

**Literatur im Netz - Sammeln, Erschließen, Archivieren. Praxisbericht über eine neue Herausforderung für die klassische Bibliothek**

**Vortrag auf der ASpB-Tagung an der Universitätsbibliothek Karlsruhe am 23.09.2009**

**Karin Schmidgall, Jochen Walter (Bibliothek des Deutschen Literaturarchivs Marbach)**

Die Bibliothek des Deutschen Literaturarchivs sammelt, erschließt und archiviert die neuere deutschsprachige Literatur. Sie ist mit weit über 700.000 Bänden eine der größten Spezialbibliotheken für diesen Bereich und wird von Forschern aus aller Welt frequentiert. Zu ihren Besonderheiten zählen über 130 Autoren-, Gelehrten- und Sammlerbibliotheken, Verlagsarchivproduktionen und gattungsspezifische Sammlungen. Der Bestand an literarischen Zeitschriften der Moderne seit 1880 gehört zu den umfangreichsten in der Bundesrepublik. Seit 1959 werden literarische und literaturwissenschaftliche Zeitschriften, Jahrbücher und Sammelwerke retrospektiv und in einem fortlaufenden Programm ausgewertet. Monographien und Beiträge werden bibliographisch, dokumentarisch und sachlich erschlossen und sind im Onlinekatalog differenziert suchbar.

In seinen Sammlungen bildet das Deutsche Literaturarchiv Marbach (DLA) das Netzwerk des literarischen Lebens in all seinen Facetten ab. Im Zentrum des quellenorientierten Sammelns und der Erschließung steht der Autor (bzw. die Autorin). Die Literatur wird dokumentiert vom Entstehungsprozess eines Werkes über die verschiedenen Ausgaben und dessen Rezeption in der Literaturkritik, seine dramaturgische Umsetzung in Hörfunk, Film, auf der Bühne und in der Musik. Seit 2008 bezieht das DLA auch Internetquellen wie literarische Zeitschriften, Netzliteratur und Weblogs in sein Spektrum mit ein und reagiert damit auf die zunehmende Bedeutung des Internets als Publikationsforum.

Sammeln, Erschließen und Archivieren bilden eine notwendige Einheit; gerade die Flüchtigkeit der netzbasierten Ressourcen macht eine langfristige Sicherung der Verfügbarkeit erforderlich. Notwendig sind daher mehrere Säulen, auf denen diese neue Sammlung von „Literatur im Netz“ basiert. Mit der Marbacher Datenbank Kallias steht ein EDV-Standardsystem zur Verfügung, das den spezifischen Anforderungen des Deutschen Literaturarchivs angepasst wurde und über eine Schnittstelle zum Katalog des Südwestdeutschen Bibliotheksverbundes (SWB) verfügt. Seine Aufgabe ist der integrierte Nachweis aller Bestände. Diese sind über den SWB, die Zeitschriftendatenbank (ZDB) und

Kalliope, den Verbundkatalog für Nachlässe und Autographen, zugleich in das nationale Bibliotheks- und Archivnetzwerk eingebunden.

Mit der Virtuellen Fachbibliothek „Germanistik im Netz“<sup>1</sup> ist zudem ein fachspezifischer Zugang zu Printmedien, Datenbanken, Online-Ressourcen und digitalen Volltexten für die deutsche Sprach- und Literaturwissenschaft geschaffen worden. Das Portal wurde mit Unterstützung der Deutschen Forschungsgemeinschaft unter Federführung der Sondersammelgebietsbibliothek UB Frankfurt/Main aufgebaut. Im Fachinformationsführer werden Internetquellen mit bibliothekarischen Metadaten und einer kurzen Inhaltsbeschreibung erschlossen. Als Kooperationspartner der Virtuellen Fachbibliothek pflegt das DLA Marbach seit 2006 online angebotene literarische Zeitschriften, Weblogs und Netzliteratur ein.

Im Zuge der Maßnahmen des Landes Baden-Württemberg zur Langzeitarchivierung hat das Bibliotheksservice-Zentrum (BSZ) Konstanz mit der Software SWBcontent eine technische Infrastruktur für die langfristige Verfügbarkeit und Archivierung von Online-Publikationen aufgebaut, die sich in die nationalen und internationalen Bestrebungen durch die konsequente Orientierung an Standards nahtlos einbettet. Sie umfasst die Komponenten Spiegelung, persistente Identifizierung über URNs, automatische Erhebung technischer Metadaten, Bereitstellung der Metadaten über OAI sowie eine Bereitstellung für die Langzeitarchivierung (vgl. BSZ-Kompakt: 31). Seit 2009 setzt das DLA die Software SWBcontent in der Installation „Literatur im Netz“ produktiv ein; die gespiegelten Objekte werden auf dem BSZ-Server archiviert.

