

Compact Memory – Ein DFG-Projekt zur retrospektiven Digitalisierung jüdischer Periodika im deutschsprachigen Raum

Im Internetarchiv jüdischer Periodika im deutschsprachigen Raum wird die Mehrzahl der jüdischen Zeitschriften und Zeitungen des 19. und 20. Jh. online kostenlos bereitgestellt. Die Periodika umfassen alle religiösen, kulturellen und politischen Aspekte des jüdischen Lebens in Mitteleuropa und stellen somit für die Erforschung des Judentums in der Neuzeit ein unschätzbares Quellenreservoir dar. Die Zeitschriften werden teilweise als Volltext, teilweise als Graphik mit entsprechender Indexierung zugänglich gemacht. Eine erweiterte Datenbanksuche eröffnet verschiedene Recherchemöglichkeiten. Zurzeit sind über 100 Periodika mit einer Gesamtzahl von ca. 700.000 Seiten im Netz.

Das Projekt ist ein Kooperationsprojekt des Lehr- und Forschungsgebiets Deutsch-jüdische Literaturgeschichte der RWTH Aachen, der Universitätsbibliothek Frankfurt am Main und der Bibliothek Germania Judaica Köln.

In the internet archive portal »Compact Memory« for Jewish periodicals in German-speaking countries the vast majority of Jewish newspapers and magazines published in the 19th and 20th century are available on-line at no charge. These periodicals encompass all religious, cultural and political aspects of Jewish life in Central Europe and are hence an invaluable original resource for the research of Judaism in modern times. The texts are being made available partly as full-text, and partly as graphic documents with the corresponding index options. An expanded search machine provides various points of access. At present there are over 100 periodicals with about 700,000 pages online. Compact Memory is a cooperative project of the Teaching and Research Group for German-Judaic Literary History at the Technical University of Aachen, the University of Frankfurt and the Germania Judaica library in Cologne.

DIE DATENBANK DER JÜDISCHEN ZEITSCHRIFTEN

Die jüdischen Periodika, von denen es weltweit nach einer Statistik des *Jüdischen Lexikons* von 1667 bis 1929 rund 5.000 gegeben hat, verteilt über 70 Länder in allen fünf Erdteilen, stellen für die Erforschung des Judentums seit der beginnenden Neuzeit ein nicht überschätzbares Quellenreservoir dar, auf das wohl alle mit Jüdischen Studien befassten Fächer immer wieder rekurrieren. Gerade weil in diesen Periodika alle religiösen, politischen und sozialen Richtungen innerhalb des Judentums vertreten sind und alle Bedürfnisse – wissenschaftliche, berufliche, literarische, pädagogisch-didaktische – artikuliert werden, lassen sie sich, wie es im *Jüdischen Lexikon* heißt, als »ein getreues Abbild des jüdischen Lebens« interpretieren.

Selbstverständlich wendet zwar jede der Disziplinen ihre eigenen Methoden an, um die jeweils interessierenden Informationen zu erschließen und zu deuten, doch eines ist ihnen gemeinsam: dass nämlich die Periodika – insbesondere infolge der Verluste im Dritten Reich – in ihrer großen Mehrzahl nur schwer zugänglich und vollständige Jahrgänge in der Regel nur an wenigen Bibliotheken zu finden sind. Selbst diese sind meist nur in verschiedenen Formen (Original-Papierausgabe, Reprint, Mikrofilm und -fiche) zu nutzen, was sowohl für die Benutzer (die für die Durchsicht

verschiedene Materialien bestellen und an verschiedenen Arbeitsplätzen einsehen müssen) als auch für das Bibliothekspersonal einen erheblichen zusätzlichen Arbeits- und Zeitaufwand bedeutet. Von daher bieten sich die Digitalisierung, die Erschließung und die zentrale Bereitstellung der jüdischen Periodika im Internet geradezu an, um die Jüdischen Studien weltweit zu fördern: Digitalisierungsprojekte sind die einzige ökonomisch vertretbare Möglichkeit, Archiv- und Bibliotheksbestände der Wissenschaft zur Verfügung zu stellen und ihrem drohenden Verfall und damit endgültigen Verlust entgegenzuwirken.

