind. So generiert beispielsweise das Blogsystem „Wordpress“ unter der Bezeichnung „Permalink“ Direktlinks auf die einzelnen Blogbeiträge. Um echte Permalinks handelt es sich aber nur dann, wenn diejenigen, die den Blog betreiben, die Links selbst und die darunter veröffentlichten Beiträge nicht nachträglich ändern und sich dazu in ihren Richtlinien verpflichten. Ein Blog- oder Content-Management-System, das eine Permanentlink-Funktion anbietet, stellt damit also lediglich Direktlinks als technische Voraussetzung für echte Permalinks bereit.

Taucht der Begriff in der Dokumentation eines Blogsystems auf, so ist das vielleicht auch nicht anders zu erwarten. Aber auch in anderen Kontexten wird der Ausdruck „Permalink“ manchmal lediglich als Synonym für einen Direktlink verwendet. So sind beispielsweise die als solche bezeichneten Permalinks im bayerischen Verbundkatalog wie gateway-bayern.de/BV003390811 keine Permalinks in unserem Sinne, denn ihre Permanenz ist nicht garantiert. Vielmehr können sie im Rahmen der Konsolidierung des Bibliothekskatalogs, namentlich der Zusammenführung von Dubletten, verschwinden und damit ins Leere führen (dazu unten mehr).

4.2 Kurzlinks

Relativ offensichtlich ist es, dass Permalinks von Kurzlinks zu unterscheiden sind, wie sie von

Diensten wie tinyurl.com, goo.gl oder bit.ly bereit gestellt werden. Technisch gesehen beruhen Kurzlinks auf ganz ähnlichen Mechanismen wie die Linkauflöser für Permalinks, aber ihr Sinn und Zweck ist ein anderer. So sind Kurzlinks ihrer Kürze wegen vor allem für Twitter oder Handy-Nachrichten hilfreich. Zum Zitieren eignen sich Kurz-URLs von kommerziellen Diensteanbietern eher nicht, da -- selbst wenn sie eine Langlebigkeitsgarantie abgeben -- nicht sicher ist, wie lange der Dienst sich am Markt halten kann. Auch Großunternehmen, bei denen ein schnelles Verschwinden vom Markt nicht zu erwarten ist, neigen in dieser Branchen dazu als unrentabel empfundene Geschäftsbereiche vom einen auf den anderen Tag aufzugeben.

Etwas Anderes ist es, wenn wissenschaftliche Informationsdienstleister eigene Kurzlink-Dienste mit Permalink-Leitlinien anbieten. Dabei müsste allerdings sichergestellt werden, dass die zu kürzenden Netzadressen Permalinks repräsentieren. Am einfachsten lässt sich das realisieren, wenn der Dienst als zu kürzende Links nur permanente URLs innerhalb der eigenen Domäne der bereitstellenden Institution zulässt. Unabhängig von diesen Überlegungen ist es in jedem Fall sinnvoll, die URLs von Permalinks eher kurz zu halten, damit sie nicht die Zeilenlänge in einem Buch oder Aufsatz überschreiten und ggf. bequem abgetippt werden können.

4.3 Semantische Links

Permalinks sind nicht dasselbe wie semantische Links. Unter „semantischen Links“ versteht man aussagekräftige Netzadressen, die in der URL bereits für Menschen sinnvolle und verständliche Angaben zu der damit adressierten digitalen Ressource enthalten (z.B. de.wikipedia.org/wiki/Der_Schmied_von_Koche!). Zu der Verwechslung kommt es deshalb so leicht, da semantische Links sich, gerade weil sie so gut lesbar sind, zum Zitieren so schön anzubieten scheinen. Es ist zwar möglich und auch sinnvoll, für Permalinks aussagekräftige URLs vorzusehen. Im Zweifelsfall ist die Permanenz aber wichtiger als die Aussagekräftigkeit -- und wird von ihr leider weit öfter gefährdet, als zunächst scheinen mag. Beispiele für Angaben, deren Verwendung in einem Permalink Schwierigkeiten machen kann:

- Eine historische Datierung (z.B. einer Urkunde) kann sich im Laufe der Forschung ändern.
- Titel sind nicht sicher eindeutig; und es kann sich der für ein titellostes historisches Werk eingeführte Titel ändern.
- Ein entsprechender Fall: Der einem historischen Werk zugeschriebene Autor kann sich ändern.
- Personennamen-GNDs sind nicht sicher eindeutig. Personen-GNDs wiederum gefährden die Beständigkeit insofern, als sie falsch zugeordnet sein können und dann geändert werden müssten.
- Lemmaansätze sind nicht sicher eindeutig (wegen möglicher Homonymie) und müssen möglicherweise geändert werden (z.B. bei Vereinheitlichungen in der Lautwiedergabe oder der Behandlung von Fugenelementen). Vorzuziehen sind eigene Kennungen, die man seinen Lemmaansätzen zuordnet.

- Sogar Bibliothekssignaturen können sich ändern oder verschwinden. Verwendet man die früheren Signaturen weiter, entsteht eine Mischung aus semantischen und scheinbar semantischen Kennungen (veralteten Signaturen). Noch ungünstiger ist es, wenn die Bibliothek eine veraltete Signatur neu verwendet.

