

Kai Eckert, Rachel Heuberger und Marko Knepper

JudaicaLink und der FID Jüdische Studien

JudaicaLink and the specialist information service (FID) Jewish studies

<http://doi.org/10.1515/bd-2018-0032>

Zusammenfassung: In der Forschungslandschaft Deutschlands fehlte es bislang an adäquaten fachspezifischen wissenschaftlichen Informationsdienstleistungen zu Jüdischen Studien und Israel-Studien. Der Fachinformationsdienst (FID) Jüdische Studien soll dazu beitragen, diese Lücke zu schließen. Im Vordergrund steht das FID-Portal als zentrale Anlaufstelle und einzigartiges Nachweis- und Recherche-Tool, das die Fachinformationen der Jüdischen Studien in ihrer gesamten Reichweite bündelt und durch die Aufbereitung und Anreicherung der Metadaten sowie der Verknüpfung externer Ressourcen durch JudaicaLink den Wert der zur Verfügung gestellten Informationen vervielfacht.

Schlüsselwörter: Fachinformationsdienst, Jüdische Studien, JudaicaLink

Abstract: Up to now, there were no adequate specialist information services for scientists concerning Jewish studies and Israel studies in Germany's research scene. The specialist information service (Fachinformationsdienst (FID)) Jewish studies is a contribution to close this gap. Most important is the FID portal as central platform and unique tool for research and proof-finding which unites specialist information of Jewish studies in their entire range. It greatly increases the worth of the information available by adapting and accumulating metadata as well as interlinking external resources via JudaicaLink.

Keywords: specialist information service, Jewish studies, JudaicaLink

Kai Eckert: eckert@hdm-stuttgart.de

Rachel Heuberger: r.heuberger@ub.uni-frankfurt.de

Marko Knepper: m.knepper@ub.uni-mainz.de

1 Einleitung

Der Fachinformationsdienst fokussiert ganz darauf, ein Recherchewerkzeug zum Nachweis fachspezifischer Informationen bereitzustellen. Der Grundgedanke beim Entwurf des Fachinformationsdienstes bestand darin, sich auch in technischer Hinsicht den spezifischen Problemstellungen des Fachgebietes zu stellen. Es sollte ein Werkzeug entstehen, das über die Aggregation relevanter Datenbestände in Kombination mit der Verwendung zeitgemäßer Nachweis- und Recherchetechniken hinaus technische Lösungen in Angriff nimmt, die für die Jüdischen Studien von besonderem Interesse sind.

Unter diesem Aspekt wurden zwei spezifische Anforderungsfelder identifiziert: Zum einen ist der Umgang mit transliterierten Metadaten aus dem Hebräischen problematisch. Eine technische Lösung für die Rückführung transliteriert vorliegender hebräischer Metadaten in die Originalschrift würde nicht nur im vorliegenden Projekt neue Optionen für die Recherche in den Jüdischen Studien bieten.

Zum anderen sind die Nachweise für Literatur und andere Informationsquellen zu den Jüdischen Studien ausgeprägt über den Globus verteilt, so dass eine Vernetzung über technische und räumliche Grenzen hinweg einen besonderen Mehrwert für die Beschaffung von Informationen zu bestimmten Fragestellungen darstellt. Die Verknüpfung der Daten als Linked Open Data (LOD) mit externen Ressourcen kann hier einen spezifischen Nutzen für das ganze Fachgebiet stiften und dient gleichzeitig als Use Case für die stetige Entwicklung offener Datenbestände hin zu global verknüpften Ressourcen.

Die Arbeiten zur Gewinnung der notwendigen Daten umfassen vor dem Hintergrund dieser Überlegungen im Wesentlichen drei Arbeitspakete, deren Resultate sich zum FID-Portal ergänzen. In den drei Arbeitspaketen kann jedoch weitgehend unabhängig vom Fortschritt der jeweils anderen beiden Arbeitspakete gearbeitet werden.