Die Vorteile, große Textbestände im Internet zu veröffentlichen, liegen vor allem für Quellen auf der Hand, die der Forscher wegen ihres hohen Informationswerts häufig frequentiert, die jedoch aufgrund ihres Umfangs oder ihrer ausschließlich wissenschaftlichen Nutzung wie auch wegen der damit verbundenen immensen Kosten nicht neu aufgelegt werden. Beschaffungsprobleme entfallen vollständig, da die Textcorpora per Internet von jedem privaten oder öffentlichen Arbeitsplatz abgerufen werden können. Im Gegensatz zum Buch oder Mikrofilm/-fiche kann der Bestand eines Internet-Textcorpus laufend erweitert und aktualisiert werden; selbst Software-Updates, wie sie im Fall der CD-ROM/DVD-Angebote üblich sind, entfallen.

Aus den skizzierten Vorteilen ergeben sich drei qualitative und quantitative Kriterien, an denen Digitalisierungsprojekte generell und auch dieses spezifische zu messen sind:

➤ **Datenangebot:** Der Aufbau eines digitalen Archivs kann nur dann eine Alternative zu anderen Publikationsarten bilden, wenn ein Textcorpus angeboten wird, das (a) einen bedeutenden Stellenwert für die Forschung besitzt, (b) schwer verfügbar ist, (c) häufig frequentiert werden muss sowie (d) langfristig nicht als Buch oder Mikrofilm/-fiche bereitgestellt werden kann.

➤ **Datenmenge:** Die Stärke digitaler Archive liegt in der technischen Möglichkeit, große bzw. größte Textmengen kostengünstiger erfassen, strukturierter verwalten und zügiger veröffentlichen zu können, als dies in herkömmlichen Publikationsformen der Fall ist. Die weitgehend automatisierte Erfassung und Digitalisierung umfangreicher Textbestände ist nicht nur in der



Hans Otto Horch

Foto privat

1667–1929:
5.000 jüdische Periodika

Verluste in der NS-Zeit

Herstellung preiswerter als Editionen in Buch- oder Mikroform – sie steht dem Nutzer zudem unentgeltlich zur Verfügung. Digitalisierungsprojekte sollten daher bewusst auf ›Masse‹ setzen und Textcorpora bereitstellen, die mindestens einige zehn- oder hunderttausend Seiten umfassen.

► **Datenzugang:** Ein Digitalisierungsprojekt darf sich nicht darauf beschränken, Quellentexte als schlichte ›Loseblattsammlung‹ anzubieten. Vielmehr haben die Art und Weise der Erschließung und Bereitstellung der Quellen sowie der Zugriff auf die Textcorpora die hergebrachten Möglichkeiten des Buchs, der Mikroform oder der CD-ROM zu übertreffen. Die Entwicklung technologisch innovativer Lösungen, z. B. auf der Basis spezieller Formatsprachen, hoch strukturierter Datenbankmodelle oder kombinierter Darstellungsverfahren, muss eine grundsätzlich ›neue Sicht‹ auf die Texte und Materialien unterstützen, die etwa Register, Indices oder Volltextsuche nicht oder nur eingeschränkt vermitteln können. Es versteht sich dabei von selbst, dass der Hauptvorteil gegenüber gängigen Publikationsformen – die globale Verfügbarkeit der Textcorpora – nicht zugunsten hoch technisierter Anforderungsmerkmale geopfert werden darf, sondern die in der Regel technisch mindere Ausstattung der Nutzer bzw. deren eingeschränkte Kenntnisse zu berücksichtigen sind.

Im Aachener Lehr- und Forschungsgebiet Deutsch-Jüdische Literaturgeschichte standen jüdische Periodika seit seiner Gründung 1992 im Mittelpunkt des Interesses. Bereits vorher gab es am Aachener Germanistischen Institut im Zusammenhang mit der Erstellung von Autoren- und Werkindices sowie Wörterbüchern Erfahrungen mit EDV.¹ So lag es nahe, im Anschluss an ein von Ende 1998 bis Ende 1999 vom Ministerium für Schule, Wissenschaft und Forschung des Landes Nordrhein-Westfalen gefördertes Pilotprojekt zur Digitalisierung der alljüdischen Revue »Die Freistatt« bei der Deutschen Forschungsgemeinschaft ein Großprojekt zu beantragen: das auf sechs Jahre angelegte, im Frühjahr 2006 abgeschlossene Projekt »Retrospektive Digitalisierung jüdischer Periodika im deutschsprachigen Raum« (www.compactmemory.de). Das Projekt wurde allerdings nicht im Rahmen des allgemeinen Antragsverfahrens, sondern im Bereich der Bibliotheksförderprogramme, speziell in dem eigens geförderten Bereich »Wissenschaftliche Literaturversorgungs- und Informationssysteme / Kulturelle Überlieferung« angesiedelt, in dessen Rahmen der sukzessive, nachhaltige Aufbau und Ausbau einer *Verteilten Digitalen Forschungsbibliothek* angestrebt wird.² Dies bedeutete die Fokussierung auf die mediale Grundlage der Periodika, was eine weitergehende semantische Aufberei-

