Erfahrungen beim Aufbau eines Webseitenarchivs für die Internet-Veröffentlichungen Jüdischer Gemeinden und Verbände in der Bundesrepublik

Vortrag von Peter Honigmann

auf der Tagung über *Elektronisches Archivgut*
am 5. und 6. März in Dresden
(= 6. Tagung des Arbeitskreises *Archivierung von Unterlagen aus digitalen Systemen*)

1987 hat der *Zentralrat der Juden in Deutschland* mit Unterstützung der Bundesregierung in Heidelberg ein Archiv eingerichtet, das alle dokumentarischen und archivischen Aufgaben wahrnehmen soll, die sich aus den Anforderungen des jüdischen Lebens in der Bundesrepublik ergeben. In erster Linie geht es um die Übernahme und fachlich korrekte Bearbeitung des nach 1945 in den neugegründeten jüdischen Gemeinden und Verbänden entstandenen Schriftguts. Inzwischen sind an die 800 lfm Altakten von ungefähr 20 Registraturbildnern in die Heidelberger Magazine gebracht worden, was ungefähr einem Drittel der geschätzten Gesamtmenge entspricht. Daneben kümmert sich das Zentralarchiv um Nachlässe jüdischer Autoren und um die Dokumentation jüdischer Grabinschriften. [1]

In Ergänzung zu den verschiedenen Beständen werden systematisch archivische Sammlungen aufgebaut. In Hinblick auf das eigentliche Thema meines heutigen Vortrags verdient die **Sammlung von Mitteilungsblättern** jüdischer Gemeinden und Verbände insofern etwas mehr Aufmerksamkeit, als es sich dabei um die papierne Ausformung eines Kommmunkationsbedürfnisses handelt, das inzwischen mehr und mehr die Wege des Internet beschreitet. Vom Tag seiner Gründung an ist das Zentralarchiv bestrebt, alle in der Bundesrepublik von jüdischen Gemeinden, Verbänden, Einrichtungen und losen Gruppierungen mehr oder weniger periodisch herausgegebenen Publikationen möglichst vollständig zu sammeln. Unter Publikationen werden in diesem Zusamnmenhang schriftliche Mitteilungen verstanden, die wenigstens in dem Sinne öffentlich sind, daß sie den Mitgliedern einer Gemeinde oder eines Vereins zugeschickt werden. Die Sammlung des Zentralarchivs, die inzwischen 182 Titel umfaßt, reicht also von Zeitungen und Zeitschriften, die öffentlich vertrieben werden, über Mitteilungsblätter Jüdischer Gemeinden und

[1] Einen ausführlichen Überblick über die Arbeit des Zentralarchivs findet man in: Menora. Jahrbuch für deutsch-jüdische Geschichte, Bd. 12 (2001), S. 345-370.

Programmhefte Jüdischer Volkshochschulen bis hin zu xerokopierten Rundbriefen.

Informationen, die von einigen Gruppen inzwischen per **Fax** oder **e-mail** verschickt werden, lassen sich noch ohne besondere Mühe in den bisherigen Rahmen dieser Sammlung einfügen. Es war beim Erwerb eines Fax-Gerätes lediglich darauf zu achten, daß nicht auf Thermopapier, sondern auf Normalpapier ausgedruckt wird. Und auch die Umwandlung von e-mail in bedrucktes Papier hat weder zu einer Veränderung der ursprünglichen Struktur des Dokuments noch zu konservatorischen Problemen geführt.