Grundsätzlich sind Angaben, die möglicherweise einmal geändert werden müssten, kaum permalinkgeeignet, es sei denn, dass man bei solchen Änderungen eine neue Auflage veröffentlicht und dieser geänderten Auflage ohnehin einen neuen Permalink erteilt.² Man kann auch dulden, dass ein Permalink veraltete Angaben enthält; allerdings ist er dann nicht länger aussagekräftig, sondern sogar geradezu irreführend geworden.

Daher ist es unter Umständen vorzuziehen, statt bedeutungstragender Angaben von vornherein völlig bedeutungslose und daher nie veraltende IDs zu verwenden. So verfahren denn auch die öffentlichen Linkauflösersysteme wie z.B. DOIs (siehe dazu weiter unten die Diskussion von Linkauflösern).

4.4 Kanonische Links

Mindestens ebenso leicht kann es zur Verwechslung von Permalinks mit kanonischen Links kommen. Ein kanonischer Link ist die im Zweifelsfalle vorzuziehende URL eines elektronischen Dokuments. Die anderen Fassungen bieten z.B. denselben Inhalt an (nur unter einer anderen URL) oder denselben Hauptinhalt mit z.B. einem Menü in anderer Sprache.³

Häufig, aber nicht immer, verweist ein kanonischer Link auf die aktuellste Fassung einer elektronischen Publikation. (Manchmal ist die aktuellste Fassung noch eine beta-Version oder eine Diskussionsvorlage, die erst später finalisiert wird und deshalb gerade nicht kanonisch ist.) Kanonische Links spielen vor allem bei Nachschlagewerken eine Rolle, deren Beiträge mit der Zeit überarbeitet werden und bei denen die Nutzer so gut wie immer an der aktuellsten Fassung interessiert sind.

Außer der schon erwähnten Wikipedia liefert die Stanford Encyclopedia of Philosophy ein gutes Beispiel. Unter dem kanonischen Link (z.B.: <plato.stanford.edu/entries/aristotle/>) ist die jeweils aktuellste Fassung eines Artikels abrufbar. Für jeden Artikel existiert aber auch ein Permalink (z.B.: <plato.stanford.edu/archives/fall2015/entries/aristotle/>), der durch die Datumsangabe in der URL bereits als solcher erkennbar ist. Eine Zitationshilfe liefert die korrekten bibliographischen Angaben mit Permalink. Die Permalink-Leitlinien werden auf einer Dokumentationsseite (<plato.stanford.edu/cite.html>) in angemessen knapper Form (eine Bildschirmseite) erläutert.

Die klare Unterscheidung zwischen Permalinks und kanonischen Links wird auch dadurch erschwert, dass der Ausdruck „kanonischer Link“ manchmal fälschlich im Sinne von „Permalink“

² Dazu oben im Abschnitt „Ein Permalink muss immer zu demselben Ziel führen“. Die dort erwähnte Datierung steht anders als die oben gemeinte historische Datierung (zum Beispiel einer Urkunde) nicht in der Gefahr zu veralten.

³ Für Suchmaschinen sind als kanonisch ausgezeichnete Links der Verweis auf eine Seite, die sie für ihre Indizierung der vorliegenden Seite vorziehen sollen. Vgl. M. Ohye (2012).

verwendet wird.⁴

4.5 Andere permanente Identifikatoren

Schließlich sollten Permalinks nicht mit anderen Arten von permanenten Identifikatoren verwechselt werden, die nicht durch eine Netzadresse dargestellt werden. Anders als bei den zuvor besprochenen Linktypen ist die Verwechslung in diesem Fall aber weniger schädlich, da andere permanente Identifikatoren weitgehend dieselben Funktionen erfüllen können wie Permalinks. Insbesondere können sie ebenso wie Permalinks mittels des Identifikators die Zitierbarkeit von elektronischen Dokumenten ermöglichen -- sofern sich dem Identifikator jederzeit ohne große Mühe eine Netzadresse zuordnen lässt. Darin sehen wir einen, wenn auch nur geringfügigen Nachteil von permanenten Identifikatoren, die keine Permalinks sind. Ein Permalink kann immer so, wie er ist, in die Adresszeile eines Internetbrowsers eingegeben werden und führt dann zu der verknüpften Ressource. Ein permanenter Identifikator wie z.B. eine DOI erfordert einen Webservice, der einem zu dem Identifikator eine Netzadresse liefert oder auch gleich auf die dem Identifikator zugeordnete Netzseite weiterleitet.

5 Sind permanente Identifikatoren (z.B. DOI) eine bessere Alternative?

Seltsamerweise erfreuen sich einige der eben beschriebenen alternativen Identifikatoren, insbesondere die DOI-Kennungen („Digital Object Identifier“) eines gewissen Prestiges, so, als würde ein digitales Dokument erst durch die Verleihung einer DOI zu einer respektablen elektronischen Publikation. Diese Prestigezuschreibung ist unserer Ansicht nach in der Sache nicht begründet und wohl eher durch die Analogie zur ISBN von Druckpublikationen zu erklären, deren Prestige als vermeintliches Mindestqualitätsmerkmal einer wissenschaftlichen Buchpublikationen allerdings auf ähnlichen Missverständnissen beruht. Denn sowohl technisch als auch administrativ ist die DOI-Kennung lediglich ein vorgeschaltetes Linkauflösersystem wie andere auch, z.B. PURL. Linkauflösersysteme (DOI, PURL, etc.) sind im Prinzip nichts anderes als Verzeichnisse, in denen einem digitalen Objekt-Identifikator ein Link auf das Objekt zugeordnet wird. Sie garantieren nicht und können auch nicht garantieren, dass dieser Link gültig bleibt. Dies hängt wiederum allein von der Institution ab, die die Seite hinter dem Link bereitgestellt hat. Ein möglicher Vorteil ist, dass DOI, PURL, URN nicht von der oben beschriebenen Mehrdeutigkeit des Wortes „Permalink“ betroffen sind. Andererseits sollte man sich sowieso nie auf die bloße Bezeichnung verlassen, sondern auf die ausdrücklichen, veröffentlichten Permalink-Leitlinien achten (siehe oben, Abschnitt 3).