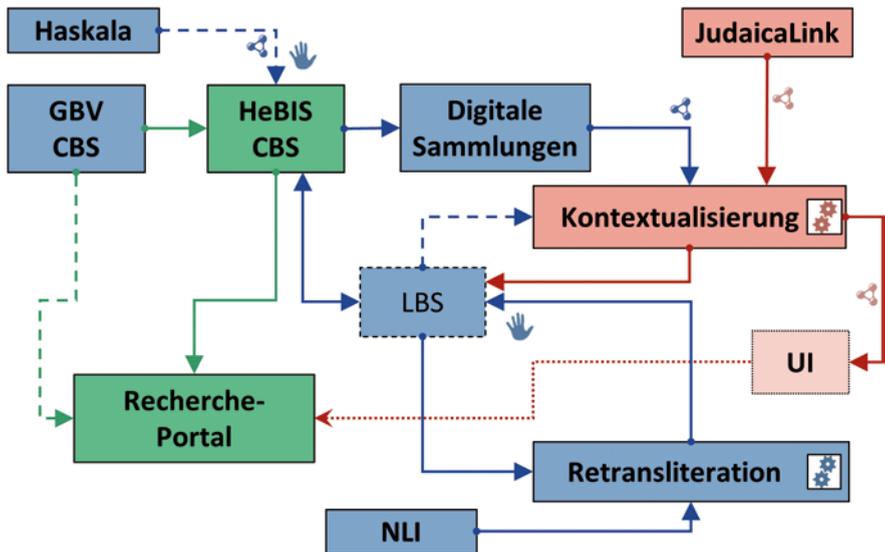


Abb. 1: Übersicht über die Datenflüsse der drei Arbeitspakete im Projekt.

Die drei Arbeitspakete werden nachfolgend im Einzelnen erläutert:

2 Entwicklung eines automatisierten Retrokonversionsverfahrens für transliterierte hebräische Titelinformationen der hebräischen und jiddischen Literatur in die hebräische Originalschrift

Im europäischen und anglo-amerikanischen Sprachraum erfolgt die Katalogisierung von Werken in hebräischen Schriftzeichen – so wie generell von Werken in anderen Schriftzeichen – auf der Grundlage der Konversion dieser Schrift in die lateinische als dominante Schrift zum Nachweis der Materialien in einem zentralen Katalogsystem. Im deutschen Bibliothekswesen handelte es sich bis zum Jahr 2006 um eine spezifische, für den deutschen Sprachraum eigens entwickelte Umschrift, die auf den Transkriptionstabellen der 1899 erschienenen „Instruktionen für die Alphabetischen Kataloge der Preußischen Bibliotheken“ basierte. Diese ursprünglich nur für die Preußischen Bibliotheken einschließlich

der Frankfurter UB verbindlichen Regeln wurden von anderen Bibliotheken übernommen und behielten bis zur Ablösung durch die Regeln für die Alphabetische Katalogisierung (RAK) im Jahre 1977 bzw. die Einführung der Regeln für die Alphabetische Katalogisierung in wissenschaftlichen Bibliotheken (RAK-WB) 1983 ihre Gültigkeit. Die mit den Preußischen Instruktionen entwickelten Transkriptionsregeln waren unter Berücksichtigung des Neuhebräischen die Grundlage der im April 1982 erlassenen Norm DIN 31636 *Umschrift des hebräischen Alphabets*, auf die in den RAK-WB verwiesen wird. Mit Einführung der DIN 31636 wird die hebräische Sprache in den Bibliothekskatalogen des deutschen Sprachbereichs (Deutschland, Österreich, deutschsprachige Schweiz) verbindlich umschrieben.

Die hebräische Sprache stellt eine Konsonantensprache dar, deren Vokalisierung mit Hilfe von Vokalzeichen unter den Konsonanten oder mit Konsonanten in der Funktion von Vokalen erfolgt. Deshalb bedarf es der genauen Kenntnis der hebräischen Grammatik, der Lexik und Phonetik sowie der Kenntnis der DIN sowohl zur Erstellung einer der Aussprache korrekten Umschrift bei der Erfassung der Daten als auch für den Nutzer zum Auffinden der Titel im Katalog. Da die DIN die Möglichkeit der eindeutigen Rückführung, der Retransliteration in die Ursprungssprache Hebräisch beabsichtigt, ergeben sich ungewohnte Buchstabenkombinationen in der lateinischen Schrift, die Folge komplizierter grammatikalischer Regeln sind und in der heutigen Praxis in Israel den meisten nicht geläufig sind. Während die rabbinischen und religionsgesetzlichen Texte in der Regel punktiert und damit klar vokalisiert sind, ist die Schreibweise des modernen Hebräisch, in der die gesamte wissenschaftliche Literatur im 20. Jahrhundert gedruckt ist, unpunktiert und zu großen Teilen auch in einer defektiven, d. h. die als Vokale fungierenden Konsonanten weglassenden Schreibweise. In der Regel ist ein Abgleich des Wortes in einem hebräischen Wörterbuch erforderlich, um die genaue transkribierte Schreibweise zu eruieren.¹