---

<sup>1</sup> <http://www.germanistik-im-netz.de/>



Von den 140 Zeitschriften und etwa 180 Weblogs, die mittlerweile in Kallias und im Webverzeichnis GinFix von „Germanistik im Netz“ verzeichnet und beschrieben sind, werden derzeit 31 Weblogs und 9 Zeitschriften in zeitlich festgelegten Intervallen regelmäßig gespiegelt<sup>2</sup>; 2009 sollen weitere 20 Ressourcen in die Spiegelung einbezogen werden. Auch in Zukunft kann nur eine überschaubare Menge qualitativ gepflegt werden. Mit der Aufnahme des Produktionsbetriebs sind die Qualitätssicherung, d.h. die Überprüfung von Kriterien wie Authentizität, Adressierbarkeit, Singularität und inhaltliche Qualität sowie die regelmäßige Aktualisierung der Speicherung (Versionierung) in Geschäftsgängen geregelt worden.

## 2. Auswahl

Am Anfang steht das Bemühen um einen Überblick über das Angebot an literarischen Zeitschriften und Weblogs, sodann die Auswahl gemäß dem Sammelauftrag des DLA. Probleme ergeben sich zum einen aus der schieren Menge der relevanten Internetressourcen, die auch dann groß bleibt, wenn man sich auf den deutschen Sprachkreis beschränkt. Auch bleibt es abzuwarten, ob die Marbacher Auswahlkriterien, die bislang überwiegend im

<sup>2</sup><http://literatur-im-netz.dla-marbach.de/>

Bereich der Print- und Rundfunkmedien bzw. der Verlagsproduktionen entwickelt und angewendet werden, analog auf Medien übertragbar sind, die sich in Produktion, Distribution und Rezeption von klassischen Veröffentlichungen wesentlich unterscheiden. Bei der Erwerbung greifen wissenschaftliche Bibliotheken normalerweise auf die umfassenden bibliographischen Dienstleistungen der Deutschen Nationalbibliothek zurück, die Marbacher Spezialbibliothek wertet zusätzlich viele entlegene Quellen und Hinweise aus. Doch sind in der Nationalbibliographie die gesuchten elektronischen Quellen des WWW bislang nicht zu finden bzw. nicht gezielt recherchierbar. Die Suche im Netz zielt daher zunächst auf die Existenz kumulierender Verzeichnisse und Listen, auf „Webliographien“.

Da die meisten dieser Listen<sup>3</sup> auf die Initiativen einzelner Personen zurückgehen und zum Zeitpunkt der Auswertung bereits nicht mehr aktualisiert werden, kommen sie für eine bibliographische Routine nicht in Frage, da sich ihr Wert bereits nach der ersten Auswertung erschöpft hat. Bedenkenswert ist im Übrigen, dass auch solche Verzeichnungen zur Überlieferungsbildung im Netz beitragen und deshalb gleichfalls archivierungswürdig sind. Zweifel an der Vollständigkeit der jeweiligen Listen resultieren nicht nur aus einer unterstellten Menge an Publikationen, sondern haben ihren Grund auch in der unterschiedlichen Definition der literarischen Zeitschrift bzw. des literarischen Weblogs. Die Netzpublikationen sollen in erster Linie literarische Quellen sein, d.h. hauptsächlich deutschsprachige belletristische Textsorten enthalten. Reine Rezensionsorgane beispielsweise finden im Rahmen von "Literatur im Netz" keine Berücksichtigung. Weitere Ausschlusskriterien gibt es im Rahmen des Projektes nicht. Diesen Auswahlkriterien folgt keine der erwähnten Listen (Ausnahme: [litblogs.net](http://litblogs.net)<sup>4</sup>). Problematisch ist die Anwendung der Kriterien auf Publikationen, die ohne herkömmliche redaktionelle Verfahren wie Zeitschriftenredaktionen und/oder Verlagslektorate auskommen. Gerade Weblogs erweisen sich mitunter als Grauzone zwischen literarischen, tagebuchartigen, journalistischen und sachlichen Texten, zwischen Veröffentlichung und privatem Inhalt.

---

<sup>3</sup> z. B. <http://www.berlinerzimmer.de/eliteratur/magazine960129.html>, <http://literaturwelt.de/brennenderbusch/links.html>, <http://home.rhein-zeitung.de/~nettwill/main.htm> etc.