Der Nutzen von Linkauflösersystem besteht darin, dass die Adresse des Zielobjekts stillschweigend geändert werden kann und der Link immer noch funktioniert. Sofern man annimmt, dass das Linkauflösersystem von größerer Dauerhaftigkeit ist als die Permalinks der bereitstellenden Institutionen, bietet das einen Vorteil. Das könnte z.B. dann der Fall sein, wenn die Permalinks der bereitstellenden Institution einen Domain-Namen enthalten, der durch Verkauf oder Abtretung an eine andere Institution übergeht und dann nicht mehr verwendet

⁴ So bei Van de Sompel et al. (2016a, 2), wenn sie den Attributwert „canonical“ für das „rel“-Attribut des Verknüpfungstags vorschlagen, obwohl sie eigentlich einen Permalink kennzeichnen wollen.

werden darf. Es ist aber auch nur dann der Fall, wenn der neue digitale Ort des Zielobjekts an das Linkauflösersystem zurückgemeldet wird. Entsprechende Leitlinien dürften sich aber nur schwer durch den Betreiber des Linkauflösersystems erzwingen lassen, weil gerade der häufigste Fall, in dem dies nötig wäre, die Auflösung einer Institution ist. Es hängt nicht vom Betreiber des Linkauflösers, sondern von dessen Kunden ab, ob durch die Entkoppelungsfunktion eine größere Permanenz der (entkoppelten) digitalen Identifikatoren gegenüber selbsterstellten Permalinks der Kunden möglich ist.

Wir haben daher Zweifel, ob die digitalen Identifikatoren einer zentralen Organisation wie der DOI-Foundation tatsächlich eine größere Permanenz verbürgen können als gewöhnliche Permalinks. Theoretisch wird die Permanenz durch die Einführung eines zusätzlichen Zwischenschrittes zunächst einmal geschwächt, da sie nun von zwei Institutionen, dem Linkauflöser und dem Bereitsteller der digitalen Ressource, abhängig ist. Sollte dieser Nachteil durch die Entkoppelungsfunktion hinreichend kompensiert werden können, dann würden wir, schon um Bürokratie und Abhängigkeiten zu vermeiden, ihre Realisierung durch eine Form von peer-to-peer-System bevorzugen.⁵ Ein weiterer Grund ist, dass das (geschäftliche) Eigeninteresse entsprechender Institutionen nicht immer mit den Interessen der Nutzer harmonisieren muss.⁶

Insgesamt sind wir aber der Überzeugung, dass Identifikatoren wie DOI, PURL, etc. nicht grundsätzlich besser oder zitierfähiger sind als jeder andere Permalink. Dazu ein Zitat von Tim Berners-Lee (dem Vater des WWW):

„This is [...] probably one of the worst side-effects of the URN discussions. Some seem to think that because there is research about namespaces which will be more persistent, that they can be as lax about dangling links as they like as "URNs will fix all that". If you are one of these folks, then allow me to disillusion you.

Most URN schemes I have seen look something like an authority ID followed by either a date and a string you choose, or just a string you choose. This looks very like an HTTP URI. In other words, if you think your organization will be capable of creating URNs which will last, then prove it by doing it now and using them for your HTTP URIs. There is nothing about HTTP which makes your URIs unstable. It is your organization. Make a database which maps document URN to current filename, and let the web server use that to actually retrieve files.“ (Berners-Lee 1998)

6 Permalinks in der Praxis

6.1 Permalinks und Zitierbarkeit

Die Frage der Zitierbarkeit von digitalen Ressourcen hängt von ihrer Identifizierbarkeit und Wiederauffindbarkeit ab. Permalinks sind ein Mittel, um digitale Ressourcen zu identifizieren

5 So strebt das beispielsweise auch die „Permanent Identifier Community Group“ unter dem Dach des World-Wide-Web-Konsortiums an <www.w3.org/community/perma-id/2013/02/16/launch/>.

6 Vgl. zu Kosten und Geschäftsmodellen etwa den Abschnitt über die „Registration Agencies“ im DOI-Handbuch (DOI-Foundation 2016). Vgl. auch den Clarin-Report zu „Persistent and Unique Identifiers“ von 2008 (Broeder et al. 2008), der das durchaus kritisch sieht.

und wiederzufinden, und damit den Nachvollzug einer wissenschaftlichen Argumentation, die sich auf andere Quellen stützt, nachzuvollziehen. Weitere Mittel sind aber zumindest denkbar.⁷ Insofern können digitale Ressourcen auch ohne Permalink als zitierbar gelten, wenn die im zitierenden Werk gegebenen Angaben dazu ausreichen, um sie ohne zu großen Aufwand zu identifizieren und wieder zu finden. Ist aber ein Permalink zu einer digitalen Ressource vorhanden, dann wäre es im Sinne einer wissenschaftlich sorgfältigen Zitierweise eine Nachlässigkeit, keinen Permalink oder anderen permanenten Identifikator anzugeben.