Der seit den Institutsgründungen zur Judaistik und Jüdischen Studien in den 80er und 90er Jahren in Deutschland stetig angestiegene Literaturbestand in neuhebräischer Sprache machte eine Überarbeitung der DIN 31636 erforderlich und führte zu einer Neuversion im Jahre 2006. Diese ist sowohl in der Wiedergabe der Grundbuchstaben des lateinischen Alphabets als auch in den Transliterationsregeln identisch mit den Regeln der American Library Association/Library of Congress (ALA/LoC Romanization Rules for Hebrew) aus dem Jahre 1987 und trägt

¹ Zusammenfassung in: Marquardt, Susanne: Transliteration und Retrieval. Zur Problematik des Auffindens hebräischsprachiger Medien in Online-Katalogen. Berlin 2005 (Berliner Handreichungen zur Bibliothekswissenschaft, Heft 157).

den Entwicklungen des modernen Hebräisch Rechnung.² Seit 2010 wird an Universitätsbibliotheken und in den Verbänden zusätzlich in bestimmten Feldern eine Titelaufnahme in der Originalschrift erstellt, so dass auch eine Recherche in der Originalschrift möglich ist.

In einem Zeitraum von über 100 Jahren, seit der Einführung der Transliterationsregeln im Jahre 1899 bis zur Revision in der DIN 31636 von 2006, wurden die hebräischen Bestände kontinuierlich nach den damals gültigen, d. h. „alten“ Umschriftregeln erfasst, so dass eine beträchtliche Anzahl hebräischer Titel in dieser mittlerweile schwer zugänglichen Form in den Katalogen nachgewiesen und auf Grund der wachsenden Unkenntnis der spezifischen Transliterationsregeln nicht auffindbar ist. Zahlenmäßig handelt es sich um rund 22.000 Titel mit dem Sprachcode Hebräisch in der Universitätsbibliothek Frankfurt aus den Jahren 1901–2006. Aus Personal- und Kostengründen ist eine intellektuelle Umarbeitung der Titel von der alten in die neue Transliterationsversion nicht realistisch.

Zur automatisierten Umarbeitung wurde im ersten Schritt eine Software entwickelt, die zunächst im analytischen Verfahren unter umgekehrter Anwendung der oben beschriebenen Regeln für die transliterierten hebräischen Worte des Hauptsachtitels die hebräische Originalschrift ableitet.

Die Umsetzung der nicht geschriebenen Vokale des Hebräischen führt in der Praxis häufig zu mehrdeutigen Ergebnissen, deren Plausibilität – beruhend auf der Wahrscheinlichkeit von Buchstabenfolgen – in einem mehrstufigen Verfahren festgelegt wird. Hierbei dienen punktierte, d. h. durch Diakritika vokalisierte hebräische Worte, wie sie in den hebräischen Bibeltexten geschrieben werden, als Richtschnur. Zur Festlegung des Rankings der korrekten Schreibweise innerhalb der möglichen Varianten wird auch auf „Hspell“, eine Open-Source-Datenbank für die hebräische Schreibweise und Wortbildungslehre, zurückgegriffen. Die erfolgten Ergebnisse werden mit Unterstützung der FID-Bibliothekarin überprüft und die Korrekturen vom Programm übernommen. Die bislang erzielten Ergebnisse sehen wie folgt aus:

- Überprüfte Worte = 1.344, davon richtig = 1.233, korrekte Trefferquote = 91,7%,
- Überprüfte Titel = 269, davon richtig = 189, korrekte Trefferquote = 70,4%.

2 Mahler, Paul: *Hebraica Cataloging*. Washington D.C. 1987.

Berücksichtigt man hierbei, dass 51 Worte der Ausgangstexte nicht entsprechend der DIN korrekt transliteriert wurden, dann erhöht sich die korrekte Trefferquote bei den Worten auf 95,5% und bei den Titeln auf 84,1%.³

Die Ergebnisse werden schließlich mit den Katalogdaten der Nationalbibliothek Israels (NLI) abgeglichen, welche die hebräischen Titel stets in der Originalschrift katalogisiert. Hierfür wurde der Datenabzug der relevanten Katalogdaten mit der NLI vereinbart. Anhand der positiven Treffer wird das Feld für den Eintrag „Verfasser“ überprüft und ergänzt. In weiteren Abfragen mit normiertem Vokabular werden die Verfasser mit der Normdatenbank VIAF⁴ abgeglichen.