tung ausschloss – eine Zurückhaltung, die für moderne Editionsprojekte aller Art typisch ist. Als Kooperationspartner auf der Bibliotheksseite konnten das Sondersammelgebiet ›Wissenschaft vom Judentum‹ der Universitätsbibliothek Frankfurt am Main (Leitung: Dr. Rachel Heuberger) und die Bibliothek Germania Judaica in Köln (Leitung: Dr. Annette Haller) gewonnen werden, deren umfangreiche Bestände an jüdischen Periodika nicht nur die Materialbasis des Unternehmens boten, sondern die schwerpunktmäßig die bibliothekarische Vorbereitung des Projekts und die Bereitstellung der Periodika übernommen, die laufende bibliografische regelgerechte Erschließung garantiert sowie die Überprüfung der Digitalisate sichergestellt haben. Hierzu zählten u. a. die akribische Sichtung der zu digitalisierenden und verfilmenden Periodika, die Protokollierung der Lücken und mangelhaften Digitalisate sowie die Ergänzung fehlender Hefte und die Beschaffung von Ersatzvorlagen, was äußerst zeitintensive Recherchen und Korrespondenzen erforderte.

DIE TECHNISCHE UMSETZUNG

Neben der bibliothekarischen Organisation des Projekts, d. h. der Zusammenführung, Sichtung und Katalogisierung der verstreut und lückenhaft überlieferten Quellen zu einem weltweit einzigartigen, annähernd vollständigen Corpus, stellte die technische Durchführung des Digitalisierungsvorhabens eine wesentliche Herausforderung dar. Mit Blick auf die oben aufgeführten Kriterien hinsichtlich des zu realisierenden Datenangebots, der zu verarbeitenden Datenmenge sowie der unterschiedlichen Datenzugänge war es erforderlich, eine effiziente Produktionslinie aufzubauen, um die zeitnahe Bereitstellung großer, heterogener Datenbestände – Originalgrafiken und bibliothekarische Metadaten – zur frei zugänglichen Nutzung im Internet zu gewährleisten.

Eine zentrale Anforderung bestand vor allem darin, der zum Projektende enorm angewachsenen Datenmenge – insgesamt rund 700.000 Einzelseiten, verteilt auf 118 Periodika – Herr zu werden und den Nutzern per Internet am individuellen Arbeitsplatz einen gleichermaßen ökonomisch wie ergonomisch sinnvollen und intuitiven Zugriff zu ermöglichen. Zugleich war sicherzustellen, dass die Nutzer ohne Zeitverzug bald nach Projektstart bereits auf erste Ergebnisse in Form eines strukturierten Archivs navigier- bzw. skalierbarer Grafiksammlungen zugreifen konnten. Maßgebend war in diesem Kontext das etablierte Prinzip des ›Browsing and Searching‹: Der Benutzer sollte einerseits wie in einer ›realen‹ Bibliothek ein beliebiges Periodikum über einen Navigationsbaum auswählen können, um über die einzelnen Jahrgänge zu den ge-

wünschten Einzelheften zu gelangen; andererseits mussten verschieden differenzierte Suchmasken die gezielte Recherche nach gesuchten Materialien erlauben.

Die angemessene Präsentation des Corpus, dessen Bereitstellung zudem zeit- und kostensparend erfolgen musste und die technischen Fertigkeiten der Projektmitarbeiter zu berücksichtigen hatte, beruht maßgeblich auf der Qualität der zugrunde gelegten Datenbanklösung und Anwendungssoftware. Zum Einsatz kam das spezifisch auf den Bedarf von Digitalisierungsprojekten zugeschnittene Produkt »Visual Library« der Aachener Firma *semantics*, dessen Einzelmodule die strukturierte Erfassung, Indizierung, Volltexterkennung, Bearbeitung und Bereitstellung beliebiger grafischer und textueller Materialien im Internet ermöglicht.³

Als Arbeitswerkzeug des Digitalisierungsprojekts können mit dem zentralen Modul, dem *Library Manager*, große Mengen von Grafiken übersichtlich, strukturiert und schnell per drag-and-drop auf einen lokalen Datenbankserver überspielt werden (ca. 1.000 images pro Stunde) und mit Hilfe eines integrierten Grafikbetrachters im Zielverzeichnis einzeln oder in Form von thumbnails aufgerufen und nachbearbeitet werden. Über einen Navigationsbaum, der die serverinterne Zielverzeichnisstruktur abbildet, legt der Bearbeiter neue, annotierbare Zeitschriftentitel, Jahrgänge oder Hefte an, wobei jedes image zudem typisierbar ist (Titelblatt, Inhaltsverzeichnis, Artikel usw.).

