

Online-Workshop „Computerunterstützte Inhaltserschließung“ 2020

Zum ersten Mal in digitaler Form und mit 230 Teilnehmer*innen fand am 11. und 12. November 2020 der 4. Workshop „Computerunterstützte Inhaltserschließung“ statt, organisiert von der Deutschen Nationalbibliothek (DNB), der Firma Eurospider Information Technology, der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (SBB), der UB Stuttgart und dem Bibliotheksservice-Zentrum Baden-Württemberg (BSZ).¹

Im Mittelpunkt stand der „Digitale Assistent DA-3“²: In elf Vorträgen wurden Anwendungsszenarien und Erfahrungen mit dem System vorgestellt, das Bibliotheken und andere Wissenschafts- und Kultureinrichtungen bei der Inhaltserschließung unterstützen soll.

Die Begrüßung und Einführung in die beiden Workshop-Tage übernahm Frank Scholze (Generaldirektor der DNB). Er sieht den DA-3 als Baustein für die Verzahnung der intellektuellen und der maschinellen Erschließung.

Programm des ersten Tages

Regine Beckmann (Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, SBB) betonte in ihrem gemeinsam mit Imma Hinrichs (UB Stuttgart) entwickelten Auftaktvortrag „Organisation und Kommunikation rund um den Digitalen Assistenten DA-3“, wie wichtig eine gute Organisation und Kommunikation zwischen den beteiligten Personen und Institutionen sind und welche Herausforderungen bei der bibliotheks-, verbund- und länderübergreifenden Zusammenarbeit zu meistern sind.

Derzeit gibt es drei verschiedene Profile für die drei großen Anwendergruppen des DA-3. Das gemeinsame K10plus-Profil für die teilnehmenden Bibliotheken aus GBV und SWB ist seit November 2019 produktiv, das OBV-Profil seit Januar 2020 und das der DNB seit März 2020. Teilprojekte sind momentan in der Vorbereitungsphase: beim Fachinformationsdienst (FID) Theologie an der UB Tübingen, bei der ZBW – Leibniz Informationszentrum Wirtschaft und beim Fachinformationsverbund Internationale Beziehungen und Länderkunde (FIV).

Der DA-3-Lenkungsausschuss übernimmt die Organisation. Er setzt sich zusammen aus Verantwortlichen der beteiligten Verbände, der DNB, der Firma Eurospider und dem Advisory Board der Firma Eurospider, zu dem wiederum Koordinator*innen aus der DNB, dem GBV, dem OBV sowie dem SWB gehören. Im Advisory Board werden anstehende Entwicklungsschritte besprochen und dem Lenkungsausschuss vorgeschlagen. Der Lenkungsausschuss hat Entscheidungsbefugnis und diskutiert über neue Aufgaben und Wünsche aus dem Nutzer*innenkreis sowie deren Finanzierung.

- 1 Website zur Veranstaltung: <<https://wiki.dnb.de/pages/viewpage.action?pageId=181735291>>, Stand: 10.03.2021. Materialien zu den ersten drei Workshops werden von der UB Stuttgart bereitgestellt: <<https://blog.ub.uni-stuttgart.de/veranstaltungen/>>, Stand: 10.03.2021.
- 2 Vgl. <<https://www.eurospider.com/de/relevancy-produkt/digitaler-assistent-da-3>>, Stand: 10.03.2021.

Ansprechpartnerinnen im K10plus sind die Fachkoordinatorinnen Imma Hinrichs (UB Stuttgart) für die SWB-Bibliotheken und Regine Beckmann (SBB) für die GBV-Bibliotheken. Die jeweiligen Verbundzentralen VZG und BSZ fungieren als Ansprechstellen für Verträge, Schnittstellen und Importprogrammierungen; Ansprechpartner für Sonderentwicklungen und Verträge sind Eurospider sowie die Verbundzentralen.

Im Rahmen des seit 2017 jährlich stattfindenden Workshops „Computerunterstützte Inhaltserschließung“ findet ein jährliches DA-3-Anwendertreffen der teilnehmenden sowie interessierten Einrichtungen statt. Geplant ist ein Ticketsystem zur Dokumentation von Bugs, Erweiterungswünschen und Ideen allgemeiner Verbesserungen. Eine öffentliche DA-3-Webseite und ein Wiki, in dem Tool-Beschreibung, Handreichungen, technische Informationen, FAQs sowie Ansprechpartner*innen und Organisatorisches rund um den DA-3 abgelegt werden, befinden sich im Aufbau. So sollen Informationen gebündelt und zusätzlich Nachrichten zu Updates und neuen Funktionalitäten über eine Mailingliste ähnlich wie K10plus-SE verbreitet werden.