<sup>4</sup> <http://www.litblogs.net/>

Zeitschriften sind formal konsistenter und daher einfacher zu kategorisieren. Ihnen kommt jedoch auch die geringere Bedeutung bei "Literatur im Netz" zu, was verschiedene Gründe hat: Hinter den Titeln stehen oft kleine bis kleinste Verlage oder gar Einmannunternehmen, deren Fortbestand vermutlich nur durch Enthusiasmus und ein großes Maß an Selbstaussbeutung zu sichern ist. Außerdem handelt es sich oftmals um Webauftritte der gedruckten Ausgaben, die diese werbend und mit Leseproben begleiten. Dabei zeichnet es sich nicht ab, dass die Online-Veröffentlichung an die Stelle des konventionellen Mediums tritt. Dies ist ein deutlicher Unterschied zum Markt der wissenschaftlichen Periodika, deren Entwicklung in der völlig anders gearteten wissenschaftlichen Kommunikation und Rezeption begründet liegt. Im Gegensatz zu elektronischen Zeitschriften werden Weblogs auch in Zukunft die größere Rolle als literarische Publikationsform spielen.<sup>5</sup>

Aus der oben beschriebenen schwierigen bibliographischen Lage heraus ergibt sich vor allem für Weblogs die Notwendigkeit von weiteren Auswahlverfahren: Dem ersten bibliographischen Durchgang liegen die oben beschriebenen Listen zugrunde. Anschließend werden die Blogrolls der 54 Titel aus der ersten Runde vollständig ausgewertet. Das Ergebnis sind 63 weitere Weblogs. Die folgenden Durchgänge basieren auf einer Mischung aus den bereits bestehenden Listen im Internet, Blogrolls, Hinweisen aus Blogpostings sowie direkten Hinweisen von Autorinnen und Autoren. Die aktuelle interne Blogsammlung umfasst ca. 430 Titel, von denen etwa 230 als relevant im Sinne des Projektes einzustufen sind. Das Einsammeln der in Frage kommenden Weblogs geschieht als Feedsammlung im FeedReader Bloglines<sup>6</sup>. Für die Zeitschriftentitel erfolgt eine erste Sammlung mittels der Bookmark-Funktionen im Browser Firefox. Im Bewusstsein einer eigenen déformation professionnelle bedingt durch die nach wie vor konventionellen Maßstäbe bleibt eine gewisse Unsicherheit und der seltsame, in der Bibliotheksarbeit seltene Eindruck, Realitäten nicht nur zu beschreiben, sondern auch zu schaffen.

### **3. Rechtliche Aspekte**

In der Regel gibt es gegenüber der Bibliothek des DLA außer beim Anspruch auf ein Belegexemplar keinerlei Abgabepflicht von Publikationen, wie es etwa der Fall bei den

---

<sup>5</sup>Ein Beispiel für diese Ablösung der Zeitschriften durch Weblogs im literarischen Bereich mag die Zeitschrift "Sieb.10 : e-zine für Literatur" sein, die als Zeitschrift im Jahr 2007 endet und als Weblog fortgesetzt wird.

<sup>6</sup> <http://www.bloglines.com/>

Regionalbibliotheken ist. Daher und weil es sich bei den erwähnten Netzpublikationen nicht um käuflich erworbene Verlagsprodukte handelt, sind sowohl bei der Vervielfältigung (Speicherung einer archivierten Fassung der Online-Ressource) als auch im Rahmen der Nutzung (öffentliche Präsentation via Nachweis in den o. a. Katalogen) die Verwertungsrechte der jeweiligen Rechteinhaber zu berücksichtigen, die sich aus dem Urheberrechtsgesetz ergeben: das Recht auf Vervielfältigung und das auf Verbreitung. Die Rechteinhaber werden deshalb angeschrieben und über die Archivierungsabsicht, den Archivierungszweck und die Tätigkeit des DLA allgemein informiert. Mit einem entsprechenden Formular werden sie sodann schriftlich um die Zustimmung zur Archivierung ihres Online-Angebotes gebeten. Dies geschieht auch in den Fällen, in denen die vorliegende Online-Ressource unter einer Creative-Commons-Lizenz steht, die beispielsweise eine Vervielfältigung und Verbreitung unter den Bedingungen "Namensnennung", "keine kommerzielle Nutzung" und "Weitergabe unter den gleichen Bedingungen" gestattet. Für die Archivierung erhalten die Angeschriebenen keine Vergütung; umgekehrt ist die Langzeitarchivierung für die Rechteinhaber ein kostenloser Service, aus dem für sie keine weiteren Arbeiten oder Verpflichtungen resultieren. Gleichzeitig stellt sich die Bibliothek auch als Ansprechpartnerin für Fragen und Kritik vor: Nach erfolgter Spiegelung werden die Betroffenen informiert und um ihre Meinung gebeten, auch um Fehler und Unstimmigkeiten zu beseitigen. Diese Erfahrungen werden ggf. als Anforderungen in die Entwicklung der beteiligten Software SWBcontent einfließen.