Mehr und mehr gehen die von uns archivisch betreuten jüdischen Gemeinden, Verbände und Gruppen jedoch dazu über, eigene *Webseiten* einzurichten. Die damit einhergehenden Veränderungen sowohl bei den Kommunikationsabläufen als auch beim Aufbau der Dokumente lassen sich mit den hergebrachten Techniken der archivischen Sicherung nicht mehr bewältigen. Die ersten *Webeiten* konnte man noch ausdrucken, sie waren weder umfangreich, noch vielschichtig verlinkt oder aufwendig graphisch gestaltet. Mit der Zunahme von Qualität und Umfang der *Webseiten* wurde bald klar, daß dies keine auf Dauer generell anwendbare Methode ist. Auch wenn kleinere technische Probleme beim **Ausdrucken** wie etwa der immer wieder vorkommende Textverlust am rechten Rand erfolgreich bewältigt werden können, gehen wesentliche Funktionselemente einfach dadurch verloren, daß man eine mehrdimensionale Struktur auf zwei Dimensionen plattdrückt. Die Argumente zugunsten des Ausdrucks sind damit nicht gänzlich ausgeräumt. Die Unwägbarkeiten einer rasanten technischen Entwicklung auf der einen Seite und die gewohnte Stabilität von Papier als Datenträger auf der anderen Seite lassen es immer wieder passend erscheinen, von wichtigen digitalen Dokumenten für Archivzwecke Papierkopien anzufertigen. Mag eine solche Sicherungs- und Konservierungsmethode für elektronische Fachzeitschriften durchaus erwägenswert sein, [2] so schien sie für die *Webseiten* der Jüdischen Gemeinden, über deren dauerhafte Sicherung das Zentralarchiv anfing nachzudenken, nicht wirklich geeignet zu sein. Die jüdischen *Webseiten* ähneln eher einem Puzzle von kurzen Notizen als einer Sammlung von gewichtigen Texten oder Katalogen, die lediglich elektronisch verbreitet werden. Wenn man bemüht ist, ein Bild von der Art und Weise festzuhalten, wie die jüdischen Gemeinden und Gruppen mit Hilfe des Internets öffentlich kommunizieren, dann kommt eigentlich nur die dauerhafte Erhaltung der *Webseiten* in ihrer ursprünglichen digitalen Form ernsthaft in Betracht.

---

[2] Siehe z.B. den Artikel *Les avantages comparés du papier et de la Toile* in Le Monde vom 22.1.1999, p. 21, wo es vor allem um elektronisch publizierte Fachzeitschriften geht.

Recht bald wurde auch klar, daß der Archivar bei dieser Aufgabe viel mehr unter **Zeitdruck** steht, als er es von seinen traditionellen Erledigungen her gewohnt war. Es ist bekannt, daß sich die Lebensdauer der erzeugten Informationen mit jeder neuen Technik weiter verringert. [3] Ich möchte nicht bis zu den Tontafeln der Assyrer zurückgehen. Für uns ist schon der Unterschied zwischen *Webseiten* und herkömmlich gedruckten Mitteilungsblättern beeindruckend. Mitteilungsblätter kann man noch Jahrzehnte nach ihrem Erscheinen in eine Sammlung übernehmen. Wie erste Erfahrungen gezeigt haben, kann man bei Internet-Publikationen nicht so lange warten. Was nicht unmittelbar nach seiner Erzeugung in ein archivisches Sicherungssytem gelangt, ist schon kurze Zeit später nur noch mit den allergrößten Mühen oder manchmal auch überhaupt nicht mehr zu erlangen. Auch wenn von den Betreibern vor Ort meist Abzüge ihrer *Webseiten* auf CD abgelegt werden, ist der Zugriff darauf mit einem viel zu hohen Verhandlungsaufwand verbunden. Manche studentischen oder anderen Gruppen sind von einer solchen Instabilität, daß ihre *Webseiten* nach wenigen Monaten spurlos aus dem Internet verschwinden und der Archivar auch nicht weiß, an wen er sich danach noch wenden soll. Es gibt natürlich auch Gemeinden, Verbände und Einrichtungen, die schon einige Jahrzehnt existieren und bei deren *Webseiten* man auf eine ebensolche Beständigkeit rechnen kann.

Trotzdem läßt sich wohl mit einiger Sicherheit sagen, daß in Abwesenheit verbindlicher Regelungen eine systematische **Übernahme** der im jüdischen Bereich vorhandenen *Webseiten* direkt vom Erzeuger eine Illusion ist. Nur ein ganz geringer Teil der inzwischen fast siebzig Erzeuger wäre in der Lage und interessiert, sich an einem so komplizierten Anmelde- und Ablierferungsverfahren zu beteiligen, wie es etwa ***Die Deutsche Bibliothek*** für einen ganz anderen Kundenkreis entwickelt hat. [4] Verleger elektronischer Fachzeitschriften oder Universitätsbibliotheken, die Dissertationen über das Internet zugänglich machen, sind in dieser Hinsicht kaum zu vergleichen mit Studentengruppen die lediglich ein Interesse an schneller und wirkungsvoller Verbreitung aktueller Informationen über Vortragsabende und Wochenendseminare haben. Das einzige für uns denkbare Verfahren war also eine Sammlung von Kopien der einschlägigen *Webseiten* direkt aus dem Internet heraus.