Die früher zeitweise übliche Praxis, beim Verweis auf Webseiten ein Zugriffsdatum anzugeben, hat sich dagegen überlebt. Wenn eine digitale Ressource nicht mehr existiert, dann hilft auch der Hinweis auf einen in der Vergangenheit möglichen Zugriff den Rezipientinnen und Rezipienten nicht weiter. Lediglich bei der Zitation von „lebenden Dokumenten“ (siehe unten) kann das Zugriffsdatum als redundante Kennzeichnung der zitierten „Zeitstelle“ hilfreich sein.

Anders verhält es sich mit dem Anlegen eigener Kopien von Internetquellen. Zumindest, wenn es um wissenschaftliche Primärquellen geht, kann es manchmal unvermeidlich sein, flüchtige Internetquellen zu zitieren. Das wäre z.B. bei einer Forschungsarbeit über die Bewertung der Flüchtlingskrise in der Blogsphäre der Fall, denn nicht jeder Internetblog bietet Permalinks an, und darüber hinaus erfüllt die Permalinkfunktion gängiger Blogsysteme nicht automatisch schon wissenschaftliche Ansprüche an die Persistenz von Inhalten -- von der Frage ganz zu schweigen, wie sehr man darauf vertrauen kann, dass Blogbeiträge oder Beiträge zu Diskussionsforen nicht nachträglich verändert werden. In solchen Fällen ist es nach wie vor hilfreich, ein eigenes Forschungsarchiv anzulegen, in dem man Kopien der nicht mit (vertrauenswürdigen) Permalink versehenen digitalen Ressourcen sammelt, oder auf einen Archiv-Anbieter wie z.B. <http://archive.org> zurückzugreifen. Wie oben bereits erwähnt, ist es bei letzteren leider nicht immer leicht, sich über die Permalink-Leitlinien und relevante Details der technischen Umsetzung (z.B., inwieweit Unterseiten und verlinkte Inhalte mitarchiviert werden) Klarheit zu verschaffen. Sollte der Bedarf groß genug sein, erscheint es auch denkbar, dass die Informationsdienste wissenschaftlicher Institutionen in Zukunft entsprechende Archivierungsservices für digitale Quellensammlungen als Cloud-Service anbieten.

Wie weiter oben beschrieben wurde, kann es je nach Striktheit der Permalink-Richtlinien vorkommen, dass unter einem Permalink erreichbare digitale Ressourcen immer noch bestimmte flüchtige Eigenschaften aufweisen können. Für die Zitierbarkeit hängt es dann davon ab, ob die wissenschaftliche Argumentation, in deren Zusammenhang auf die digitalen Ressource verwiesen wird, sich maßgeblich auf diese flüchtigen Eigenschaften stützt. Für diesen Ausnahmefall bietet sich als Ausweg ebenfalls die Eigenarchivierung an.

Ebenfalls relevant für die Zitierbarkeit von digitalen ebenso wie analogen Primärquellen ist die Frage ihrer Echtheit. Bei digitalen Quellen verschärft sich die Frage insofern noch, als digitale Quellen sich im Prinzip sehr leicht nachträglich ändern lassen. Zumindest bei sensiblem Material, etwa archivierten Regierungsakten, könnte es daher in Zukunft notwendig werden, die Permalinks mit den verlinkten Dokumenten kryptographisch abzusichern, um die

⁷ Zur Frage, wie man Online-Quellen zitiert, vgl. die ausführliche und schöne Zusammenstellung von Graf (2011).

Vertrauenswürdigkeit zu ermöglichen.

Von der Zitierbarkeit im weiteren Sinne ist die Zitierfähigkeit im engeren Sinne zu unterscheiden. Die Zitierfähigkeit bezieht sich nur auf wissenschaftliche Sekundärliteratur. Gemeint ist damit, dass die eigene wissenschaftliche Argumentation sich nicht auf Forschungsergebnisse aus dubiosen Quellen stützen sollte. Als zitierfähig gilt daher in der Regel nur diejenige Sekundärliteratur, die einen den wissenschaftlichen Gepflogenheiten entsprechenden Qualitätssicherungsprozess wie z.B. das bei Fachjournalen übliche Begutachtungsverfahren („peer review“) durchlaufen hat. Die Zitierbarkeit im Sinne der Identifizierbarkeit und Wiederauffindbarkeit ist eine notwendige, aber keine hinreichende Bedingung für die Zitierfähigkeit. Insofern sollten digitale Publikationen, um zitierfähig zu sein, am Besten über einen Permalink verfügen. Zitierfähig als Sekundärliteratur wird die verlinkte Publikation aber nur durch ihre wissenschaftliche Qualität, nicht durch das Vorhandensein eines Permalinks.