Etwa 90 Prozent der Rechteanfragen werden positiv beantwortet. Die restlichen zehn Prozent lehnen eine Archivierung entweder ab oder stimmen nur unter Bedingungen zu, die zur Zeit nicht erfüllt werden können. Die Sicherung und Langzeitarchivierung eines flüchtigen Mediums ist in den Fällen nicht im Interesse der Autorinnen und Autoren, in denen die Flüchtigkeit nicht Nachteil, sondern gewollte oder erwünschte Eigenschaft ist. Der bisweilen geäußerte Wunsch, die archivierte Quelle gänzlich oder in bestimmten Teilen löschen zu können, ist derzeit technisch nicht realisierbar. Dabei handelt es sich in der Regel um persönliche Bedenken der Schreibenden, die Kontrolle über ältere Texte, Fassungen oder Diskussionsbeiträge aus der Hand zu geben. An dieser Stelle sei noch einmal an die Position der Weblogs zwischen Veröffentlichung und privatem Tagebuch erinnert. Nicht wenige Blogger schreiben unter Pseudonym. Die Wahrung dieser Pseudonyme ist eine häufig genannte (und leicht zu erfüllende) Archivierungsbedingung. Eine Indexierung der

archivierten Quellen durch Suchmaschinen wird zwar vielfach abgelehnt, ihre Verhinderung jedoch bislang nicht als zwingende Voraussetzung für eine Teilnahme bei "Literatur im Netz" genannt. Neben die persönlichen und technischen Einwände treten bei Weblogs und Websites urheberrechtliche Bedenken dergestalt, dass man nicht Rechte gewähren könne, über die man nicht verfüge. Obgleich die Zahl der betroffenen Rechteinhaber auch weitere Autoren, Fotografen, Illustratoren, Verfasser von Style sheets und nicht zuletzt die Diskussionsteilnehmer umfassen kann, wird in der Regel nur die hauptverantwortliche Person angeschrieben. Eine vollständige Ermittlung und Kontaktierung aller Beteiligten ist auch wegen der Pseudonyme nicht machbar, so dass hier urheberrechtliche Bedenken nicht vollständig ausgeräumt werden können. Auf manchen Weblogs gibt es mittlerweile den Hinweis, dass Archivierungsprojekte wie Dilimag<sup>7</sup> oder "Literatur im Netz" die vorliegende Quelle archivieren<sup>8</sup>. Die rechtliche Wirksamkeit solcher Disclaimer ist jedoch zumindest in Bezug auf eine rückwirkende Gültigkeit fraglich.

#### **4. Technische Aspekte**

Sind Katalogisierung und Rechteinholung erfolgreich abgeschlossen, erfolgt die eigentliche Archivierung, d. h. die Spiegelung der Ressource inklusive Vergabe von Persistent Identifiers in Form von URNs.

Ist ein Titeldatensatz abschließend bearbeitet, enthält er in den drei Nachweisinstrumenten Kallias, ZDB und SWB sowohl die URL des Original-Angebots im WWW als auch die URN der auf dem Server des Bibliotheksservice-Zentrums Baden-Württemberg archivierten Fassung. Von nun an steht die archivierte Ressource für die Nutzung zur Verfügung<sup>9</sup>.