Je größer die Gemeinschaft der Teilnehmenden am DA-3 wird, umso mehr wird eine Arbeitsteilung im Bereich Vermittlungskonzept erforderlich. Bisher gab es einzelne Schulungen bei der Einführung des DA-3 in einer Einrichtung. Für K10plus-Bibliotheken fand zudem im Herbst 2020 eine einführende Online-Seminar-Reihe statt.

Da aus Zeitgründen nicht alle Fragen und Anregungen zum Vortrag aus dem Chat aufgegriffen und beantwortet werden konnten, wurde auf das DA-3-Anwendertreffen am Ende des Workshops verwiesen.

Im ersten Vortrag aus Anwendersicht stellte Martin Faßnacht vom FID Theologie an der UB Tübingen die Artikelerschließung im Index Theologicus (IxTheo) auf der Basis des DA-3 und die multilinguale Volltextsuche im IxTheo vor. IxTheo³ ist eine umfangreiche, internationale, frei zugängliche Bibliografie für Theologie und Religionswissenschaft.

Um eine mehrsprachige Suche in IxTheo zu ermöglichen, werden drei Instrumente genutzt: ein internes Übersetzungstool, der DA-3 und eine Volltextsuche. Im „IxTheo Translation Tool“ werden fachlich relevante normierte Sachschlagwörter und zum Teil die Verweisungsformen aus der GND zu einem Thesaurus zusammengestellt und von Fachwissenschaftler*innen und -übersetzer*innen in mehrere Sprachen übersetzt. Die Übersetzungen werden täglich in den IxTheo-SOLR-Index und in regelmäßigen Abständen in das „GND-Derivat“ des K10plus eingespielt und stehen so schnell für die Suche zur Verfügung. Das Übersetzungstool soll zukünftig für die Fachcommunity geöffnet werden, um z.B. für neu angesetzte Schlagwörter zügig Übersetzungen zur Verfügung zu haben. Für die Selektion noch nicht erschlossener Titel können Trefferlisten einer IxTheo-Suche auf noch nicht verschlagwortete Titel eingegrenzt werden (Option „[Unassigend]“ im „Thema [Schlagwort]“-Filter und im „IxTheo-Klassifikation“-Filter). Diese Treffer können direkt in den DA-3 übernommen und dort bearbeitet werden – ein komfortables Verfahren für die Erschließung. In der Volltextsuche werden drei

3 <<https://ixtheo.de>>, Stand: 10.03.2021.

Volltextarten unterschieden: der Volltext eines Aufsatzes oder einer Monografie, Inhaltsverzeichnisse und Zusammenfassungen/Abstracts. Es wird zurzeit mit einigen großen Verlagen darüber verhandelt, die kompletten theologischen und religionswissenschaftlichen Verlagsvolltexte für die Volltextsuche zur Verfügung gestellt zu bekommen; das wäre eine sehr wertvolle Ergänzung der IxTheo-Datenbasis. Die Volltextsuche wird in der Grundeinstellung ohne Synonyme durchgeführt; es kann alternativ eine Suche „mit Synonymen“ oder „mit mehrsprachigen Synonymen“ gewählt werden. Für die mehrsprachigen Synonyme wird auf die Übersetzungen der normierten GND-Schlagwörter zurückgegriffen.

In der Diskussion zum Vortrag wurden ausgehend von der Frage, warum die Übersetzungen nicht in die GND eingespielt werden, verschiedene Probleme thematisiert, die sich rund um die Übersetzung von Schlagwörtern, Thesauri und Normdaten ergeben können. Faßnacht und seine Kollegen vom IxTheo-Team sind sich gewisser Probleme bewusst, halten bei einer Abwägung aber die Vorteile einer Verankerung von Übersetzungen im GND-Normdatensatz für einen großen Mehrwert. Englischsprachige Schlagwörter (wie z.B. das im politischen Kontext verwendete „Appeasement“), die mit einer Übersetzung eines anderen deutschsprachigen Schlagworts (z.B. „Versöhnung“) kollidieren, könnten durch Homonym-Zusatz vereindeutigt werden. Die Lösung mit der Eintragung der Übersetzungen ins K10plus-„GND-Derivat“ mit entsprechender Kennzeichnung als IxTheo-Übersetzung ist ein Weg, der für diese Bibliografie erst einmal gut gangbar ist.