6.2 Permalinks auf Datenbanken und lebende Dokumente

Während Druck-Publikationen nach den Fahnenkorrekturen finalisiert sind, besteht bei digitalen Dokumenten und Daten die Möglichkeit, dass sie -- solange sie verwendet werden -- auch kontinuierlich geändert werden. Man kann sogar soweit gehen zu behaupten, dass gerade die Auflösung der scharfen Trennung zwischen Erstellungs- und Nutzungsphase neue Möglichkeiten für den wissenschaftlichen Forschungsprozess eröffnet. Neben der von Hubertus Kohle prägnant beschriebenen „publish first -- filter later“-Strategie wären hier vor allem Wissenschaftsblogs zu nennen, die nicht nur den Austausch und die Wissenschaftskommunikation im Vorfeld späterer Fachveröffentlichungen verändern, sondern sich -- wie das Polymath-Projekt (Gowers et al. 2016; Boulton 2012) vor Augen führt -- tatsächlich auch schon als Mittel der massiv kollaborativen Forschung bei der Lösung schwieriger wissenschaftlicher Probleme bewährt haben, die im Rahmen der traditionellen Organisationsform der Wissenschaft in derselben Zeit kaum hätten gelöst werden können.

Für digitale Dokumente, die nach ihrer Veröffentlichung noch verändert werden, hat sich die Bezeichnung „lebende Dokumente“ eingebürgert. Wenn lebende Dokumente -- wie das Beispiel des Polymath-Projekts zeigt -- bei der Genese wissenschaftlicher Erkenntnis eine tragende Rolle spielen können, dann ist es wichtig, sie zitierbar zu gestalten. Ein wissenschaftspolitisches Eigentor wäre es dagegen, wenn man umgekehrt lebenden Dokumenten wegen ihres veränderlichen Charakters die Zitierfähigkeit kategorisch absprechen wollte. Für die Zitierbarkeit lebender Dokumente ist es erforderlich, Permalinks bereit zu stellen, die auf den Zustand eines lebenden Dokuments zu einem bestimmten Zeitpunkt verweisen. Eine Erläuterung, wie das technisch realisiert werden kann, würde an dieser Stelle zu weit führen. Wir möchten deshalb nur darauf hinweisen, dass dafür in Form von versionierenden Datenbanken und den aus der Softwareentwicklung bekannten Versionsverwaltungssystemen wohlbewährte Lösungen existieren.

Ähnliches gilt für die zunehmend wichtiger werdenden Forschungsdatenbanken. In Zukunft werden auch in den Geisteswissenschaften selbst die Forschungsergebnisse nicht mehr nur

Textform haben, sondern teilweise auch in Form von Daten vorliegen. Damit wird es aber auch erforderlich, die entsprechenden Datenbanken zitierbar zu gestalten. Da Zitierungen eine der Währungen sind, in denen wissenschaftliche Anerkennung ausgetauscht wird, besteht ein sehr erwünschter Nebeneffekt darin, dass auf diese Weise auch die Arbeit der Recherche, Sammlung und Erarbeitung von wissenschaftlichen Datenbeständen als wissenschaftliche Leistung gewürdigt und ihren Urhebern zugeschrieben werden kann.

Die Bereitstellung von Permalinks für Datenbanken ist eine etwas komplexere Angelegenheit als die für lebende Dokumente, da hier nicht nur der zeitliche Verlauf zu berücksichtigen ist, sondern auch, dass sich die für eine wissenschaftliche Argumentation relevante „Datenbankstelle“ (in Analogie zur zitierten Textstelle) in der Regel aus bestimmten Datenbankabfragen ergibt. Ein Permalink auf die gesamte Datenbank wäre zu grobkörnig, ein Link auf einen einzelnen Datensatz zu feinkörnig, da in einer Abfrage ja in der Regel Daten aus mehreren Datensätzen (bei relationalen Datenbanken: mehreren Zeilen aus unterschiedlichen Tabellen) zusammengefasst werden. Inzwischen existieren aber auch hierfür durchdachte Konzepte, wie man die Zitierbarkeit von Datenbanken gestalten kann (Rauber et al. 2015; Rauber et al. 2016). So ist es im Prinzip möglich, Permalinks für konkrete Datenbankabfragen dynamisch zu erzeugen. Die technische Voraussetzung dafür ist, dass sich eine Abfrage über den Zustand der Datenbank zu einem bestimmten Zeitpunkt jederzeit rekonstruieren lässt. Das ist mit versionierenden Datenbanken leicht möglich, sofern neben der Datenbank noch ein Verzeichnis der Abfragen und der ihnen zugeordneten Permalinks geführt wird.

Die Frage der „Granularität“ von Permalinks stellt sich nicht nur bei Datenbanken, sondern bereits bei Texten. Bei Texten kann es jedoch genügen, wenn der Permalink auf ein Dokument von vielen hundert Seiten verweist und die Seitenzahl lediglich im Zitat genannt wird -- wobei eine Verknüpfung, die direkt auf eine bestimmte Seite führt, natürlich die Bequemlichkeit erhöht. Bei Datenbanken erscheint es dagegen unpraktikabel, die konkrete Abfrage, die die Gestalt von vielen Zeilen Programmcode in einer Abfragesprache wie SQL haben kann, im Text eines wissenschaftlichen Aufsatzes anzuführen.