Die gewonnen originalschriftlichen Daten werden in den Katalogbestand eingespielt und stehen so im FID-Portal zur Verfügung. Neben dem authentischen Nachweis steht in diesem Arbeitspaket als Ziel vor allem die Recherchierbarkeit im Vordergrund. Die Indexierung in Originalschrift als klar definiertem Standard macht eine Recherche in den nach unterschiedlichen Regeln transliterierten Titeln erst möglich.

3 Aggregation der fachspezifischen bibliografischen Nachweise aus den Katalogen der beiden Bibliotheken mit relevanten Sammelschwerpunkten UB Frankfurt und ULB Sachsen Anhalt in Bezug auf Israel-Studien und Jüdische Studien, sowie die Einbindung der bibliografischen Datenbank zur Literatur der Jüdischen Aufklärung (Library of the Haskala)

Ein umfassender Bestand an gedruckten und elektronischen Veröffentlichungen aus dem Frankfurter Sammelschwerpunkt zum Staat Israel ist unter besonderer Berücksichtigung der hebräischen Sprache im Katalog der UB Frankfurt nach-

³ Diese Abweichungen beruhen zum Teil auf durchaus zulässigen subjektiven Interpretationen der hebräischen Worte sowie auf einer von der Akademie der Hebräischen Sprache abweichenden Praxis der Autoren und Verleger. Die Open Source Datenbank HSpell richtet sich jedoch strikt nach den Regeln der Sprachakademie.

⁴ <https://viaf.org/> [Zugriff: 14.01.2018].

gewiesen. In der Virtuellen Fachbibliothek Vorderer Orient MENALIB der ULB Sachsen Anhalt in Halle⁵ sind die Publikationsnachweise mehrerer Bibliotheken für das dort früher angesiedelte Sondersammelgebiet (SSG) „Vorderer Orient einschließlich Nordafrika“ enthalten. Der Katalog hat sich auf Publikationen auf Arabisch (sowie Türkisch und Persisch) spezialisiert. Bei den Beständen beider Bibliotheken sowie den Nachweisen in der MENALIB handelt es sich um unterschiedliche Sammlungen, die sich gegenseitig ergänzen. Zudem wurde 2006 eine bereits bestehende Absprache zwischen der UB Frankfurt und der ULB Halle bezüglich der Erwerbung der wissenschaftlichen Literatur zum Thema Israel und der Nahostregion konkretisiert. Dabei wurde festgelegt, die wissenschaftliche Literatur nach sprachlichen Gesichtspunkten zu erwerben, so dass aus Gründen der Effizienz und Kompetenz die Erwerbung hebräischer Literatur in Frankfurt und arabischer Literatur in Halle erfolgt.

Aus den Metadaten der ULB Halle und dem Frankfurter Katalog wird ein gemeinsamer Index erstellt, der unter einer Weboberfläche mit Suchfunktionen die unterschiedlichen Sammlungen insbesondere auf Arabisch und Hebräisch zusammenführt und komfortabel nutzbar macht. Durch den Aufbau integrierter Indizierung, die auf den entsprechend ausgewiesenen Identifizierungsmerkmalen in Halle und Frankfurt beruhen, werden Resultate auch für ganz spezifische Anfragen generiert.

Somit wird ein Recherche- und Nachweisportal errichtet, das den Fachwissenschaftlern einen umfassenden Überblick über die Spezialliteratur zu Israel ohne Unterschied der Sprache und Anbieterbibliothek bietet.

Da der Datenbestand im Kern aus Verbunddatensätzen von HeBIS und GBV bestehen soll, die beide mit dem System PICA von OCLC arbeiten, ist eine Normalisierung des Datenbestandes außerhalb dieser Systeme für das Projekt nicht notwendig. Die Daten werden vielmehr im HeBIS-System aggregiert, d. h. der entsprechende Abzug der der ULB Halle-Daten wird in die HeBIS-Verbunddatenbank eingespielt, die dann als „Backend“ des Recherchetools dient. Als Rechercheoberfläche wird eine Instanz des gemeinschaftlich unter Beteiligung der UB Frankfurt auf VuFind-Basis entwickelten Discovery-Systems des Verbundes (HeBIS Discovery System)⁶ genutzt. Datenbasis der individuellen Sichten der Verbundbibliotheken ist ein gemeinsamer Gesamtindex, in dem dann auch die ULB Halle-Daten enthalten sein werden. Die Rechercheoberfläche für den FID Jüdische Studien ist als weitere Sicht realisiert, die um fachspezifische Funkti-

⁵ <https://www.menalib.de/> [Zugriff: 14.1.2018].