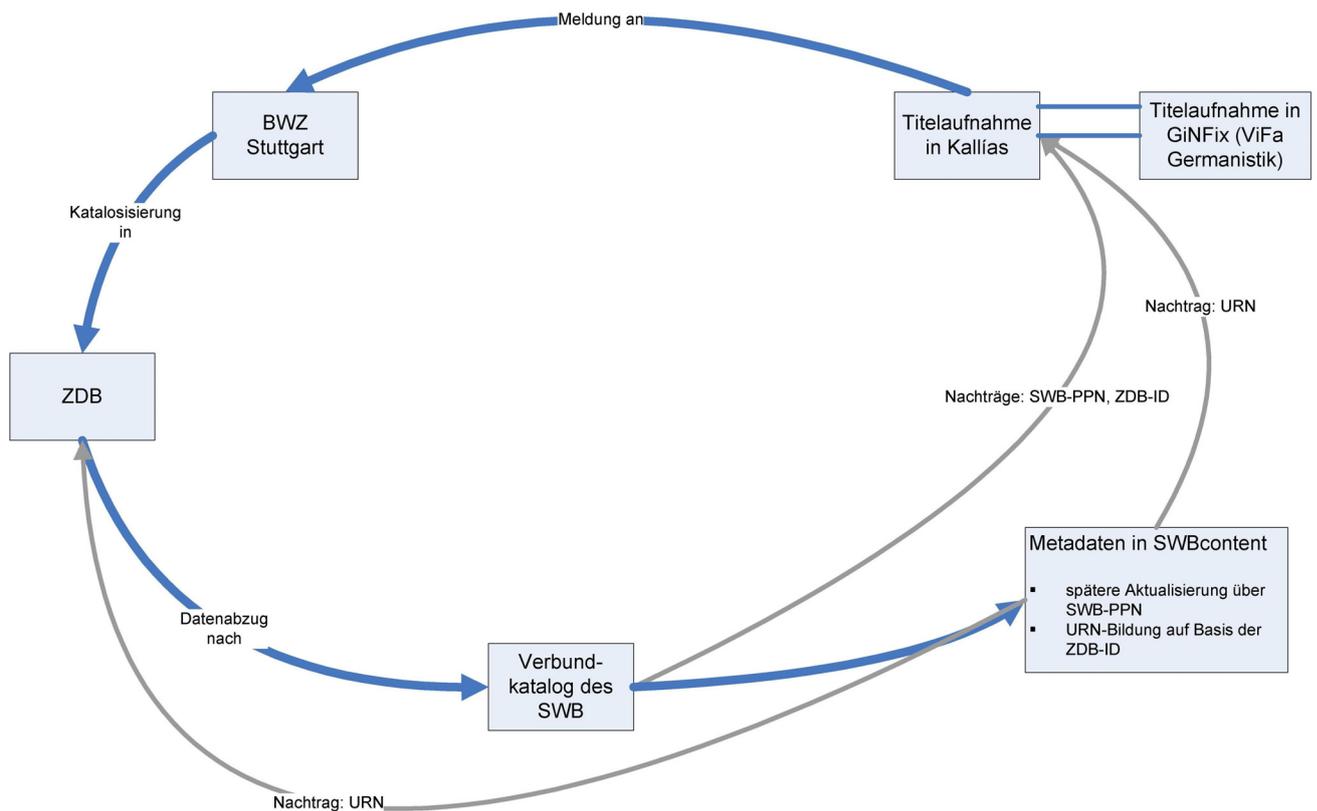
Zu den technischen Grundvoraussetzungen für Spiegelungen in SWBcontent gehören die Titelaufnahmen der zu archivierenden Ressourcen. Wie diese in SWBcontent gelangen, zeigt ausgehend von der Titelaufnahme in Kallias das folgende Schaubild:

---

<sup>7</sup> <http://www.uibk.ac.at/germanistik/dilimag/>

<sup>8</sup> z. B. <http://albannikolaiherbst.twoday.net/>, <http://www.litblogs.net/>, <http://www.kenyamamoto.de/wordpress/>, <http://littalk1.blogspot.com/> u. a.

<sup>9</sup> Ein Nachtrag der URNs in der Virtuellen Fachbibliothek Germanistik steht bislang noch aus.



1. Titelaufnahme in Kallias (parallel dazu auch in der ViFa Germanistik)
2. Katalogisierung in der ZDB via Baden-Württembergische Zeitschriften-Redaktion Stuttgart
3. Katalogisierung im SWB über Datenabzug aus der ZDB (plus Nachträge)
4. Upload der SWB-Daten in SWB-Content (plus Nachträge)

Die eingesetzte Software SWBcontent fungiert hauptsächlich als Workflowtool, während die eigentliche "Ernte" der Netzquellen mit dem Programm httrack<sup>10</sup> erfolgt, das über die webbasierte Oberfläche von SWBcontent konfigurierbar ist. Zu den dort vorgenommenen Downloadparametern gehören: Ausgangsadresse der Quelle, Filter für ein- bzw. auszuschließende Adressen, zeitliche und quantitative Limitierungen, interne und externe Spiegelungstiefe sowie das Verhalten gegenüber Anweisungen wie robots.txt. Die Spiegelung einer konventionell strukturierten Netz-Zeitschrift erfolgt "heft"-weise. Ein lediglich begleitender Webauftritt wird als Website in einem regelmäßigen Turnus komplett gespiegelt. Dieses Verfahren, Zeitschnitte zu archivieren, das auch das Internet Archive<sup>11</sup> praktiziert,

