



Separatum aus:

THEMENHEFT 12

*Elisabeth Lienert / Joachim Hamm
Albrecht Hausmann / Gabriel Viehhauser (Hrsg.)*

Digitale Mediävistik

Perspektiven der Digital Humanities für die Altgermanistik

Publiziert im November 2022.

Die BmE Themenhefte erscheinen online im BIS-Verlag der Carl von Ossietzky Universität Oldenburg unter der Creative Commons Lizenz [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/). Die ›Beiträge zur mediävistischen Erzählforschung‹ (BmE) werden herausgegeben von PD Dr. Anja Becker (München) und Prof. Dr. Albrecht Hausmann (Oldenburg). Die inhaltliche und editorische Verantwortung für das einzelne Themenheft liegt bei den jeweiligen Hefherausgebern.

<http://www.erzaehlforschung.de> – Kontakt: herausgeber@erzaehlforschung.de
ISSN 2568-9967

Zitiervorschlag für diesen Beitrag:

Zeppezauer-Wachauer, Katharina: 50 Jahre Mittelhochdeutsche Begriffsdatenbank (MHDBDB). Eine Jubiläums-Zeitreise zwischen Lochkarten, Pixel-Drachen, relationaler Datenbank und Graph Daten, in: Lienert, Elisabeth/Hamm, Joachim/Hausmann, Albrecht/Viehhauser, Gabriel (Hrsg.): Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik, Oldenburg 2022 (BmE Themenheft 12), S. 161–186 (online).

Katharina Zeppezauer-Wachauer

50 Jahre Mittelhochdeutsche Begriffsdatenbank (MHDBDB)

Eine Jubiläums-Zeitreise zwischen Lochkarten, Pixel-
Drachen, relationaler Datenbank und Graphdaten

Abstract. Die seit 1972 betriebene Mittelhochdeutsche Begriffsdatenbank (MHDBDB) der Universität Salzburg ist ein umfassendes Recherchetool für das Mittelhochdeutsche. Kernelemente sind eine komplexe Suchmaschine und ein Wörterbuch, in dem mittels eines Begriffssystems Bedeutungen von korpusbasierten Wortartikeln erschlossen werden. Die einzelnen Vorkommensformen im Korpus werden auf diese Wortartikel und die im Kontext gültige Bedeutung sowie auf weitere Annotationsebenen wie beispielsweise grammatikalische Daten oder Metadaten bezogen. User können somit nicht nur nach Wörtern, Zeichenketten und Begriffen suchen, sondern auch wesentlich komplexere Fragestellungen auswerten. Dieser Beitrag liefert eine Rückschau über die letzten 50 Jahre sowie einen Ausblick auf zukünftige Entwicklungen.

Die Mittelhochdeutsche Begriffsdatenbank (MHDBDB) der Universität Salzburg ist seit den frühen 1970er-Jahren in Betrieb und inzwischen zu einem unerlässlichen Forschungswerkzeug der Fachcommunity geworden, wie erfreulicherweise auch dieser Sammelband zeigt. Mehrere im vorliegenden Band vorgestellten Digital-Humanities-Projekte arbeiten mit dem MHDBDB-Korpus: Dimpel [u. a.]; Brandes [u. a.]; Viehhauser. Der Dank ergeht an alle Kolleginnen und Kollegen, die das Feld um neue, lebendige Forschungsfragen und -methoden erweitern.

Als umfassendes Recherchetool für das Mittelhochdeutsche ist die MHDBDB auf der Grundlage digitaler Editionen und E-Texte aufgebaut. Ihr Korpus umfasst derzeit über 10,6 Millionen Tokens in fast 29.000 Wortartikeln und 666 Werken, wird jedoch beständig erweitert. Jährlich verzeichnet die Datenbank rund 28.000 individuelle User aus zahlreichen Ländern.

Kernelemente sind eine komplexe Suchmaschine und ein onomasiologisches Wörterbuch, in dem mittels eines Begriffssystems Bedeutungen von korpusbasierten Lemmata bzw. Wortartikeln erschlossen werden. Die einzelnen Vorkommensformen im Korpus (Tokens) werden auf diese Lemmata und die im Kontext gültige Bedeutung sowie auf weitere Annotations-ebenen wie beispielsweise grammatikalische Daten oder Metadaten (s. u.) bezogen. User können somit nicht nur nach Wörtern, Zeichenketten und Begriffen suchen, sondern auch wesentlich komplexere Fragestellungen auswerten, etwa zur linguistischen, onomasiologischen, grammatikalischen, sprachhistorischen, Wortschatz- oder autorspezifischen Forschung.