⁶ Sunckel, Bettina; Reh, Uwe und Niernerza, Heike: Das HeBIS Discovery System. In: Bibliotheksdienst 48 (2014), S. 784–794.

onen (Sach-Facetten, Berücksichtigung der Schreibweise von rechts nach links in der Anzeige) erweitert wird.

Unter Verwendung von Anker-Datensätzen im Verbundkatalog wird weiterhin die Spezialbibliographie *Library of the Haskala* (Bibliothek der Jüdischen Aufklärung)⁷ der Werke der Jüdischen Aufklärung des 17. und 18. Jahrhunderts in hebräischer und Deutscher Sprache eingebunden.

4 Kontextualisierung der Metadaten der Digitalen Sammlungen Judaica zur Verknüpfung von unterschiedlichen Judaica-Ressourcen und Optimierung der Recherchemöglichkeiten

Bei der Kontextualisierung geht es zunächst darum, Entitäten in den Metadaten zu identifizieren und persistente Identifier zu vergeben. Dabei wird auf die Daten und Identifier externer Datenquellen zurückgegriffen, was einerseits für eine semantische Eindeutigkeit sorgt und gleichzeitig erste Links zu weiteren Informationen liefert. Entitäten sind dabei vor allem Personen, Orte, Ereignisse und thematische Begriffe.

Die Quellen lassen sich grob in die folgenden Kategorien einteilen:

1. Bibliothekarische Normdaten: Im deutschsprachigen Raum ist das vor allem die Gemeinsame Normdatei der Deutschen Nationalbibliothek (GND), die unter anderem Personen und Sachschlagworte enthält. Die Personendaten sind eingebunden in das Virtual International Authority File (VIAF), worüber sich Links zu vielen weiteren Normdatensätzen von Nationalbibliotheken weltweit finden lassen. Auf der thematischen Ebene sind international vor allem die Library of Congress Subject Headings (LCSH) interessant, für die teilweise auch eine Konkordanz zur GND besteht.
2. Offene Daten aus dem Linked-Open-Data-Web (LOD)⁸: Die LOD-Quellen eignen sich sehr gut, um einen wesentlich breiteren und reichhaltigeren Kontext herzustellen. Viele Datenquellen werden hier von mehr oder weniger großen Communities gepflegt, bzw. leiten sich aus Community-Projekten ab,

⁷ <http://www.haskala-library.net/> [Zugriff: 14.01.2018].

⁸ Natürlich sind auch die bibliothekarischen Normdaten Teil des LOD-Webs, hier sind also die weiteren Quellen gemeint.

allen voran die auf Wikipedia basierende DBpedia⁹, das auf OpenStreetMap basierende Linked-Geodata-Projekt¹⁰, sowie Geonames¹¹, eine umfangreiche Datenbasis zu (historischen und multilingualen) Ortsnamen inklusive Geodaten.

3. Fachspezifische Datenquellen: Hierzu zählen Online-Enzyklopädien mit einem thematischen Fokus, aber auch spezielle Datenbanken, wie z. B. Namenslisten von Grabinschriften. Diese Quellen zeichnen sich vor allem dadurch aus, dass sie noch nicht in einer Form vorliegen, die für die Kontextualisierung geeignet ist, insbesondere fehlt es an persistenten Identifiern und einer Aufbereitung der Daten zur maschinellen Verarbeitung. Aufgrund des thematischen Bezugs und der hohen Qualität sind sie allerdings gerade für Fachwissenschaftler sehr interessant.

Die erste prototypische Kontextualisierung beschränkte sich auf Titel aus der Zeitschriften-Datenbank „Compact Memory“ mit der GND und DBpedia als Datenquellen zur Verlinkung. Der Fokus lag vor allem auf dem Gesamtprozess innerhalb des FID Jüdische Studien; vom Zugriff auf die Daten über die Datenbereitstellung für das Portal bis hin zur Nutzung und Anzeige der Ergebnisse im Portal selbst.