<sup>10</sup> <http://www.httrack.com/>

<sup>11</sup> <http://www.archive.org/index.php>

wird auch auf Weblogs angewendet. Im Gegensatz zur abschnittswisen (heftweisen) Methode führt diese Vorgehensweise zu einem hohen Maß an redundanten Daten. Will man jedoch die jeweilige Quelle möglichst vollständig erhalten, ist diese Datenhaltung alternativlos, da die neueste Fassung einer Website nicht zugleich die vollständigste sein muß. Eine genaue Überprüfung dieser Vollständigkeit ist aus weiteren Gründen nicht möglich: Die zu spiegelnde Vorlage, um einen bibliothekarischen Terminus zu gebrauchen, ist im Gegensatz zu konventionellen Medien in ihrem Umfang nicht fest definiert. Vielmehr werden ihre Grenzen erst vom Archivierenden durch die Downloadparameter gesetzt. Bei Spiegelungen im Rahmen von "Literatur im Netz" werden beispielsweise aus Gründen der Rechteverwaltung und des Speicherplatzes externe Links nicht verfolgt. Die Downloadparameter lauten an dieser Stelle also: „Externe Spiegelungstiefe = null“. Sollte dies zu Ergebnissen führen, die für die Rechteinhaber nicht akzeptabel sind, z. B. zum Fehlen kommentierter Bildern aus einer externen Quelle, können differenziertere Anweisungen beraten werden, die dann wiederum zu einer anderen, umfassenderen Definition der Vorlage führen. Jede Spiegelung einer Netzressource kann als Unikat betrachtet werden: Sie ist einmal auf den jeweiligen Zeitpunkt bezogen, lässt sich also nicht wiederholen, und es ist eher unwahrscheinlich, dass zwei gleichzeitige Archivierungsvorgänge einer Netzressource gleichartig angelegt sind, wenn sie von verschiedenen Personen in unterschiedlichen Situationen vorgenommen werden. Das liegt einerseits in den vielfältigen Konfigurationsmöglichkeiten von Harvestingprogrammen wie httrack oder Heritrix<sup>12</sup>, zum anderen in der Natur des Mediums selbst begründet. Daraus ergeben sich gerade im Blick auf eine wissenschaftliche Nutzung der archivierten Quellen auch Konsequenzen für die Katalogisierung, für Struktur und Inhalt der Metadaten, da im Vergleich zu Büchern eine objektive Kollationierung nicht zu bewerkstelligen ist und der Zeitfaktor hinzukommt. Auf der Titelebene werden die hinlänglich bekannten Daten wie Autor/en und/oder sonstige beteiligte Personen, Titel, Erscheinungsvermerk, Ort, Verlag, Jahr, Erscheinungsverlauf, Sprache, Land festgehalten. Auf der Ebene der einzelnen Spiegelung werden derzeit die (technischen) Metadaten Archivierungszeitpunkt, Format, Gesamtgröße nicht nur im Archivsystem erfasst, sondern auch auf der Benutzeroberfläche angezeigt. Es wäre wünschenswert, diese zukünftig um die vorgenommenen Downloadparameter zu ergänzen. So wäre der Archivierungsvorgang genau beschrieben und objektiv nachvollziehbar, nicht nur

---

<sup>12</sup> <http://crawler.archive.org/>

für den netzphilologisch interessierten Forscher, sondern auch für die Archivare im Blick auf künftige Spiegelungsvorgänge. Und nur auf diese Weise wird die archivierte Ressource genau beschrieben, da anders als bei Büchern der Archivierungsvorgang selbst zu ihren Merkmalen gehört. Eine Alternative zu diesem Kollationierungsvorgehen könnte die Definition der Vorlage durch die Betreiber selbst sein mit detaillierten Spiegelungsanweisungen in Form von Sitemaps. Der hierbei deutlich höhere Aufwand für die Akteure macht dieses Szenario jedoch wenig wahrscheinlich für eine breite Anwendung.

Für den eindeutigen Zugriff und die Zitierbarkeit werden den Metadaten in SWBcontent URNs mit korrelierenden URLs zugeordnet. Dabei wird jeder in SWBcontent angelegte "Knoten" separat adressiert wie das folgende Schema (Stand Juli 2009) zeigt:

- **Titelfrontpage**  
URN(1)/URL(1)
  - **Frontpage der ersten Einzelspiegelung**  
URN(2)/URL(2)  
[URL (Ausgangs-URL des Originals)]
    - **Archiviertes Objekt**  
URL(3)
  - **Frontpage der zweiten Einzelspiegelung**  
URN(4)/URL(4)  
[URL (Ausgangs-URL des Originals)]
    - **Archiviertes Objekt**  
URL(5)
  - **Frontpage der dritten Einzelspiegelung**  
etc.
- Zuordnungen URL und URN:  
URN(1) -> URL(1)  
URN(2) -> URL(2) und URL(3)  
URN(4) -> URL(4) und URL(5)

Die Registrierung und Auflösung der URNs erfolgt durch die Deutsche Nationalbibliothek<sup>13</sup> im URN-Namensraum nbn:de, Unternamensraum bsz:mar1. Im Anschluss werden die URNs

---

<sup>13</sup> <http://www.persistent-identifier.de/>

in Kallias<sup>14</sup> und in der ZDB<sup>15</sup> manuell nachgeführt. Der Nachtrag im Verbundkatalog des SWB<sup>16</sup> geschieht automatisch mittels Datenabzug aus der ZDB.