Des Unterschiedes zwischen Permalinks auf elektronische Dokumente und Permalinks auf Datenbankabfragen sollte man sich bewusst sein. Während bei der Zitation von Permalinks auf Dokumente die bibliographischen Angaben das Dokument beschreiben und Permalinks lediglich den Ort des Dokuments im Internet angeben, ist bei Permalinks auf Datenbanksuchen nach dem eben beschriebenen Konzept die Suchanfrage und damit die Beschreibung des referenzierten Objekts im Permalink versteckt. Ein zerbrochener Permalink würde es daher schwerer machen, die Referenz zu rekonstruieren als im Falle eines elektronischen Dokuments. Die Praxis muss zeigen, ob das zu Problemen führt. Möglicherweise wird sich dann die Konvention herausbilden, die konkreten Suchanfragen im Anhang oder in den Ergänzungen (den sogenannten „supplements“), die bei elektronisch publizierten Artikeln ohnehin zunehmend üblich werden, zu dokumentieren.

6.3 Die Abwicklung von Permalinks

Die eigentlich etwas paradox erscheinende Abwicklung von Permalinks kann in zwei Fällen

erforderlich werden. Der erste Fall ist der, dass das digitale Objekt, auf das der Permalink verweist, einen Lebenszyklus hat und dessen Ende erreicht ist. Ein Beispiel dafür wäre ein Permalink auf einen Katalogeintrag, der im Rahmen einer Dublettenkorrektur gelöscht wird. In diesem Fall könnte die Abwicklung darin bestehen, dass der Permalink des gelöschten Eintrags nun auch auf den nicht gelöschten verweist, insoweit die Dublette in zwei Einträgen bestand, die beide dasselbe Buch verzeichnen. Wenn aber ein Katalogeintrag als falsch herauskorrigiert werden muss, könnte der Permalink auf einen Nachfolger oder Stellvertreter führen, z.B. auf eine Seite, die die Korrektur transparent macht.

Der zweite Fall ist derjenige, bei dem eine wissenschaftliche Institution, die Permalinks ausgegeben hat, aufhört zu existieren. In diesem Fall sollte ein neuer Besitzer für die Permalinks gesucht werden. Ein Problem stellt dann noch der in den Permalinks enthaltene Domänenname dar. (Das ist, wie oben beschrieben, unserer Ansicht nach der einzige Nachteil von Permalinks gegenüber anderen Arten von digitalen Identifikatoren, die Identifikator und Netzadresse entkoppeln.) Sofern der Domänenname frei wird, kann er vom neuen Besitzer übernommen und weitergeführt werden. Wichtig ist nur, dass dies im Rahmen der Abwicklung auch mitberücksichtigt wird.

Nur wenn der Domänenname an einen neuen Besitzer übergeht, der nicht bereit ist, die Permalinks weiterzuführen bzw. eine Weiterleitung einzurichten, ist das Zerschneiden der Permalinks bei der Abwicklung einer wissenschaftlichen Einrichtung unvermeidlich. Wir gehen davon aus, dass dieser Fall nur selten eintreten wird, denn in der Regel dürfte der neue Domänenbesitzer, schon um lästige Anfragen zu vermeiden, über eine Weiterleitung mit sich reden lassen. Ausschließen lässt sich dieser Fall aber nicht.

6.4 Zwei offene Fragen: Vorschaltseiten und Maschinenlesbarkeit von Permalinks

6.4.1 Sollen Permalinks auf Vorschaltseiten oder direkt auf Dokumente verweisen? Die Abwicklung von Permalinks

Eine öfters diskutierte Frage ist die, ob Permalinks besser auf Vorschaltseiten oder auf das gemeinte digitale Objekt selbst verweisen sollten.⁸ Wir haben den Eindruck, dass es zunehmend üblich wird, Permalinks nicht direkt auf die verlinkten digitalen Objekte verweisen zu lassen, sondern auf eine Vorschaltseite, die Informationen (sprich „Metadaten“) über das Objekt und den Link auf das eigentliche Objekt enthält oder mehrere Links auf unterschiedliche Repräsentationen des digitalen Objekts z.B. in unterschiedlichen Dateiformaten. Vorschaltseiten sind für die Nutzerinnen und Nutzer zunächst ein Nachteil, weil sie einen weiteren Klick und die Suche nach dem endgültigen Link erfordern. Aber die Vorteile, unterschiedliche Dateiformate, Ausprägungen (z.B. hochauflösend oder komprimiert) anbieten oder auch Metadaten oder Lizenzinformationen kenntlich machen zu können, wiegen in unseren Augen die Nachteile auf. Hinzu kommt: Führt ein Permalink unmittelbar auf ein PDF, so fehlt das für eine weitere Recherche vielleicht erwünschte Menü der das PDF bereitstellenden Seite. Besonders bei retrodigitalisiertem Material erscheint uns überdies die Einbettung einer Vorschaltseite in die

⁸ Auf die weitergehenden Implikationen dieses Themas, insbesondere auch, was den automatischen Abruf von Dokumenten in großer Menge angeht („harvesting“) können wir hier nicht eingehen. Vgl. dazu Van de Sompel et al. (2016b).

Dateien selbst als noch störender. Im Übrigen kann man der inzwischen schon häufig anzutreffenden Konvention folgen, dass der Permalink keine Dateiendung enthält und auf eine Vorschaltseite verweist, während derselbe Link mit Dateiendung direkt zum digitalen Objekt führt.

