Zitierbarkeit

Eine wichtige Anforderung an digitale Publikationen

Sabine Kunze
sabine.kunze@ub.hu-berlin.de

Ähnlich wie die ISBN garantiert, dass ein Leser das von ihm zitierte Buch im Buchhandel immer wiederfinden kann, so garantieren Persistent Identifiers den dauerhaften Zugang zu Netzpublikationen.

Einleitung

Das größte Problem der Online-Dokumente ist die Flüchtigkeit: Heute noch im Netz gefunden – morgen schon daraus verschwunden. Wie kann man elektronische Publikationen dauerhaft zugriffsfähig und damit zitierbar machen?

An dieser Stelle kommen dauerhafte Bezeichnungen, sogenannte *Persistent Identifiers* (PI) zum Einsatz. Sie sollen, ähnlich wie die ISBN bei gedruckten Büchern, garantieren, dass immer dieselbe Publikation referenziert wird.

Normale URLs (Uniform Ressource Locator), die wir häufig als http-Adressen in der Form *http://edoc.hu-berlin.de* verwenden, können diese Funktion nicht übernehmen. Sie ändern sich z. B. bei einem Namenswechsel des Servers oder bei einer Umstrukturierung der digitalen Angebote durch Verschieben der Dokumente im Dateibaum.

So kann es also passieren, dass ein Dokument, das vor einem Jahr noch unter der Adresse *http://www.agrar.hu-berlin.de/pflanzenbau/agrarmet/hefte.html* verfügbar war, nun nicht mehr darunter gefunden werden kann.

Die Idee ist es nun, diese URLs zu ersetzen durch einen Ausdruck, der unabhängig vom jeweiligen Server ist. So könnte dann oben genanntes Dokument durch einen Persistent Identifier referenziert werden.

Dieser wird in einer Datenbank, Resolver (Auflöser) genannt, gemeinsam mit dem aktuell dahinter liegenden URL gespeichert. So wird sichergestellt, dass bei der Eingabe des PI in das Browserfenster automatisch zum richtigen Dokument gelinkt wird.

Uniform Resource Names (URN) an Der Deutschen Bibliothek

Um einen dauerhaften Zugriff auf elektronische Ressourcen zu gewährleisten, hat Die Deutsche Bibliothek (DDB) einen Dienst zur Vergabe und Verwaltung von Persistent Identifiern aufgebaut und wendet als Persistent Identifier den *Uniform Resource Name* (URN) in Form der *National Bibliographic Number* (NBN) an. Dadurch sind elektronische Dokumente stabil referenzierbar und auch bei einem sich ändernden URL erreichbar.

»Der Uniform Resource Name (URN) existiert seit 1992 und ist ein Standard zur Adressierung von Objekten mit einer institutionellen Verpflichtung zur persistenten, standortunabhängigen Identifizierung der Ressourcen. URNs wurden mit dem Ziel konzipiert, die Kosten für die Bereitstellung von Gateways sowie die Nutzung von URNs so gering wie möglich zu halten – vergleichbar mit existierenden Namensräumen, wie z. B. URLs. Aus diesem Grund wurde in Standards festgelegt, wie bereits existierende oder angewendete Namensräume bzw. Nummernsysteme einfach in das URN-Schema sowie die gängigen Protokolle, wie z. B. HTTP (Hypertext Transfer Protocol), oder Schemas, wie z. B. URLs, integriert werden können.« [1]

URNs, die an der DDB administriert werden, leiten sich aus dem internationalen Namensraum NBN (National Bibliographic Number) ab. Die DDB bietet die Verwaltung von URNs mit dem Namensraum `nbn:de` an. Ein solcher URN wird nach dem folgenden Muster gebildet:

```
urn:nbn:de:[Verbundabkürzung]:
      [Sigelnummer]-[Nummer][P]

Beispiel: urn:nbn:de:kobv:
      11-10041149
```

Dabei bedeutet [Verbundabkürzung] ein Buchstabenkürzel des jeweiligen Bibliotheksverbundes, [Sigelnummer] steht für das Bibliothekssigel der (Hochschul-) Bibliothek und [Nummer] muss eine eindeutige (Produktions-) Nummer sein. [P] bezeichnet die Prüfziffer zur Validierung des Strings.