Nicht immer gelingt es, die Quellen korrekt zu spiegeln. Am leichtesten fällt das bei klassischen HTML-Strukturen oder statischen Formaten wie PDF. Schwierigkeiten in diesen Bereichen resultieren z. B. aus der Verwendung von Frames, lassen sich aber über den Quelltext der entsprechenden Seiten klären. Größere Probleme bereiten externe Inhalte. Diese sind zum einen wegen der erwähnten Downloadeinstellungen für die externe Spiegelungstiefe selbst produziert. Zum anderen gibt es Fälle, in denen die Einbindung externer style sheets trotz entsprechender zusätzlicher Filter nicht gelingt. Auch cookies können zu unzureichenden Spiegelungsergebnissen führen, bei denen mehrfach nachjustiert werden muss, bevor das Ergebnis annehmbar ist. Die Adressierung von Seiten in Content Management Systemen und die klare Trennung von Inhalt, Struktur und Gestaltung können zu Kopien führen, in denen das Original kaum wiederzuerkennen ist. Ein weiteres Problemfeld stellen dynamisch generierte Inhalte und Datenbankausgaben dar. Grundsätzlich lässt sich sagen: Je breiter (in dimensionaler und technischer Hinsicht) die einzelnen Komponenten der Quelle verteilt sind, desto schwieriger ist ihre Zusammenführung in der Archivkopie.

Spätestens an dieser Stelle wird deutlich, dass der Archivierungsvorgang dem Wesen des Mediums Internet völlig zuwiderläuft. Die jeweilige Fehleranalyse erfolgt über die htrack-Fehlermeldungen in SWBcontent, die Quelltexte, robots.txt, über lokale Kontrollspiegelungen und Recherchen im hilfreichen htrack-Forum<sup>17</sup>. Sie kann vor dem Hintergrund von mitunter wenig transparenten Strukturen und einer großen Vielfalt beteiligter Formate und Objektarten im Einzelfall längere Zeit in Anspruch nehmen. Im Anschluss an eine Spiegelung wird all dies genau dokumentiert und im Fall einer gelungenen Archivierung eine Gebrauchsanweisung für die Zukunft erstellt. Der geschilderte Aufwand macht deutlich, dass diese Arbeiten beim augenblicklichen Stand der Technik nicht völlig automatisiert werden können. Vor allem im Blick auf Netzliteratur, bei der ggf. noch komplexere Funktionalitäten abzubilden sind, kann es hier nur um Einzelarchivierungen gehen.

---

<sup>14</sup> [http://www.dla-marbach.de/opac\\_kallias/index.html](http://www.dla-marbach.de/opac_kallias/index.html)

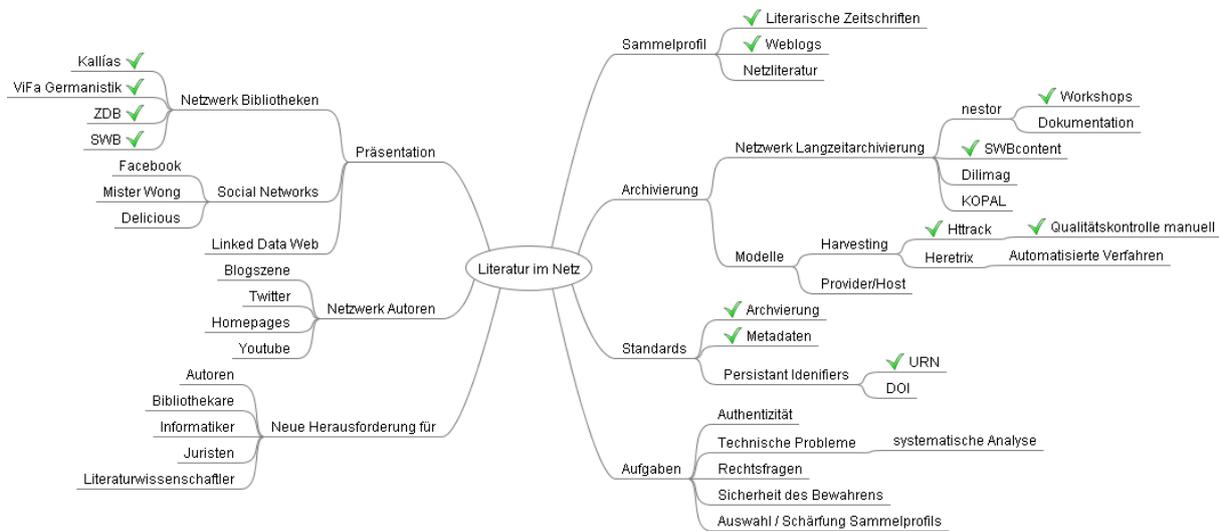
<sup>15</sup> <http://www.zeitschriftendatenbank.de/>