Ein Beispiel soll das Problem der Kontextualisierung verdeutlichen. In Compact Memory sind viele Zeitschriften mit rudimentären Metadaten beschrieben, oft kommt jedoch eine textuelle Beschreibung hinzu, wie etwa die folgende:

Bar Kochba: Blätter für die heranwachsende jüdische Jugend

Herausgeber: Cheskel Zwi Klötzel

Die Jugendzeitschrift Bar Kochba erschien vierzehntägig seit Frühjahr 1919. Das Blatt wurde im Juni 1921 aus wirtschaftlichen Gründen eingestellt.

Die kurzlebige Jugendzeitschrift Bar Kochba, die der zionistische Journalist Cheskel Zwi Klötzel (1891–1951) im renommierten Berliner Welt-Verlag herausgab, veröffentlichte hauptsächlich Erzählungen, Märchen und Sagen; Rätsel sowie Berichte zur Lage der jüdischen Jugend bildeten weitere Schwerpunkte des Blattes.

Hier ginge es zum Beispiel darum, die Person Cheskel Zwi Klötzel zu identifizieren, die im Text als Herausgeber genannt wird und für die im weiteren Verlauf auch Lebensdaten zu finden sind. Weiteren Kontext liefert z. B. der Wikipedia-Ar-

⁹ Lehmann, Jens; Isele, Robert; Jakob, Max et al.: DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. In: Semantic Web 6 (2015), S. 167–195.

¹⁰ Stadler, Claus; Lehmann, Jens; Höffner, Konrad; Auer, Sören: LinkedGeoData: A core for a web of spatial open data. In: Semantic Web 3 (2012), S. 333–354.

¹¹ <http://www.geonames.org/> [Zugriff: 14.01.2018].

tikel zu C. Z. Klötzel¹², in dem wiederum Bar Kochba erwähnt wird, so dass ein Link sehr sicher korrekt wäre. Über den Artikel findet man auch die GND-Nummer 11623234X¹³, in allen drei Quellen stimmen die Lebensdaten überein.

Bei der Kontextualisierung von Compact Memory wurden zwei wesentliche Herausforderungen identifiziert:

1. Die Kontextualisierung profitiert von der Ausnutzung von Querbeziehungen zwischen den zu verlinkenden Datenquellen. Das macht es allerdings erforderlich, diese Datenquellen zur lokalen Verarbeitung zu laden und die Querbeziehungen entweder vollständig oder explorativ während der Kontextualisierung herzustellen. Bei der Vielzahl und Größe der Datenquellen ist das nicht trivial, alleine das Laden der GND erfordert einen performanten Server mit ausreichend Hauptspeicher und einen optimierten Ladeprozess, die DBpedia ist nochmal deutlich größer.
2. Die Anzeige aller gefundenen Links im Portal ist unübersichtlich, zumal viele der Datenquellen Seiten präsentieren, die nicht sehr benutzerfreundlich sind.

Beide Problematiken sind aus anderen Projekten bekannt. Den Ansatz zur Lösung liefert *JudaicaLink*¹⁴, ein RDF-basierter Knowledge Graph, der eine vereinheitlichte Sicht auf die einzelnen Quellen enthält und diese durch Querverweise bündelt. Dabei entsteht schon durch diese Bündelung eine wesentlich reichere Datenquelle, die die Kontextualisierung der Metadaten verbessert gegenüber der Nutzung der einzelnen Quellen. Gleichzeitig fungiert *JudaicaLink* als Hub, so dass ausgehend von einer *JudaicaLink*-URL, die als Identifier zur Anreicherung genutzt wird, alle weiteren Datenquellen erreicht und zur Kontextbildung herangezogen werden können.

Im FID Portal werden Links zu *JudaicaLink* hinterlegt, so dass auf die Präsentation der Kontextualisierungsdaten auch Einfluss genommen werden kann.

Zur Erzeugung der Daten in *JudaicaLink* werden zwei Strategien verfolgt:

1. Aus jeder Datenquelle werden die für die Domäne der jüdischen Studien relevanten Daten extrahiert. Dazu kommen verschiedene Seed-Listen zum Einsatz sowie Techniken aus dem Focused Crawling, um weitere Daten zu finden und auf Relevanz zu prüfen. Für diese domänenspezifischen Untermengen der Daten werden anschließend Querbezüge hergestellt, um den Kontextualisierungsprozess zu unterstützen. Ein Beispiel zur Verdeutlichung:

12 https://de.wikipedia.org/wiki/C._Z._Kl%C3%B6tzel [Zugriff: 14.01.2018].