Im Kern liegt die Stärke des *Library Manager* darin, den Zeitaufwand zwischen der Erfassung bzw. Einspeisung der Digitalisate und deren Präsentation im Internet auf ein Minimum zu verringern und den Nutzern das bearbeitete Material ohne nennenswerten Verzögerung in Form dynamisch generierter Webseiten zu Recherchezwecken zur Verfügung zu stellen. Gegliedert nach Zeitschriftentitel, Jahrgängen und Einzelheften können die Nutzer bereits in der unmittelbaren Anfangsphase eines Digitalisierungsprojekts durch das im Aufbau befindliche Corpus »browsen«. Waren bislang aufwändige Bibliotheks- und Archivaufenthalte nötig, werden die Nutzer dieses bloße Vorhandensein bislang schwer zugänglicher Quellen, seien diese auch vergleichsweise flach erschlossen, bereits als wichtige Arbeitserleichterung zu schätzen wissen. Die Möglichkeit des »Searching«, d. h. der gezielten Suche nach einzelnen Textstellen, wird anschließend in einem zweiten, optionalen Schritt durch die automatisierte Volltexterkennung erreicht, die der *Library Manager* ebenfalls bereitstellt.

Dieser Grad der Erschließung – die strukturierte Präsentation der Originalgrafiken nebst integrierter

Volltextsuche – mag hinsichtlich jener Corpora genügen, die vor allem monografisches Schrifttum zusammenführen. Als separater Produktionsschritt wurden in *Compact Memory* darüber hinaus die technischen Möglichkeiten geschaffen, getrennt von den Originalgrafiken und den Volltexten direkt auch auf die bibliografischen Kerndaten der in den Periodika erschienenen Einzelbeiträge zuzugreifen. Zwar erfordert diese normkonforme Katalogisierung unselbstständig erschienener Literatur einen weitaus höheren Arbeitsaufwand, die Vorteile sind jedoch evident: Zum einen werden die Nutzer entlastet, insofern sie nicht bei jeder neuen Fragestellung immer wieder eine Unzahl von Grafiken oder Volltexten nach den gesuchten Materialien autopsieren müssen – eine zeitraubende Prozedur, die bereits die Arbeit mit Mikrofilmen oder papierenen Vorlagen erschwert hat. Zum anderen erhöht die Erfassung der Einzelbeiträge nach den »Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken« (RAK-WB) den wissenschaftlichen Wert der Datenbank um ein Vielfaches, zumal in *Compact Memory* Metadaten und Verknüpfungen berücksichtigt werden, die RAK-WB nicht oder nur ungenügend abdeckt (Rubriken, Inserate, Illustrationen etc.). Resultat der bibliografischen Erfassung ist eine Datenbank, die bereits eine volle inhaltliche Erschließung der einzelnen Komponenten der Zeitschrift leistet und damit Informationen liefert, die weit über die reine Formalerschließung hinausgehen. So werden inhaltliche Bezüge zwischen den einzelnen Aufsätzen hergestellt, indem nachfolgende Berichtigungen, Ergänzungen oder Reaktionen miteinander verknüpft werden. Da ferner die Rubriken mit ihren unterschiedlichen Ebenen bibliografisch als Sachtitel erfasst bzw. ihre Inhalte durch Schlagworte wiedergegeben werden, erfolgt eine systematische Sacherschließung, die den Nutzern weitaus exaktere Suchmöglichkeiten und -ergebnisse als bisher üblich liefern kann.

Das kompromisslose Festhalten an den Regeln von RAK-WB trotz des großen Zeitaufwandes hat dazu geführt, dass eine Datenbank entstanden ist, deren Aufnahmen den Normen der alphabetischen Katalogisierung des wissenschaftlichen Bibliothekswesens entsprechen und die damit Modellcharakter für weitere Datenbanken hat. Die detaillierte bibliografische Erfassung der einzelnen Komponenten der Zeitschriften macht die Datenbank zudem zu einem Informationsmedium für die *Wissenschaft des Judentums*, welches die bibliografischen Informationen vergleichbarer Zeitschriftendatenbanken wesentlich übersteigt.