Soweit vorhanden, werden im Online-Katalog der DDB nur URNs angezeigt und keine Herkunfts-URLs mehr (stets angezeigt wird jedoch der Link zum Archivserver der DDB). Damit wird die Anzeige ungültiger URLs vermieden und die Vorteile der URNs genutzt.

Mit Richtlinien für weltweit eindeutige URNs wurde eine sich organisatorisch selbst tragende Struktur eingeführt, bei der diese URNs vergeben werden. Gleichzeitig werden administrative Zwischenschritte vermieden, wodurch die Universitätsbibliotheken in der Lage sind, unmittelbar mit der URN-Vergabe zu beginnen. Außerdem lassen sich lokale Besonderheiten berücksichtigen, die in der Struktur der URNs abgebildet werden können.

Die DDB betreibt den URN-Resolver für die de-Domain. Die anderen Nationalbibliotheken administrieren die URN-Domains für ihr Land. Zusammengefasst werden diese Resolver dann über die *Library of Congress*, die die richtige Auflösung der Haupt-Domains, d. h. »de«, »se«, etc., anbietet.

Die DDB stellt auch ein im Rahmen des EPICUR-Projektes [2] entwickeltes Browser-Plugin zur Verfügung, das die direkte Eingabe einer URN in ein Browserfenster und die Weiterleitung an den Resolver und die Auflösung des URN-Terms und somit die Adressierung des Dokumentes ermöglicht.

So wird im obigen Beispiel bei der Eingabe des URN `urn:nbn:de:kobv:11-10041149` in die Adresszeile des Browsers automatisch auf den Resolver gelinkt: *http://www.nbn-resolving.de/urn:nbn:de:kobv:11-10041149*, der dann die Verbindung zur echten HTTP-Adresse *http://edoc.hu-berlin.de/dissertationen/hubertus-jochen-2005-01-11/HTML/* herstellt.

Für die Erstregistrierung von URNs wurde zum einen ein eigenständiger Geschäftsgang entwickelt. Zum anderen können Online-Dissertationen und deren eindeutige URN über die bestehenden Transferschnittstellen zusammen mit anderen Metadaten im MetaDiss- bzw. XMetaDiss-Format an Die Deutsche Bibliothek gemeldet werden.

Nutzung

Die DDB administriert seit September 2001 URNs für Netzpublikationen. Seitdem wurden insgesamt 38 402 URNs (Stand: 01.08.2005) registriert. Gegenwärtig haben sich 124 Institutionen angemeldet, davon nutzen 116 aktiv dieses Verfahren. Wie man aus Tabelle 1 ersehen kann, hat die Anzahl der registrierten URNs in den vergangenen Jahren rapide zugenommen.

Da die DDB und die Arbeitsgruppe »Elektronisches Publizieren« schon im Projekt »Dissertationen Online« [3] Kooperationspartner sind und URNs eine sehr kostengünstige Variante der Persistent Identifier (z.B. im Vergleich zu DOIs) darstellen, hat sich die AG Elektronisches Publizieren für die Verwendung von URNs entschieden.

2001	158
2002	2 880
2003	7 827
2004	14 168
2005 (Stand August)	13 322

Tab. 1: Anzahl registrierter URNs bei der DDB pro Jahr (Stand: 01.08.2005)

URNs für die digitalen Publikationen des edoc-Servers der HU

Jede Publikation auf dem edoc-Server erhält einen URN.

Auf dem edoc-Server werden URNs automatisch durch das Metadatensystem erzeugt. Sie haben immer die im o. a. Beispiel gezeigte Kodierung, dabei bedeutet das Kürzel `kobv` Kooperativer Bibliotheksverbund Berlin-Brandenburg, und das Sigel `11` steht für die Universitätsbibliothek der Humboldt-Universität zu Berlin.