Im anschließenden Vortrag „Synergien nutzen – Maschinelle Unterstützung der intellektuellen Inhaltserschließung. Einsatz des DA-3 in der Deutschen Nationalbibliothek“ wurde neben der Anwendung des DA-3 und seiner technischen Umgebung die Plattform Culturegraph von Helga Karg, Hans-Jürgen Becker und Angela Vorndran vorgestellt. Das Projekt Culturegraph kann unterschiedliche Publikationen des gleichen Werkes in Werkbündeln zusammenführen.

Die DNB möchte alle Materialien, die sie sammelt, in hoher Qualität erschließen. Durch die Nutzung technischer Möglichkeiten, z. B. Fremddaten und automatische Vorschlagssysteme, soll Doppelarbeit vermieden werden. Der DA-3 setzt stark auf eine Vernetzung zwischen verschiedenen Arten der Inhaltserschließung. Konkordanzen können vorhandene Erschließungsdaten nachnutzbar machen, indem sie sie in andere Erschließungssprachen übersetzen. Der DA-3 bietet unter einer Oberfläche die meisten der benötigten Instrumente zur Inhaltserschließung an. Die DNB kann damit z.B. Schlagwörter oder Formangaben vergeben. Der DA-3 wurde für die Bedürfnisse der DNB angepasst. So sind die Anzeige von z.B. Titeldaten, GND-Normdaten und maschinellen Erschließungsdaten überarbeitet bzw. eingeführt worden. Bei der Anzeige der GND-Normdaten ist u.a. das Katalogisierungslevel wichtig, weil nur Daten mit dem Level 1 für die Inhaltserschließung verwendet werden. Im DA-3 kann direkt in der GND getrennt nach Schlagwortarten gesucht werden, und bei Hinweissätzen werden automatisch die zu verwendenden GND-Sätze angezeigt.

Das Vorschlagssystem des DA-3 bietet Fremddaten aus anderen Verbänden, aus den Daten des Marketing- und Verlagsservice des Buchhandels (MVB), Broadcastsearch und Ergebnisse der maschinellen Verschlagwortung an. Die Herkunft der Vorschläge wird deutlich angezeigt; Ergebnisse aus Mapping-Verfahren werden mit einem @ gekennzeichnet. Der DA-3 hat bei der Verarbeitung viele Vorteile; so kann unter einer Oberfläche die Recherche und die Bearbeitung eines Titels vorgenommen werden.

Daten lassen sich aus allen Quellen mit einem Klick übernehmen, und es können Titel auch in Stapeln verarbeitet werden. Zur Optimierung des DA-3 wurde von Seiten der DNB unter anderem eine Verbesserung der Schlagwortsuche angeregt. So kann jetzt nach Titeln mit bestimmten Schlagwortfolgen gesucht werden. Der WebDewey sollte zudem noch für die Vergabe der DDC eingebunden werden. Darüber hinaus sollte ein Qualitätsmanagement der maschinellen Erschließung möglich sein und die Pflege der Normdaten selbst eingebunden werden.

Für die Exportschnittstelle des DA-3 wurde die vorhandene technische Infrastruktur nachgenutzt, nur der Importer in den DA-3 musste neu erstellt werden. Geänderte Daten werden automatisch alle 5 Sekunden abgeholt, in PICA+-XML konvertiert und dann verpackt in einem SRU-RU-Request über technische Verarbeitungsstufen an das Zielsystem versandt. In der WinIBW wird die Schlagwortfolge dem Titel dann hinzugefügt.

Culturegraph könnte eine weitere Quelle für den DA-3 sein. Er bietet aggregierte Daten, sogenannte Werkbündel, an. In den Werkbündeln werden möglichst alle Manifestationen eines Werkes zusammengeführt, damit inhaltserschließende Daten bei allen Ausgaben des gleichen Werkes identisch sind. Der Abgleich erfolgt über Schlüssel, die verschiedene Metadaten-elemente repräsentieren. Derzeit wird Culturegraph für die retrospektive Anreicherung eingesetzt; nach der Einrichtung einer API könnte er auch im DA-3 zur Unterstützung bei Neuzugängen eingesetzt werden.

In der Diskussion ging es vor allem um die Möglichkeit, den DA-3 für das Qualitätsmanagement der maschinellen Erschließung der DNB einzusetzen. Die DNB führt laufend Kontrollen der Ergebnisse der maschinellen Erschließung durch; dabei bewerten die Erschließer*innen die Nützlichkeit der vergebenen Schlagworte in einem vierstufigen System.