Diese Orientierung wird noch durch eine andere Beobachtung nahe gelegt. Bei den *Webseiten* aus dem Leben der jüdischen Gemeinschaft in der Bundesrepublik handelt es sich in der Regel nicht um gewichtige Dokumente oder Dateien, die einmal erzeugt und dann für lange Zeit elektronisch verfügbar gehalten werden. Sie zeichnen sich vielmehr durch schnelle und

---

[3] Dieter E. Zimmer, Das große Datensterben, in: Die Zeit vom 18. November 1999, S. 45f.
[4] http://deposit.ddb.de/netzpub/web_abgabe_np_verfahren.htm

unübersichtliche Veränderlichkeit aus. Im Unterschied zu den auf Papier verschickten Mitteilungen gibt es bei den Mitteilungen im Internet meist keine klar erkennbare **Periodizität**. Die verschiedenen Teile einer Internet-Präsenz werden auch selten zum gleichen Zeitpunkt geändert, es gibt weitgehend konstante Bereiche und ausgesprochene News-Seiten. Die Periodizität, die sonst vom Herausgeber etwa eines Mitteilungsblattes vorgegeben wird, muß bei den *Webseiten* erst im Zuge der Archivierung ermittelt bzw. festgelegt werden. Dies ist der Punkt, wo eine neue Art von archivischer Bewertung einsetzen müßte. Es ist weder sinnvoll noch mit vertretbarem Aufwand machbar, alle Ergänzungen oder Veränderungen, die an den zur Archivierung ausgewählten *Webseiten* vorgenommen werden, für die Ewigkeit festzuhalten. Ähnlich wie bei massenhaft gleichförmigen Einzelfallakten sollte also auch für die Aufbewahrung von *Webseiten* ein Archivierungsmodell erarbeitet werden. Bisher verfügen wir nur über Ansätze dazu. So gehen wir z.B. davon aus, daß es vor allem darum gehen sollte, von Zeit zu Zeit festzuhalten, auf welche Art und Weise die jüdischen Gemeinden und Gruppen mit ihren Mitgliedern und mit der Außenwelt kommunizieren. Natürlich gibt es auch immer wieder wichtige Inhalte, die nicht verloren gehen sollten. Um hier richtig abwägen und entscheiden zu können, bedarf es auf jeden Fall einer inhaltlichen Wahrnehmung der *Webseiten* im Zusammenhang mit ihrer Archivierung. Und genau das erfolgt nicht bei der in den USA entwickelten Methode zur Archivierung von *Webseiten*, die ansonsten erhebliche Vorzüge aufzuweisen hat. Seit 1996 bemüht sich das ***Internet Archive*** [5] in San Francisco um regelmäßige Archivierung sämtlicher überhaupt existierender *Webseiten*. Das gesamte Internet wird zu diesem Zweck von sogenannten Robotern abgesucht, die sich von *link* zu *link* vorarbeiten. Die Häufigkeit der Sicherung hängt bei diesem Verfahren stark vom Zufall ab. So sind auf dem Server in San Francisco z.B. von den *Webseiten* der *Jüdischen Gemeinde Düsseldorf* im Jahr 2001 insgesamt 10 Sicherungen abgelegt worden. Die meisten dieser Kopien stammen jedoch vom Anfang des Jahres, während zwischen Mai und Dezember eine Sicherungslücke von insgesamt sieben Monaten entstanden ist. Das ist auch nicht der einzige Mangel dieses Verfahrens. *Webseiten* kleiner Gruppen, auf die von anderen nicht mit einem *link* hingewiesen wird, können durch die Suchroboter nicht gefunden und damit auch nicht kopiert werden. Und auch *Webseiten*, die nicht lange existieren, können leicht der Sicherung entgehen.