Mit der Freischaltung der bedeutenden, lang erscheinenden Zeitschrift *Der Israelit* (120.000 Seiten) bietet *Compact Memory* nach der *Allgemeinen Zei-*

normkonforme
Katalogisierung

Inserate und Illustrationen:
bei den RAK nicht vor-
gesehene Metadaten

Modellcharakter für
ähnliche Vorhaben

tion des Judentums, der Jüdischen Rundschau und anderen ähnlich umfangreichen Periodika alle ›großen‹, über mehrere Jahrzehnte erschienenen jüdischen Zeitschriften und Zeitungen Deutschlands an. Das Vorhaben, ein repräsentatives Corpus jüdischer Periodika im deutschsprachigen Raum bereitzustellen, wurde in erfolgreicher Kooperation mit der Wiener Universitätsbibliothek realisiert, die maßgebende jüdische Zeitschriften und Zeitungen Österreichs für die Digitalisierung zur Verfügung stellte.

Der im Internet unter www.compactmemory.de zugängliche Datenbestand umfasst ca. 700.000 Einzelseiten bzw. 137 GB Grafiken (und 300 GB images für die Entwicklung). Ca. 81.000 Einzelbeiträge wurden bibliografisch erschlossen und stehen für Recherchezwecke zur Verfügung. Der Bestand der Digitalisate wird ergänzt durch ca. 10.370 recherchierbare Stammdaten zu Autoren (incl. Pseudonyme) und Körperschaften. Alle in Antiqua gesetzten Periodika sind volltexterkannt und in die Suchinterfaces der Website integriert. Mit Hilfe der in Kürze im Projekt zum Einsatz kommenden Version 8.0 der OCR-Software *FineReader*, von deren ausgezeichneter Erkennungsqualität sich die Projektnehmer anlässlich einer Präsentation überzeugen konnten, wird deutlich mehr Frakturtext als ursprünglich geplant für Recherchezwecke bereitgestellt werden können. Im Zuge einer mit dem Softwarehersteller ABBYY geschlossenen Vereinbarung wird dem Digitalisierungsprojekt das *FineReader*-Frakturmodul kostenlos zur Verfügung gestellt. *Compact Memory* beteiligt sich im Gegenzug an der Evaluation des OCR-Programms.

Damit bietet die Datenbank *Compact Memory* mit 118 Periodika das umfangreichste digitalisierte jüdische Zeitschriftenarchiv an und ist für die internationalen Jüdischen Studien ein unersetzliches Werkzeug geworden, was nicht zuletzt durch die zahlreichen regelmäßigen Zugriffe aus aller Welt eindrucksvoll belegt wird.

¹ Vgl. Horch, Hans Otto: Auf der Suche nach der jüdischen Erzählliteratur: Die Literaturkritik der ›Allgemeinen Zeitung des Judentums‹ (1837–1922). Bern, Berlin, Frankfurt a. M., New York, Paris, Wien: Lang 1985 (Literaturhistorische Untersuchungen; 1); Horch, Hans Otto; Dannhauer, Heinz-Martin; Schuffels, Klaus; in Verb. mit Kammer, Manfred; Rüter, Eugen: Wörterbuch zu Friedrich Hölderlin. I. Teil: Die Gedichte: Auf der Textgrundlage der Großen Stuttgarter Ausgabe. Tübingen: Niemeyer 1983 (Indices zur deutschen Literatur; 10/11); Horch, Hans Otto; Schuffels, Klaus; Kammer, Manfred; in Verb. mit Vogel, Doris; Zimmermann, Hans: dass., II. Teil: Hyperion. Auf der Textgrundlage der Großen Stuttgarter Ausgabe. Tübingen: Niemeyer 1992 (Indices zur deutschen Literatur; 19).

² Zum Förderprogramm vgl. www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/projektfoerderung/foerderziele/index.html (02.01.2006).

³ Nähere Angaben zum Funktionsumfang der *Visual Library* unter www.semantics.de/produkte/visual_library/ (02.01.2006).

DER VERFASSER

Prof. Dr. Hans Otto Horch, RWTH Aachen, Institut für Germanistische und Allgemeine Literaturgeschichte, Lehr- und Forschungsgebiet Deutsch-jüdische Literaturgeschichte, Templergraben 55, 52056 Aachen, h.horch@germlit.rwth-aachen.de / hans-otto.horch@t-online.de