Susanne Baron stellte in ihrem Vortrag „Geplante Inhaltserschließung im FIV mit dem DA-3 und dem EuroThesaurus“ zunächst den FIV vor: Der „Fachinformationsverbund Internationale Beziehungen und Länderkunde“ ist ein Zusammenschluss von zehn unabhängigen deutschen Forschungsinstituten, die als Informationsnetzwerk mit ihrer Datenbasis ein europäisches Gegenstück zur US-Datenbank PAIS bilden.

Der FIV plant ab Mai 2021 den Umstieg auf die Verschlagwortung mit dem DA-3 im K10plus. Nach einer kompletten Erneuerung der Systemarchitektur sollen sowohl Begriffe aus dem eigenen, neunsprachigen Euro-Thesaurus (ET IBLK) als auch GND-Schlagwörter aus dem DA-3 für die FIV-Datenbank (WAO-Datenbank) genutzt werden.

Im FIV-Arbeitsablauf gehören Sachklassifikation und Regionalklassifikation zur Formalerfassung der (häufig grauen) Literatur. Nach dem Beitritt zum SWB-Verbund wurde das FIV-Regelwerk durch RDA ersetzt, wobei viele Formalerschließungsprinzipien beibehalten werden konnten. Aber das FIV-Normvokabular für die Formal- und Sacherschließung musste an die GND angepasst (GND-IDs wurden in FIV-Normdaten eingetragen) und neue Begriffe in der GND angelegt werden.

Die Vergabe des FIV-Normvokabulars soll zukünftig mit dem DA-3 einfach zu handhaben sein, da größtenteils Honorarkräfte und anderes nicht WinIBW-geschultes Personal die tiefe Inhaltserschließung übernehmen.

Ohne die Einrichtung weiterer Schnittstellen und zusätzlicher (Verschlagwortungs- und Verwaltungs-) Felder auf der bibliografischen und der Exemplardatensatz-Ebene wäre der geplante DA-3-Umstieg in 2021 nicht möglich.

Im letzten Vortrag des ersten Workshoptages gaben Armin Kühn (BSZ) und Andreas Krausz (Verbundzentrale des GBV, VZG) in ihrem Vortrag „Technik des DA-3: Anschluss an K10plus“ einen Einblick in die Anwendungsgeschichte des DA in den Verbänden und in die technischen Belange, die für eine Teilnahme am DA-3 erforderlich sind.

Über die erste Version des Digitalen Assistenten in der ZB Zürich wurde auf dem Deutschen Bibliothekartag 2014 berichtet. Für das IBS|BW-Konsortium im SWB in Kooperation mit der UB Stuttgart, dem BSZ und der Firma Eurospider wurde der DA-2 entwickelt; er war ab Ende 2015 im Einsatz. Im GBV wurde der DA-2 ab Ende 2016 durch die SBB erprobt und ging 2018 produktiv. Ende 2018 stiegen beide Verbände auf den DA-3 um, um dann ab November 2019 im K10plus ein gemeinsames DA-3-Profil zu nutzen.

Für den Datenexport der im DA-3 vergebenen Sacherschließung, z.B. GND-Schlagwortfolgen, RVK oder Notationen der Basisklassifikation (BK), müssen die Daten via Perl XML-Parser zu PICA+ konvertiert werden, bevor sie im K10plus reimportiert werden können. Einspielungen erfolgen alle 70 Sekunden. Außerdem sind ggf. Lokalsätze mit Regionalcodes (SBB) anzureichern, Kategorien zu löschen (ZBW) oder Exemplardatensätze (IxTheo) anzulegen. Hinzu kommen noch Statistik- und (Fehler-)Kontrollverfahren. All dies wird über das Sigel und die Kennung (ILN) der jeweiligen Institution gesteuert.

Programm des zweiten Tages

Der Informatiker Martin Braschler von der ZHAW (Zürcher Hochschule für Angewandte Wissenschaften) berichtete von der voraussetzungsvollen Evaluierung von Retrievalergebnissen in „Spezielle Herausforderungen bei der Evaluation der Inhaltserschließung“. Er ist zudem unter anderem Mitbegründer des CLEF (Conference and Labs of the Evaluation Forum). Die Frage bei der Evaluation ist: „Welche Sacherschließungsmethode ist gut automatisierbar und bietet den Nutzern bessere Suchresultate?“

Die Sacherschließung wird im Vergleich zu den Ergebnissen aus der Volltextsuche bewertet, und es werden die Suchergebnisse untersucht. Es geht also um die Frage „Wie nützlich erweisen sich bestimmte Formen der Erschließung im Retrieval?“ und nicht „Wie gut sieht die Erschließung aus, wenn man die Indexierung auswertet?“. Zentrale Begriffe sind dabei Ausbeute und Präzision, die in einem Zielkonflikt zueinander stehen. Ausbeute steht dabei für „möglichst viele relevante Informationen“, und Präzision beschreibt die „Minimierung der ebenfalls gelieferten irrelevanten Informationen“. Schon in den frühen 1960ern stand die Frage im Raum, ob eine Volltextsuche die intellektuelle