6.4.2 Maschinenlesbarkeit von Permalinks

Maschinenlesbare Metadaten sind ein Thema für sich, das hier nicht ausführlich zu besprechen ist. Wir beschränken uns auf die Frage: Wie kann ein Programm einen Permalink als solchen erkennen? Dafür wäre die Vereinbarung eines Signalwortes wie „permalink“ nützlich, das als Wert des rel-Attributes von Link-Elementen dienen könnte.⁹ Auf jener Seite selbst, zu der der Permalink führen soll, könnte man den Permalink dann einfach mit `<link href="..." rel="permalink"/>` im head-Element der Seite vermerken (wobei für „...“ der Permalink zu schreiben wäre).

Auf anderen Seiten, von denen der Permalink zur Zielseite des Permalinks führt, müsste man deutlich machen, dass es nicht die Seite ist, auf der er steht, sondern die, zu der er führt. Dafür müsste man das Subjekt der Permalink-Aussage ausdrücklich machen, z.B. im Rahmen von RDFa mit dem about-Attribut:

```
<a about="...Permalink..." href="...Permalink..." rel="permalink">...</a>
```

(Im Rahmen von RDFa nähme man übrigens statt `rel` besser `property`, worauf hier nicht weiter einzugehen ist.)

Unseres Wissens fehlt es in den großen Vokabularen wie schema.org/docs/full.html, worin Signalwörter für die maschinelle Verarbeitung vereinbart sind, noch an einem Ausdruck für 'Permalink'.

Auch in den verbreiteten Sprachen für das Verzeichnen bibliographischer Metadaten gibt es bisher nur einen Kennungstyp für Links (z.B. die Kennung „URL“ im Falle von BibLaTeX) und keine Möglichkeit, permanente und flüchtige Links zu unterscheiden.

Insgesamt bleibt die maschinenlesbare Kennzeichnung von Permalinks noch ein Desiderat.

7 Ein Praxis-Beispiel: Warum BV-Nummern nicht für Permalinks taugen

Nach den bis hierher weitgehend theoretischen Überlegungen möchten wir nun noch kurz einige praktische Erfahrungen mitteilen, die wir mit der Bereitstellung von Permalinks gemacht haben. Beim Aufbau des Publikationsservers der Bayerischen Akademie der Wissenschaften bestand unser erster Ansatz darin, für die Permalinks Kennungen zu benutzen, die identisch mit den BV-Nummern des BV-Katalogs sind. Mehrere Gründe waren dafür ausschlaggebend:

⁹ Die Idee kommt von Van de Sompel et al. (2016a, 2). Unglücklicherweise schlagen sie den Attributwert „canonical“ vor. Dieser Wert wird aber schon verwendet, und zwar um einen kanonischen Link in dem Sinne zu bezeichnen, der oben unter „Kanonische Links“ erläutert wurde. Vgl. M. Ohye (2012)

1. Alle Werke der Bayerischen Akademie der Wissenschaften werden im BV-Katalog eingetragen. Also hat jedes Werk, das auf dem Publikationsserver erscheinen kann, eine BV-Nummer.
2. Wir benutzen diese BV-Nummer ohnehin schon, um dem BV-Katalog die bibliographischen Angaben für Publikationen auf dem Publikationsserver zu entnehmen. Denn die Einträge im BV-Katalog wurden von bibliothekarischen Fachkräften vorgenommen, was eine bedeutende Qualitätssicherung darstellt, und können über die vom Katalog dankenswerterweise bereitgestellte SRU-Schnittstelle automatisch entnommen werden, so dass die Angaben nur einmal menschliche Arbeitskraft erfordern.
3. Wir gingen davon aus, dass eine einmal vergebene BV-Nummer auf unabsehbare Zeit zu einer Beschreibung der Publikation führt, für die sie vergeben wurde.
4. Durch die Nutzung von BV-Nummern wären unsere Permalinks indirekt in einen größeren Kontext eingeordnet gewesen, nämlich den des BV-Katalogs und hätten, indem sie die Katalognummer enthalten, zugleich eine gewisse Semantik gehabt.

Aus zwei Gründen ließ sich dieses Vorhaben nicht ganz durchhalten:

1. Die BV-Nummern sind primär Identifikatoren für die Datenbankeinträge im BV-Katalog, nicht für die in den Datenbankeinträgen beschriebenen Werke. Wird nun ein Datenbankeintrag auf einen anderen umgelenkt -- was nötig sein kann, um Dubletten zusammenzuführen --, dann verschwindet die BV-Nummer des umgelenkten Eintrags und mit ihr auch der Permalink des BV-Kataloges. (Dies nach einer brieflichen Auskunft von M. Kratzer von der BSB, für die wir hier herzlich danken.) Zum Beispiel hatte das Werk BV020097138 (siehe gateway-bayern.de/BV020097138) früher auch die BV-Nummer BV003899554, die der Katalog jetzt nicht mehr kennt; weitere Beispiele für verschwundene BV-Nummern sind BV023335674 und BV023335670.
2. Gelegentlich ist es nur für ausgebildete Bibliothekare und Bibliothekarinnen sicher erkennbar, welchem Katalogeintrag und damit welcher BV-Nummer eine Publikation zuzuordnen ist. Das gilt besonders im Fall von Dubletten, die der BV-Katalog schon deswegen nicht so selten enthält, weil er aus mehreren, früher unabhängigen Katalogen zusammengeführt wurde. Die Situation, wie sie ist, führt dazu, dass ein Rückstau von Publikationen entsteht, die bloß deshalb nicht auf dem Publikationsserver veröffentlicht werden können, weil ihre BV-Nummern noch nicht kontrolliert und gegebenenfalls korrigiert sind.