Auf der anderen Seite ist die technische Leistungsfähigkeit des *Internet Archives* in San Francisco so beeindruckend, daß es nicht so leicht zu überbieten ist. Über eine sogenannte *Wayback Machine* lassen sich die Ergebnisse der Archivierung sofort *online* aufrufen und einsehen. Man muß nur die zugehörige URL eingeben und schon erhält man eine Übersicht über alle Sicherungsdaten. Bei diesem Suchverfahren gibt es Bedarf für archivische Nachbesserung. Wenn ein

[5] http://www.archive.org/

Historiker die Geschichte einer jüdischen Gruppierung studieren möchte, dann kennt er zwar ihren Namen aber noch nicht unbedingt die mitunter wechselnden Internet-Adressen, unter denen diese Gruppierung in den zurückliegenden Jahren ihre *Webseiten* veröffentlicht hat. Das hier angedeutete Informationsdefizit wird nun vom Heidelberger Zentralarchiv auf seinen eigenen *Webseiten* ausgeglichen. Nach Sachgruppen geordnet findet sich dort eine Aufstellung aller einschlägigen *Webseiten* für das organisierte jüdische Leben der Bundesrepublik. [6] Nach dem Namen des Herausgebers werden jeweils die verwendeten Internet-Adressen mitgeteilt und mit einem kurzen Kommentar zu den Zeiten ihrer Gültigkeit versehen, soweit sich das ermitteln ließ. Anschließend wird per *link* nur noch auf die zugehörige Übersicht des *Internet Archives* in San Francisco verwiesen. Auch wenn man einräumen muß, daß der erst seit Oktober letzten Jahres existierende *online*-Zugang zum *Internet Archiv* mit Hilfe der *Wayback Machine* noch einige Defizite aufweist, manchmal fehlt die Graphik oder einige *links* funktionieren nicht, so ist mit dieser neuen Technik ein Anfang gemacht, dem nur weitere Perfektion folgen kann. Es gibt wohl kein anderes Archiv auf der Welt, das seinen gesamten Inhalt, und nicht nur die Verzeichnisse, über Internet einsehbar macht.

Um nun bei der Archivierung der deutsch-jüdischen *Webseiten* nicht in vollständige Abhängigkeit von der amerikanischen Internet Community zu geraten, werden die ermittelten Webseiten auch in Heidelberg von Zeit zu Zeit mit Hilfe einer eigens dafür im Zentralarchiv installierten Software [7] heruntergeladen und auf CD-ROM [8]gespeichert. Der Aufwand, mit dem diese eigene Datensammlung betrieben wird, ist in gewisser Weise umgekehrt proportional zur Stabilität und Zugänglichkeit der Datensammlung in den USA. Gegenwärtig arbeitet das *Internet Archive* in San Francisco als *public nonprofit organization*, befindet sich aber in starker finanzieller Abhängigkeit von der *Web Information Company Alexa*, die sich wiederum als Tochtergesellschaft von *Amazon.com* definiert. Die Verhältnisse sind aus der Ferne nicht so leicht zu durchschauen. Auf jeden Fall scheint das Archiv in San Francisco von den Schwankungen und Ungewißheiten der sogenannten *new economy* nicht ganz unabhängig zu sein. Es wäre wohl gewagt, sich vollständig auf diesen Service zu verlassen. Andererseits ist bereits die Rede von der Übergabe einer Kopie der im *Internet Archive* gespeicherten Daten an die *Library of Congress*. [9] Auch wenn das Unbehagen damit nicht vollkommen beseitigt ist, muß man vielleicht im Zeitalter des Internet die auf das Internet bezüglichen Archvierungsprobleme nicht unbedingt im nationalstaatlichen Rahmen lösen. So wie sich das Internet von Anfang an als ein weltumspannendes Kommunkationssystem entwickelt

---

[6] http://zentralarchiv.uni-hd.de/B8/WWW/
[7] Das Zentralarchiv benutzt den Webseitenkopierer WinHTTrack.
[8] Das Zentralarchiv benutzt recordable Gold CDs der Firma MITSUI
[9] Brewster Kahle, Direktor des Internet Archives, am 21.1.2002 in einem Interview mit Richard Koman. Text unter http://www.oreillynet.com/pub/a/webservice/2002/01/18/brewster.html

hat, ist es möglicherweise auch am sinnvollsten, die Archvierung seiner Inhalte auf derselben Ebene zu lösen. Große Nationalbibliotheken können und müssen vielleicht ihre eigenen Datensammlungen aufbauen, [10] aber für eine kleine Einrichtung wie das Heidelberger Zentralarchiv stellt das Angebot aus San Francisco eine willkomene Dienstleistung dar, die man mit den zur Sprache gebrachten Einschränkungen durchaus zur Lösung der eigenen Aufgaben heranziehen kann.