Indexierung/Verschlagwortung ersetzen kann. Die Ansicht damals war, dass sich Volltexte positiv auf die Ausbeute auswirken und die Präzision durch intellektuelle Indexierung/Verschlagwortung erhöht werden kann. Heute sollte man versuchen, die beste Synthese aus beiden Verfahren zu finden und den Automatisierungsgrad so weit wie möglich zu erhöhen. Eine Evaluation erfolgt auf der Grundlage des Cranfield-Paradigmas. Es muss dafür sehr aufwendig eine Testkollektion erstellt werden, die eine Liste von Informationsbedürfnissen fiktiver Benutzer enthält, eine statische Datenkollektion und Relevanzbewertungen der einzelnen Titel für die in der Liste aufgeführten Informationsbedürfnisse. Weil der Aufwand für die Evaluation so hoch ist, haben sich hier Kooperationen gebildet, um die erforderlichen Datenkollektionen zu generieren. Man kann entweder bestehende Testkollektionen nachnutzen oder neue auf Volltext basierende Kollektionen mit zusätzlichen automatisch generierten Deskriptoren erstellen. Das Volltextretrieval und die Suche mit Deskriptoren lassen sich mit unterschiedlichen Retrievalmethoden untersuchen, wobei das Boolesche Retrieval bei Deskriptoren nur gut funktioniert, wenn diese aus einem kontrollierten Vokabular stammen. Ein Begriff, der in einem bestimmten Dokument häufig vorkommt, aber in vielen Dokumenten selten, lässt sich als „charakteristischer“ Begriff bezeichnen. „Und“ ist z. B. ein Wort, das sowohl in *einem* Dokument wie auch in *jedem* Dokument häufig vorkommt und sich deshalb nicht gut zur fachlichen Unterscheidung von Dokumenten eignet. Gemessen wird z. B. mit dem TF-IDF-Maß, das „lokale Häufigkeit mal globale Seltenheit“ bedeutet. Die Häufigkeitszählung macht bei Deskriptoren weniger Sinn, weil sie entweder 0 oder 1 als Wert aufweisen. Es gibt einige Studien zu Retrievalergebnissen von Deskriptoren mit unterschiedlichen Ergebnissen ihrer Nützlichkeit im Vergleich zu Volltexten. Es gibt Hinweise darauf, dass die Kombination von Volltexten und erstellten Deskriptoren die nützlichsten Suchergebnisse erbringt, wobei eine zentrale Studie von Savoy (2004) nicht frei von Kritik an der methodischen Umsetzung ist.

In der anschließenden Diskussion wurde über den Sinn von Sacherschließung debattiert. Eine intelligente Kombination von Volltexten und Deskriptoren als Grundlage für das Retrieval scheint am zielführendsten zu sein. Der Vorschlag, die Konfidenzwerte der maschinellen Erschließung in das Retrieval einzubeziehen, wurde von Braschler als sehr gute Idee bezeichnet, zu der es derzeit auch Forschungsvorhaben gibt. Daran anschließend wurde festgestellt, dass auch ein Konfidenzwert bei der intellektuellen Erschließung sinnvoll wäre, dass aber für ein solches Konzept noch einige Fragen geklärt werden müssten.

Der nächste Beitrag „Evaluation von Annif für die maschinelle Inhaltsererschließung an der Deutschen Nationalbibliothek“ von Matthias Nagelschmidt (DNB) stellte das Projekt „Erschließungsmaschine“ vor. Da die bisher eingesetzte Software zur automatisierten Inhaltsererschließung von Averbis nicht mehr weiterentwickelt wird, soll ein neues Erschließungssystem geplant und implementiert werden.