<sup>16</sup> <http://swb.bsz-bw.de/DB=2.1/>

<sup>17</sup> <http://forum.htrack.com/>

## 5. Ausblick

Wie viele Facetten Literatur im Netz hat, zeigt folgendes Brainstorming:



Dem Thema Langzeitarchivierung haben sich Bibliotheken, Archive und Museen in den letzten Jahren über theoretische Ansätze genähert. Insbesondere mit nestor, dem Kompetenznetzwerk Langzeitarchivierung ist eine Organisationsform für alle Belange der Langzeitarchivierung geschaffen worden. Praxiserfahrungen gibt es vergleichsweise sehr wenige. Auch das Projekt „Literatur im Netz“ steckt noch in den Kinderschuhen.

Die Bibliothek des DLA stellt sich der neuen Herausforderung mit den Methoden einer klassischen Bibliothek: Tradition und konventionelle Ansätze treffen auf die Schnell-Lebigkeit und Experimentierlust des Netzes. Nur knapp zehn Jahre seit der Entstehung von Netzliteratur werden für die Bereitstellung von Werken der Frühphase schon Ausgrabungsarbeiten notwendig sein. Falls sie nicht schon ganz verschwunden sind, existiert von manchen Texten vielleicht noch eine Version auf Festplatte oder CD-ROM. Mit der Archivierung von Weblogs steht das DLA dagegen mitten im literarischen Leben, wird sogar Teil der Literatur, wie im Blog „Die Dschungel. Anderswelt“ - unsere Anfrage zur Spiegelgenehmigung wurde gleich literarisch verarbeitet.

Die Archivierung der „Literatur im Netz“ ist selbst eine Werkstatt. Gerade beim Harvesting der Daten steht man vor fast unüberwindbaren Hindernissen, da eine Vielzahl von Website-Elementen zur Zeit als nicht spiegelbar gilt. Ein mit Drittmitteln gefördertes Projekt zur systematischen Untersuchung der Probleme bei der Archivierung von Internetquellen ist dringend wünschenswert. Dies ist nicht zuletzt notwendig um automatisierte Verfahren zur Spiegelung einsetzen zu können, die zu qualitativ akzeptierbaren Spiegelergebnissen führen. Ergänzend zur Archivierung durch Spiegelung von Websites sollte ein bereits im Jahr 2000 diskutiertes Provider-Modell (bei dem das DLA gegenüber Autoren und Literaturprojekten als Provider bzw. Host auftreten würde) weiterverfolgt und Autoren mehr in den Archivierungsprozess eingebunden werden.

Langzeitarchivierung kann nur in Kooperation mit starken Partnern erfolgen. Das Bibliotheksservice-Zentrum Konstanz ist Mitglied im Kooperationsverbund nestor dem deutschen Kompetenznetzwerk zur digitalen Langzeitarchivierung<sup>18</sup> und bereitet eine Kooperation mit KOPAL, dem Langzeitarchivierungssystem der Deutschen Nationalbibliothek vor. Das Marbacher Projekt ist damit integriert in das nationale Netzwerk.

„Literatur im Netz“ ist in der Bibliothekswelt präsent. Die Metadaten können im lokalen Onlinekatalog, im SWB sowie über die Zeitschriftendatenbank recherchiert werden und sind mit den archivierten Objekten verlinkt. Im Fachinformationsführer der ViFa Germanistik und auf der Homepage des DLA Marbach wird ein Browsing in Titellisten angeboten. Autoren wünschen sich darüber hinaus eine Einbindung des Angebots in Social Networks. Mit der Einhaltung von Archivierungs- und Metadatenstandards sowie mit der Vergabe von Persistent Identifiers werden auch erste Ansätze für die Nutzung der Daten im Rahmen semantischer Strukturen des Linked Data Web verfolgt. Im Web 3.0 und mit den damit verbundenen Bausteinen RDF (Resource Description Framework) und URI (Uniform Resource Identifier) sollen Strukturen geschaffen werden, die es ermöglichen, weltweit effektiver auf Informationen zuzugreifen, sie nutzen und weiterverarbeiten zu können.

---

<sup>18</sup> <http://www.langzeitarchivierung.de>

In den nächsten Jahren stehen Aspekte wie Authentizität, technische Probleme, Rechtsfragen, Sicherheit des Bewahrens sowie Schärfung des Erwerbungsprofils im Vordergrund. Im Verein mit Autoren, Bibliothekaren, Informatikern, Juristen und Literaturwissenschaftlern müssen sie erörtert und gelöst werden. Die Bewältigung dieser Aufgabe stellt neuartige und große Herausforderungen an die Arbeitsweisen, an Kapazitäten und Selbstverständnis einer klassischen Bibliothek.