Im Laufe meiner bisherigen Ausführungen sind unter dem Gesichtspunkt der Datenübernahme **drei verschiedene Verfahren** zur Sprache gekommen, die von der *Deutschen Bibliothek* praktizierte Ablieferung durch den Erzeuger, sowie die auf unterschiedliche Weise realisierten Kopierverfahren direkt aus dem Internet, einmal mit hoher technischer Perfektion vom *Internet Archive* in San Francisco mit Hilfe von Webrobotern, und dann relativ elementar, sozusagen mit einem Verfahren für den Hausgebrauch vom Heidelberger Zentralarchiv durch gezielte Anwendung eines Webkopierprogramms. Ein Vergleich dieser drei Verfahren ist auch unter anderen Gesichtspunkten von Interesse. Da ist z.B die **Frage nach der kleinsten Einheit** für die Signaturvergabe. *Die Deutsche Bibliothek* sagt ganz explizit, daß sie nicht Dateien sammelt, sondern Dokumente. [11] In dem so definierten Vorgehen artikuliert sich unmißverständlich eine bibliothekarische Traditionen der Bestandsbildung. Die Aufmerksamkeit gilt der relativ stabilen textlichen Einheit, sei es ein Buch, eine Zeitschrift oder auch nur ein einzelner Brief im Fonds der Handschriftenabteilung. Und diese Mentalität macht sich nun auch bei der Sicherung elektronische Daten bemerkbar. Übernahme- und Speicherverfahren sind auf *e-books*, elektronische Zeitschriften und andere relativ stabile Bestandteile des Internet Angebots orientiert. Im Grunde handelt es sich dabei weitgehend um klassische Erzeugnisse, mit dem einzigen Unterschied, daß sie nicht auf Papier, sondern elektronisch verbreitet werden.

Der Ansatz des *Internet Archive* in San Francisco ist davon vollkommen verschieden. Es wurde ein Archivierungsmodell entwickelt, das von der organisatorisch-technischen Struktur des Internet selbst ausging. Als des Aufbewahrens würdig befunden werden Abzüge des *World Wide Web* in seiner Gesamtheit mit allen dazugehörigen Vernetzungen, nicht nur herausgesuchte Einzelstücke. Und dementsprechend ist die Signatur, unter der ein Teil des Ganzen gefunden werden kann, auf ganz natürliche Weise seine URL, d.h. die elektronische Adresse, unter der es ursprünglich ins Netz gestellt wurde. Wie

---

[10] Zum *Nordic Web Archive*, einem Kooperationsprojekt der nordeuropäischen Nationalbibliotheken, siehe: http://nwa.nb.no/

[11] Auf den Webseiten der Deutschen Bibliothek heißt es in den einführenden Erläuterungen zur Abgae von Netzpublikationen (http://deposit.ddb.de/netzpub/web_abgabe_np_gesamt.htm) wörtlich: "Nicht die Sammlung von Dateien ist das Ziel, sondern die Sammlung von Dokumenten als logische und konsistente Einheiten."

sollte eine dauerhafte Sicherung, die von der Internet Community ausgeht, auch anders aussehen.