Das System soll verbal und klassifikatorisch erschließen können und dabei die Sprachen Deutsch und Englisch bedienen. Das Projekt befindet sich in der zweiten Phase der Evaluation möglicher Produkte. Die Entwicklung Annif der finnischen Nationalbibliothek vereint verschiedene Ansätze der Erschließung und ist als Open-Source-System verfügbar für die eigene Nachnutzung und Anpassung. Annif bietet sowohl statistische Ansätze wie TF-IDF als auch linguistische Verfahren, beispielsweise Maui und Kombinationen der angebotenen Instrumente, bei denen eigene Gewichtungen vorgenommen

werden können. Für die Trainings- und Testkorpora wurden Begriffe aus der GND und eigene Titelsätze der DNB ausgewählt. Aus der GND erfüllten 339.000 Begriffe alle Voraussetzungen (z.B. im Katalogisierungslevel eins und gleichzeitig mit mindestens einem Datensatz der DNB verknüpft zu sein). 1,18 Millionen Titeldatensätze bildeten die Grundgesamtheit; sie wurden in verschiedene Trainingskorpora bzw. einen Validierungs- und Testbereich untergliedert. Mit dem F1-Score, einem harmonischen Mittel aus Präzisions- und Vollständigkeitsquote, wurde die Übereinstimmung der Erschließung zwischen Annif und den vorhandenen Erschließungsdaten berechnet. In Bezug auf den Vortrag von Martin Braschler wurde darauf hingewiesen, dass es kein Retrievaltest war, sondern die Indexierung bewertet worden ist. Dabei schnitt Annif vor allem mit seinen kombinierten Modellen besser ab als die bisher eingesetzte Software von Averbis. Es wurden Testkorpora mit Volltexten und mit digitalisierten Inhaltsverzeichnissen eingesetzt. Daneben wurden die von Annif vergebenen Schlagwörter auch von den Erschließenden der DNB mit einem vierstufigen Nützlichkeits-Bewertungsschema evaluiert. Dabei waren 69 % der von Annif vergebenen Begriffe nützlich oder sehr nützlich. Nur 9 % der Begriffe waren falsch und im Durchschnitt fehlten 1,7 Begriffe, die hätten vergeben werden müssen, pro Titelsatz. Das Ergebnis der Software von Averbis war deutlich schlechter, diese hatte insgesamt 22 % falscher Begriffe vergeben und nur 53 % der Begriffe erwiesen sich als nützlich bzw. sehr nützlich. Immerhin zwei weitere notwendige Begriffe zur Inhaltsbeschreibung fehlten pro Dokument. Es wurden für Annif und Averbis verschiedene Korpora benutzt, die nicht fachlich spezifiziert waren.

In der Diskussion ging es um die Fehler menschlicher und maschineller Erschließung und dass ein Retrievaltest hier weitere Erkenntnisse zur Nützlichkeits von Inhaltserschließung für den Nutzer erbringen würde. Die Nützlichkeits von Begriffen hängt stark vom Informationsbedürfnis ab; es wurde das Beispiel einer Übersetzung gebracht, die auch in nicht allerbesten Qualität für bestimmte Bedürfnisse ausreichend war und deshalb deutliche Vorteile brachte gegenüber einer Nicht-Übersetzung. Auffällig war, dass wie im Vortrag von Martin Braschler kombinierte Verfahren die besten Ergebnisse brachten. Alle waren der Meinung, dass es auf dem Gebiet der Evaluation von Inhaltserschließung noch viele offene Fragen gibt.

Susanne Schmucker leitet in der ZBW unter anderem die inhaltliche Normierung. Sie hielt einen Vortrag zum „Einsatz des DA-3 in der ZBW“. Die ZBW benutzt für ihre Erschließung ihren eigenen Thesaurus, den Standard-Thesaurus Wirtschaft (STW), und für Personen und Körperschaften die GND. Sie nutzt die von den Autoren selbst eingegebenen freien Schlagwörter nach und vergibt Zeitangaben.

Für die von anderen Institutionen nicht genutzten Erschließungsinstrumente wie den STW mussten für die ZBW Anpassungen im DA-3 vorgenommen werden. Dafür wurde auch die Crosskonkordanz aus coli-conc⁴ zwischen STW und GND nachgenutzt. Bei der Besetzung der freien Schlagwörter und der Zeitangabe hat sich der DA-3 als sehr nützlich erwiesen, weil durch die gefelderte Eingabemaske Falscheingaben der Erschließenden in diesen Feldern nicht mehr vorkommen. Bei der herkömmlichen, nicht grafisch aufbereiteten Feldanzeige ist es hierbei durchaus manchmal zu Fehlern gekommen. Die

4 Coli-conc ermöglicht die „Verwaltung und Bereitstellung von Konkordanzen zwischen bibliothekarischen Wissensorganisationssystemen“ siehe <<https://coli-conc.gbv.de/de/>>, Stand: 10.03.2021.