13 <http://d-nb.info/gnd/11623234X> [Zugriff: 14.01.2018.]

14 <http://www.judaicalink.org/> [Zugriff: 14.01.2018].

Basierend auf einer Liste von Rabbinern werden Kategorien in DBpedia identifiziert, mit denen diese Rabbiner beschrieben werden. Über die Kategorien lassen sich dann weitere Rabbiner finden, die potentiell zu weiteren Kategorien führen. Die Herausforderung besteht darin, das Ergebnis nicht zu breit werden zu lassen, also z. B. eine Kategorie „Bürger von Hamburg“ nicht mit aufzunehmen, nur, weil ein Rabbiner auch in diese Kategorie fällt.

2. Zusätzlich werden weiterhin Links direkt in den Datenquellen gesucht (z. B. Ortsnamen, die nicht per se als domänenrelevant angesehen werden können). Zu verlinkende Datensätze werden dann in JudaicaLink übernommen und analog zu den Datensätzen aus Strategie 1 verarbeitet.

Dieses Vorgehen hat den großen Vorteil, dass mit *JudaicaLink* eine – im Vergleich zur Größe aller Datenquellen – handhabbare Datenquelle für die Domäne der Jüdischen Studien entsteht, die in anderen Projekten direkt nachgenutzt werden kann, insbesondere in solchen, für die das direkte Arbeiten mit den Datenquellen und die Herstellung der Querbezüge ein prohibitiver Aufwand wäre.

Die zur Entwicklung des Verfahrens genutzte Sammlung enthält mehr als 5.500 Titel, von denen etwa 2.800 keine GND-Identifizierung hatten. Von diesen konnten bislang 1.200 Autoren automatisch kontextualisiert werden, bei einer Präzision von 78% (evaluiert über eine Stichprobe).

An Datenquellen stehen über JudaicaLink derzeit zur Verfügung:

1. Die gemeinsame Normdatei (GND) der Deutschen Nationalbibliothek,
2. DBpedia,
3. die Yivo Encyclopedia of Jews in Eastern Europe,
4. die Encyclopedia of Russian Jewry,
5. die Enzyklopädie „Das Jüdische Hamburg“.

Zur Einbindung der Daten in das FID Portal und zur Weiternutzung der Links auch außerhalb des Portals für andere Nutzer der GND wird mit der Deutschen Nationalbibliothek zusammengearbeitet und das dortige Projekt „Entity Facts“¹⁵ genutzt.

Die nächsten Schritte gemäß Projektplan sind der Ausbau von JudaicaLink durch das Hinzufügen weiterer Quellen (z. B. die Jewish Encyclopedia, eine gemeinfreie Enzyklopädie in zwölf Bänden, die Library of Congress Subject Headings und Geonames). Die interne Querverlinkung, sowie die Kontextualisie-

¹⁵ Beck, Julia; Büchner, Michael; Bartholmei, Stephan; Knepper, Marko: Performing Entity Facts. In: Datenbank-Spektrum 17 (2017), S. 47–52.

rungsverfahren sollen weiter verbessert und damit schließlich die Kontextualisierung auf den gesamten Datenbestand ausgeweitet werden.

**Kai Eckert**

Hochschule der Medien Stuttgart

Nobelstr. 10

70569 Stuttgart

Deutschland

E-Mail: eckert@hdm-stuttgart.de

ORCID: orcid.org/0000-0002-5423-561X

**Rachel Heuberger**

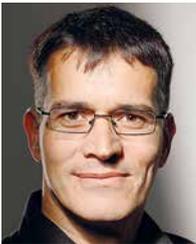
Universitätsbibliothek Frankfurt am Main

Bockenheimer Landstr. 134–138

60325 Frankfurt am Main

Deutschland

E-Mail: r.heuberger@ub.uni-frankfurt.de

**Marko Knepper**

Universitätsbibliothek Mainz

Jakob-Welder-Weg 12

55128 Mainz

Deutschland

E-Mail: m.knepper@ub.uni-mainz.de

ORCID: orcid.org/0000-0003-3905-0403