Wir haben daraus gelernt und generieren für den Publikationsserver der Akademie (publikationen.badw.de) inzwischen unsere eigenen Permalinks, deren Permanenz bis auf Mehr-Eindeutigkeit (siehe oben) von uns garantiert wird.

8 Schlussbemerkung

Die vorhergehende Diskussion von Permalinks könnte auf Grund ihrer Ausführlichkeit den Eindruck erwecken, dass Permalinks eine komplizierte Angelegenheit sind. Das ist nicht unsere Ansicht. Vielmehr glauben wir, dass sich Permalinks, wenn man ihren Zweck richtig verstanden hat, sehr einfach handhaben lassen -- sowohl was deren Bereitstellung als auch deren Nutzung angeht. Das Wesentliche dazu haben wir am Anfang in ein paar Stichpunkten zusammengefasst.

Die Ausführlichkeit der Diskussion liegt darin begründet, dass es auf Grund der Neuheit des Konzepts wie bei so vielen Begriffen der digitalen Welt leicht zu Missverständnissen kommen kann, die sich -- wie uns unsere eigenen Erfahrungen mit den BV-Nummern gezeigt haben -- auch durchaus in der Praxis auswirken können. Dieser Unsicherheiten wegen lohnt sich eine ausführliche Diskussion.

Hinzu kommt, dass mit neuen Publikationsformen wie lebenden Dokumenten und Datenbanken als Publikationen Permalinks eine zunehmende Bedeutung im Kontext der „reproduzierbaren Wissenschaft“ gewinnen. Dabei erfahren Permalinks auch eine Funktionserweiterung. Sie dienen nicht mehr nur als Ortsangabe für ein elektronisches Dokument, sondern als Stellenangabe innerhalb eines digitalen Objekts. Umso größer dürften in Zukunft daher auch die Anforderungen an die Verlässlichkeit von Permalinks werden. Wir sind aber zuversichtlich, dass sich im Laufe der Zeit und mit zunehmender Sensibilisierung für das Thema auf Seiten der Anbieter Usancen heraus bilden, die dazu führen, dass man -- anders als in den Anfangstagen des Internets -- dauerhafte von flüchtigen Links leichter unterscheiden und sich auf die Dauerhaftigkeit von Permalinks zunehmend verlassen kann.

9 Quellen

Berners-Lee, Tim (1998): Cool URIs don't change, [online], verfügbar unter: www.w3.org/Provider/Style/URI.html .

Boulton, Geoffrey (Hrsg.) (2012): Final report - Science as an open enterprise, hrsg. Von der Royal Society, [online], verfügbar unter: royalsociety.org/topics-policy/projects/science-public-enterprise/report/ .

Broeder, Daan / Dreyer, Malte / Kemps-Snijders, Marc / Witt, Andreas / Kupietz, Marc / Wittenburg, Peter (2008): Persistent and Unique Identifiers, [online], verfügbar unter: www.clarin.eu/sites/default/files/wg2-2-pid-doc-v4.pdf .

DOI-Foundation (2016): DOI Handbook, [online], verfügbar unter: www.doi.org/doi_handbook/8_Registration_Agencies.html .

Gowers, Tim / Kalai, Gil / Nielsen, Michael / Tao, Terry (2016): The polymath blog, [online], verfügbar unter: polymathprojects.org/

Graf, Klaus (2011): Wie zitiere ich ... Online-Quellen?, [online], verfügbar unter: archivalia.hypotheses.org/13353 .

Ohye, Maile / Kupke, Joachim (2012): The Canonical Link Relation, [online], verfügbar unter: tools.ietf.org/html/rfc6596 .

Rauber, Andreas / Asmi, Ari (2016): Identification of Reproducible Subsets for Data Citation, Sharing and Re-Use. Draft, [online], verfügbar unter: rd-alliance.org/system/files/documents/RDA-Guidelines_TCDL_draft.pdf .

Rauber, Andreas / Asmi, Ari / van Uytvanck, Dieter / Pröll, Stefan (2015): Data Citation of Evolving Data. Recommendations of the Working Group on Data Citation - Revision Oct. 20 2015, [online], verfügbar unter: rd-alliance.org/recommendations-working-group-data-citation-revision-oct-20-2015.html .

Van de Sompel, Herbert / Rosenthal, David S. H. / Nelson, Michael L. (2016a): Web Infrastructure to Support e-Journal Preservation (and More), in: arXiv:1605.06154 [cs], [online], verfügbar unter: arxiv.org/abs/1605.06154 .

Van de Sompel, Herbert / Klein, Martin / Jones, Shawn M. (2016b): Persistent URIs Must Be Used To Be Persistent, in: arXiv:1602.09102 [cs], [online], verfügbar unter: arxiv.org/abs/1602.09102 .

10 Autoren

Eckhart ARNOLD, Bayerische Akademie der Wissenschaften, Referat für IT und Digital Humanities, Alfons-Goppel-Str. 11, D-80539 München.

Stefan MÜLLER, Bayerische Akademie der Wissenschaften, Referat für IT und Digital Humanities, Alfons-Goppel-Str. 11, D-80539 München.