ZBW erschließt auch Aufsätze und Sammelbandbeiträge und kann dabei seltener auf vorhandene Fremddaten zugreifen. Deshalb sind ein Vorschlagssystem und eine übersichtliche Aufbereitung der vorhandenen inhaltsbeschreibenden Angaben von großem Nutzen. Für elektronische Ressourcen hat der DA-3 eine deutliche Optimierung des Workflows mit sich gebracht. Die elektronischen Publikationen können in ausgewählten Listen in den DA-3 übernommen und dort dann abgearbeitet werden. Die nicht normierten Autorenkeywords werden angezeigt und dienen zusätzlich zur automatisierten Generierung von Vorschlägen für die Erschließung.

Für die ZBW brachte der DA-3 Verbesserungen bei der Inhaltserschließung. Durch das Scratchpad⁵ konnten deutliche Effizienzsteigerungen verzeichnet werden, und die Übernahme vorgeschlagener Begriffe ist sehr komfortabel. Auch ZBW-spezifische Anforderungen wie die Einbindung automatisierter Vorschläge und die bequeme Abfrage nach Parallelausgaben und Voraufgaben ist im DA-3 gut gelöst. Der Test des DA-3 in der ZBW wird auf immer größere Gruppen ausgeweitet. So sollen im ersten Quartal 2021 alle Fachreferent*innen für den Einstieg in das Produktivsystem geschult sein.

In der Diskussion wurde die schon längere Beschäftigung der ZBW mit inhaltserschließenden Verfahren, automatisiert und halbautomatisiert, geschildert. Der Kontakt zu Anif besteht schon länger; es werden Deep-learning-Verfahren eingesetzt. Die unterschiedliche Testkollektion und das Qualitätsmanagement der DNB und der ZBW wurden angesprochen. Das Qualitätsmanagement findet bei der DNB laufend statt, während die ZBW einmal im Jahr eine größere Evaluierung der Ergebnisse der maschinellen Erschließung vornimmt. Der DA-3 würde sich dabei sehr gut für Evaluierungen eignen.

Christoph Steiner (Leiter der Sacherschließung der Österreichischen Nationalbibliothek, ÖNB) schilderte in seinem Vortrag „DA-3 zu Corona-Zeiten & DA-3-Ticket-System“ die Vorbereitungen für die Arbeit mit dem DA-3 im Österreichischen Bibliothekenverbund (OBV). Er berichtete begeistert, wie sich dessen Anwendung im OBV bzw. in der ÖNB auch besonders in Corona-Zeiten bewährt hat.

Im Februar 2019 startete die Vorbereitung des DA-3-Einstiegs mit Schnittstellenprogrammierungen und Tests. Ein Jahr später war dann der DA-3-Echtbetrieb möglich, der in 25 % der OBV-Bibliotheken läuft.

Obwohl pandemiebedingt einige einführende Veranstaltungen ausfielen, wurde im Lockdown der DA-3 rege genutzt und damit durchschnittlich 3.000 – 4.000 Datensätze pro Monat angereichert.

Im Vergleich zu ALMA bietet der DA-3 eine viel effektivere Arbeitsumgebung mit Mehrwert durch zusätzlich verfügbare Normdaten (bspw. MSC⁶ + NLM⁷), die OBV Fachgruppen sowie literarische Gattungsbegriffe.

5 Ein Scratchpad ist eine Art elektronischer Notizzettel, auf dem man z. B. oft benötigte Informationen angeben kann, damit man sie bei Bedarf schnell zur Hand hat.

6 Mathematics Subject Classification

7 National Library of Medicine classification

Da im Homeoffice keine autoptische Buchbearbeitung möglich war, wurden Scans, Kopien von Umschlägen, Inhaltsverzeichnisse und ISBN-Barcodes der Ersatz. Über Barcodelesegeräte konnte die Nummernsuche im DA-3 genutzt werden. Für Fragen zur Arbeit mit dem DA-3 wurde ein Ticketsystem (OTRS) entwickelt, dessen Prototyp Christoph Steiner vorstellte. Unter einer zentralen E-Mailadresse sind zwei Ansprechpersonen für Probleme, aber auch Anregungen zuständig.

Gewünscht ist eine Ausweitung der Kommunikation rund um den DA-3 in allen Verbänden, damit Erfahrungen, Bugs sowie Strategien geteilt werden können.

Den „Bericht aus dem Lenkungsausschuss DA-3“ übernahm Helge Steenweg (Direktor der UB Stuttgart). Der Ausschuss, an dem die jeweilige Leitung der beteiligten Verbände und Einrichtungen teilnehmen, sorgt für die Einhaltung der geplanten Ziele. Durch mehr teilnehmende Bibliotheken werden mehr Wünsche geäußert, die möglichst schnell im DA-3 umgesetzt werden sollen. Helge Steenweg erläuterte die jeweiligen Vertragsstrukturen von OBV, DNB, BSZ und GBV mit der Firma Eurospider. Teils werden die Kosten ab 2021 durch Umlage auf die Bibliotheken verteilt (OBV, BSZ), oder sie werden vom Verbund finanziert (voraussichtlich im GBV).