Es können auch bestehende URNs in die Metadatenbank übernommen werden, beispielsweise die der Schriftenreihe des Projektes »nestor – Kompetenznetzwerk Langzeitarchivierung« [4]. Hier wird der URN ursprünglich von der DDB vergeben (siehe Abb. 1).

Zusammen mit der Meldung der Metadaten an die DDB erfolgt auch die Meldung des vergebenen URN, der dann von der DDB gehostet wird.

Abb. 1: Übernahme anderer URNs, hier einer der DDB.

Abb. 2: Metadaten eines Dokumentes auf dem edoc-Server der Humboldt-Universität zu Berlin.

Abb. 3: HTML-Version eines Dokumentes auf dem edoc-Server der Humboldt-Universität zu Berlin.

Neben dem URN existieren weitere Persistent-Identifier-Systeme:

- Das Handle-System,
- der Digital Object Identifier (DOI) und
- der Persistent URL (PURL).

Das Handle System

Das *Handle*-System ist eine technische Entwicklung der *Corporation for National Research Initiatives*. Mit dem Handle-System werden Funktionen, welche die Vergabe, Administration und Auflösung von Persistent Identifiers in Form von Handles erlauben, bereitgestellt. Die technische Basis bildet ein Protokoll-Set mit Referenz-Implementationen, wie z. B. DOI, LoC. Weiterführende Informationen findet man bei CNRI (*http://www.cnri.reston.va.us/*).

Das folgende Beispiel zeigt die Struktur von Handles. Diese bestehen aus einem Präfix und einem Suffix. Das Präfix ist ein numerischer Code, der die Institution bezeichnet. Das Suffix kann sich aus einer beliebigen Zeichenkette zusammensetzen.

```
<Handle> ::= <Handle Naming Au-
thority> »/« <Handle Local Name>

    Präfix/Suffix
   10.1045/january99-bearman
```

Der Digital Object Identifier (DOI)

Anwendungen von *Digital Object Identifiers* (DOI) werden seit 1998 durch die International DOI Foundation (IDF) [5] koordiniert. Die International DOI Foundation ist eine Non-Profit-Organisation. Sie überwacht die Entwicklung, die Regulierung und die Vergabe von DOI-Lizenzen an Registrierungsagenturen und Technologieanbieter und berät hinsichtlich der Nutzung und der Entwicklung von verwandten Dienstleistungen und Technologien. »Das System des DOI dient der interoperativen Identifizierung und dem Austausch geistigen Eigentums in einer digitalen Umgebung. Es bietet ein erweiterbares Gerüst, mit dem geistige Inhalte in jeglicher Form und in jeglicher Granularität verwaltet, Kunden und Content-Anbieter direkt miteinander verbunden, E-Commerce erleichtert und die automatisierte Verwaltung von Urheberrechten für alle Medien ermöglicht werden. Das DOI-System verwendet offene Standards mit einer Standardsyntax und wird von vielen führenden, internationalen Technologie- und Content-Organisationen genutzt. »DOI« und »DOI.org« sind Warenzeichen der International DOI Foundation.« [6]

Institutionen, die einen Dienst mit einem individuellen Profil aufbauen wollen, können dies in Form von Registration Agencies umsetzen. Das bekannteste Beispiel ist CrossRef [7]. Hier können die Metadaten und Speicherorte von Referenzen verwaltet und durch externe Institutionen weiterverarbeitet werden. »CrossRef vergibt Präfixe, registriert DOIs und dazugehörige Metadaten und ist Verlagen bei der Einrichtung von