Als wir in Heidelberg daran gingen, die *Webseiten* jüdischer Gruppierungen auf CD-ROM abzulegen, mußten wir uns ebenfalls Gedanken über die Vergabe von Signaturen machen. Von vornherein war klar, daß wir uns mehr für den ganzen *site* als für einzelne auf ihm veröffentlichte Texte interessierten. Selbstverständlich konnte auch nicht die CD, auf welche wir die jeweiligen *Webseiten* umkopiert haben, Träger der Signatur sein. Die CD ist ja nur der Aufbewahrungsort, vergleichbar mit Archivkarton und Regal bei Akten aus Papier. Aber auch die Internet Adresse schien uns nicht die geeignete Einheit zu sein. Uns ging es ja mehr darum, die elektronische Publikationstätigkeit einer Gemeinde oder Gruppe zu doukmentieren. Und diese Herausgeber oder Betreiber von *Webseiten* wechseln mitunter den Server und damit auch die URL ihrer *Seiten*. Das ist vergleichbar mit dem Wechsel des Verlags durch die Herausgeber einer Fachzeitschrift. Für die Bibliotheken wäre das auch kein ausreichender Grund, der Zeitschrift eine neue Signatur zu geben. Wir haben uns also auf die archivarische Tradition besonnen und nach dem "Registraturbildner" gefragt. Und so wurde bei uns die Gemeinde oder Studentengruppe, die als Herausgeber von *Webseiten* in Erscheinung tritt, zum Träger der Signatur.

Aus der Logik dieser Entscheidung ergibt sich auch, daß wir uns nicht für alle möglichen Texte über jüdische Einrichtungen oder jüdische Themen interessieren, die irgendwo im Internet veröffentlicht werden, selbst dann nicht, wenn der Autor Vorsitzender einer Jüdischen Gemeinde ist. Dieses Verfahren, bei dem einmalig ein Text weitergegeben wird, der dann lange Zeit unverändert unter irgedeiner Adresse zu lesen ist, ähnelt der konventionellen Veröffentlichung eines Artikels in einer Zeitschrift und wir berücksichtigen solche Erzeugnisse allenfalls in unserer Pressedokumentation. Was wir für die elektronische Überlieferungsbildung im Zentralarchiv suchen, sind weniger Autoren einzelner Beiträge als vielmehr jüdische Betreiber bzw. Herausgeber von sich regelmäßig verändernden *Webseiten*. Mehr noch als bei den Druckerzeugnissen gibt es im elektronischen Bereich jedoch Grenzfälle und Übergänge. Immer wieder bringen kleinere Gruppen ihre Informationen auf den *Webseiten* größerer Gruppen oder Einrichtungen unter. Hier muß von Fall zu Fall entschieden werden. Entscheidend für die Aufnahme in unsere Sammlung ist, ob eine **jüdische Herausgebertätigkeit** feststellbar ist.

Die drei Verfahren zur Sicherung von Internet Inhalten lassen sich auch noch in anderer Hinsicht vergleichen, z.B. was die Beachtung der **Urheberrechte** anbelangt. Am leichtesten ist dies wohl bei dem Übernahmeverfahren der *Deutschen Bibliothek* möglich. Da mit jedem Erzeuger von elektronischen Publikationen ein direkter Kontakt hergestellt wird, lassen sich alle

diesbezüglichen Fragen bestens klären. Gegebenenfalls können auch Zugangbeschränkungen berücksichtigt werden. Im Unterschied dazu berücksichtigt das *Internet Archive* in San Francisco Rechtsfragen lediglich rückwirkend. Ohnehin werden nur solche Inhalte aus dem *Web* kopiert, die frei zugänglich sind. Problematisch kann jedoch die erneute Anbietung über einen eigenen Server werden. Sollte der Urheber einer *Webseite* Einspruch gegen dieses Verfahren erheben, dann ist das *Internet Archive* bereit, die strittigen Dateien sofort aus seinem Angebot herauszunehmen. [12]

Die im Heidelberger Zentralarchiv abgelegten Kopien auf CD-ROM sind in dieser Hinsicht weniger problematisch, da keine erneute Anbietung für ein großes Publikum erfolgt, sondern die Kopien ausschließlich für eine eventuelle spätere Nutzung durch einzelne Forscher bereitgehalten werden.

Aus zeitlichen Gründen will ich es bei diesen Andeutungen belassen und Ihnen jetzt nur noch rasch einen Blick auf die gegenwärtige Präsentation unseres Webseitenarchivs ermöglichen:

http://www.zentralarchiv.uni-hd.de/B8/WWW/index.html

[12] Siehe z.B. die Angaben zur Copyright Policy auf den Webseiten des *Internet Archives* (http://www.archive.org/terms/index.html) und die Äußerungen von Brewster Kahle gegen Ende seines bereits zitierten Interviews vom 21.1.2002 (s. Anm 8).