Im Herbst 2020 fand eine Online-Seminar-Reihe rund um den DA-3 statt, die von Uma Balakrishnan, Andreas Krausz, Imma Hinrichs und Regine Beckmann durchgeführt wurde. Die Veranstaltungen richteten sich in erster Linie an interessierte Kolleginnen und Kollegen, die noch nicht mit dem DA-3 arbeiten, und sollten Antworten auf Fragen zu den richtigen Ansprechpartner*innen, technischen und anderen Voraussetzungen sowie einen ersten Einblick in die Arbeit mit dem Webtool geben.

Wichtige DA-3-Infos und FAQs sollen zukünftig aber auch auf Web- bzw. Wikiseiten zu finden sein. Diese sind ebenso wie ein Ticketsystem für Fehlermeldungen und ähnliche Anliegen in Planung.

Das kollaborative Vermittlungskonzept war ein zentrales Thema des sehr gut besuchten „DA-3-Anwendertreffens“, das von Regine Beckmann (SBB) moderiert wurde und eine rege Chat-Beteiligung hatte. Eine dringende Bitte zur Beteiligung an der Durchführung von Schulungen sowie am Teilen von Präsentationen jeglicher Art, z.B. kleinen Filmen/Videos etc., richtete sich an alle DA-3-Anwender-Einrichtungen. Jürgen Fenn regte die Anwerbung von Corona-Geldern (Neustart Kultur) für die gemeinsame Erstellung von Schulungsvideos an.

Regine Beckmann schlug Arbeitsgruppen vor, um in den Bereichen Vermittlung (Aktualisieren der Schulungsunterlagen, Erstellen digitaler Schulungsvideos) und technischer Support (Wiki-Seiten) voranzukommen. Die Last ist bisher auf zu wenige Schultern verteilt. Mithilfe ist gefragt und wurde u.a. schon von ÖNB/OBV, DNB sowie ZBW zugesagt.

Weitere Themen aus dem Chat wurden erörtert und Schwerpunkte für die Weiterentwicklung des DA-3 benannt, unter anderem RVK-Kompetenz, Optimierung der Listenfunktion, Anbindung von coli-conc, Quellen-Qualität, Konfidenzwerte etc.

Besonders betont wurden folgende Wünsche für die zukünftige gemeinsame Arbeit im DA-3:

- ein einheitliches Schulungskonzept,
- die Einhaltung der Qualitätsstandards der kooperativen Sacherschließung,
- Handreichungen auf der Webseite und im Wiki.

Fazit

Die zahlreichen Fragen, die auch im Chat gestellt wurden, zeigten die aktive Beteiligung der Zuhörer*innen; es war nicht immer genug Zeit, auch alle zu beantworten, aber die Referent*innen stehen sicherlich gerne auch im Nachgang für weitere Fragen und Diskussionen zur Verfügung. Zu einigen Aspekten gab es bereits Überlegungen zur weiteren Anwendung und Entwicklung auch im Ausbildungsbereich. Der DA-3 wird von allen Anwender*innen als nützliches und effizientes Instrument beschrieben. Er könnte neben der halbautomatisierten Erschließung als Instrument des Qualitätsmanagements für die automatisierte Erschließung eingesetzt werden. Positiv ist, dass sowohl die notwendigen Informationen vorhanden sind wie auch die abschließende Bearbeitung auf einer Oberfläche möglich ist.

Ein großer Dank ging an die Organisator*innen, die für den reibungslosen Ablauf der Veranstaltung sorgten. 2021 wird der 5. Workshop Computerunterstützte Inhaltserschließung am 10./11. November stattfinden; vielleicht auch wieder digital? Aufgrund der Erfahrungen des diesjährigen Workshops sollte dann auf jeden Fall für ausreichend (Kaffee-)Pausen gesorgt werden, um in Pausengesprächen thematisch an die Vorträge anknüpfen und den Austausch vertiefen zu können.

Heike Carstensen, Staats- und Universitätsbibliothek Hamburg

Birgit Kahl, Staats- und Universitätsbibliothek Hamburg

Sarah Pielmeier, Universitäts- und Landesbibliothek Münster

Viola Voß, Universitäts- und Landesbibliothek Münster

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5685>

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](#).