	2004 Totals	2003 Totals	% Change 2003/2004
Journal Titles	10 833	9 000	+20%
Conferences	6 048		
Book Titles	4 649		
Total Records	13 917 354	9 960 631	+40%
Total DOI Resolutions	70 732 162	44 098 763	+60%
Publishers & Societies	1 400		
Members	341	264	+29%
Linkers	285	161	+77%
Depositors	326	240	+36%
Total 2004 Deposits (DOIs registered)	3 956 723	3 464 604	+14%
Total 2004 Current Deposits	1 471 718	1 006 394	+46%
Queries (references submitted)	188 927 644	101 177 977	+87%
Matches (DOIs returned)	51 430 436	18 266 364	+182%
Match %	27%	18%	+50%

Tab. 2: CrossRef: Full-Year Statistics (aus: http://www.crossref.org/01company/10newsletter.html#anchor2b) Lesedatum: 11.05.05

Links behilflich, die zu Ausgaben von Artikeln auf anderen Servern weisen, zu anderen Formaten eines Artikels weisen, Verbindungen zwischen Artikeln herstellen und zu verwandten Services in Verbindung mit einem Artikel leiten. Jeder Verlag für primäre Forschungsmaterialen kann Mitglied von CrossRef werden. Auch Bibliotheken, Sekundärverlage, Kompilators und Forscher im Bereich von digitalen Bibliotheken können den Mitgliederzugang zur Datenbank für Zeitschrifenmetadaten und zugeordneten DOIs nutzen.«[6]

Der Statistikvergleich in Tabelle 2 zeigt, dass dieses Angebot auch immer mehr angenommen wird.

Das DOI-System besteht aus den folgenden drei Komponenten:

- Metadaten,
- dem DOI als Persistent Identifier und
- der technischen Implementation des Handle-Systems.

Die Struktur von DOIs wurde seit 2001 in Form eines ANSI/NISO-Standards (Z39.84) normiert, welche die Komponenten der Handles widerspiegelt.

```
Präfix/Suffix
10.1198/april05-fischer
```

Der Zahlencode `10` bezeichnet den String als DOI, die unmittelbar an den Punkt grenzende Zahlenfolge `1198` steht für die vergebende Institution, z. B. eine *Registration Agency*. Der alphanumerische String im Anschluss an den Schrägstrich identifiziert das Objekt, z. B. einen Zeitschriftenartikel.

Persistent URL (PURL)

Ein Persistent Uniform Resource Locator oder PURL ist ein URL, der aber nicht direkt auf eine Internet-Ressource verweist, sondern auf einen Verzeichnisdienst, der den tatsächlichen, aktuellen URL der Ressource zurückgibt. Aus Benutzersicht entspricht ein PURL somit einem HTTP-redirect.

PURL (Persistent URL) wurde vom *Online Computer Library Center* (OCLC) 1995 im Rahmen des *Internet Cataloging Projects*, das durch das *U.S. Department of Education* finanziert wurde, eingeführt, um die Adressdarstellung für die Katalogisierung von Internetressourcen zu verbessern. [8]

PURLs sind – wie URNs – ein Weg, konstante, persistente Verweise auf Inhalte von Webpages bereitzustellen, während die tatsächlichen Adressen der Seiten variabel bleiben können.

Fazit

Persistent Identifiers können URLs ersetzen oder als stabile Referenzen in den Dokumenten selbst angewendet werden. Der Pflegeaufwand für die Aktualisierung der Verweise reduziert sich, da URLs nur an einer einzigen Stelle – in dem Persistent-Identifier-Dienst – automatisiert gepflegt werden, die Links können in mehrere Nachweisdienste integriert werden. Digitale Publikationen haben einen weltweit eindeutigen Identifier und sind damit zuverlässig zitierbar.

Literatur:

[1] http://www.persistent-identifier.de/ueberblick/Beispiele.php#URN
[2] EPICUR-Projekt: http://www.persistent-identifier.de/
[3] Projekt Dissertationen Online: http://www.dissonline.de/faq-x.htm
[4] nestor: http://www.langzeitarchivierung.de
[5] DOI: http://www.doi.org/
[6] http://www.doi.org/news/011009-Frank-Deutsch.pdf
[7] http://www.crossref.org/
[8] PURL: http://purl.oclc.org/