1. MHDBDB 1.0: Lochkarten

Die Anfänge der MHDBDB liegen im Jahr 1972, als Klaus M. Schmidt mit seinen Vorarbeiten zum Begriffswörterbuch der mittelhochdeutschen Literatur, einer Dissertation zur computergestützten Analyse des Wortmaterials Ulrichs von Liechtenstein, an der Universität Michigan (Ann Arbor) promoviert (vgl. Schmidt 1972; Schmidt 1978). Daher stammt auch das Logo der MHDBDB: die Helmzier Ulrichs im Codex Manesse (cpg 848, fol. 237^v). Aus dieser Dissertation entwickelten sich die ersten elektronischen MHDBDB-Grundlagen an der Bowling Green State University (Ohio, USA) als so genanntes »maschinengestütztes Begriffswörterbuch«. Damals gab es noch keine Computer-Monitore, keine Tastaturen und Mäuse. Die Ein- und auch Ausgabe der Annotationen für das mhd. Begriffswörterbuch erfolgte mit Lochkartenstanzern über Fortran-Lochkar-

ten. Drucker bzw. Fernschreiber gaben die erfassten Daten, die auf den Karten gespeichert waren, aus:

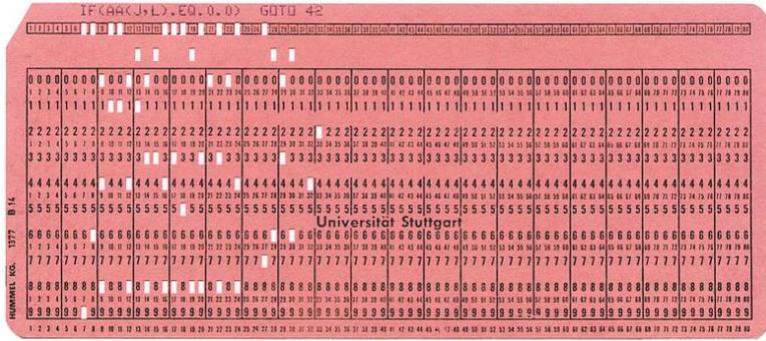


Abb. 1: Punch card Fortran Uni Stuttgart, Wikimedia Commons (CC BY-SA 3.0)

2. MHDBDB 2.0: MHDBDB goes online

Die nächste Großetappe für die MHDBDB lag in den 1990er Jahren: 1992 schlossen sich zwei Langzeitprojekte unter dem Namen »MittelHochDeutsche Begriffs-DatenBank« zusammen: Das Projekt »Namen in deutschen literarischen Texten des Mittelalters« von Horst P. Pütz (Christian-Albrechts-Universität zu Kiel) (vgl. Debus/Pütz 1987; darin insbes. Pütz, S. 287–299) und das erwähnte »Begriffswörterbuch der mittelhochdeutschen Literatur« von Schmidt. Pütz integrierte 1995 über 100 mittelhochdeutsche Texte und Namens-Annotationen und ermöglichte Schmidts Begriffswörterbuch dadurch einen gewaltigen Entwicklungsschub (vgl. Pütz/Schmidt 2001).

Auch das etwas sperrige, wenngleich mittlerweile weithin bekannte Akronym »MHDBDB« stammt aus dieser Zeit. Damals gab es allerdings auch noch eine englische Variante, die »MHGCDB« (»Middle High German Conceptual DataBase«).

Anlässlich des International Congress on Medieval Studies der Western Michigan University in Kalamazoo wurde die MHDBDB schließlich als menügesteuerte und in ihren Nutzungsmöglichkeiten noch limitierte Datenbank im Mai 1995 erstmalig über das Internet verfügbar gemacht (mhdbdb.bgsu.edu). Damit entstand ein weltweit zugängliches, einzigartiges Informationssystem zur mittelhochdeutschen Sprache und Literatur. Dank der Wayback Machine (archive.org) gibt es sogar noch Screenshots des typischen 90er-Jahre-Webdesigns, das auch eine text only-Version für den Fall anbot, dass die Nutzer*innen über keine leistungsfähige Grafikkarte verfügten:

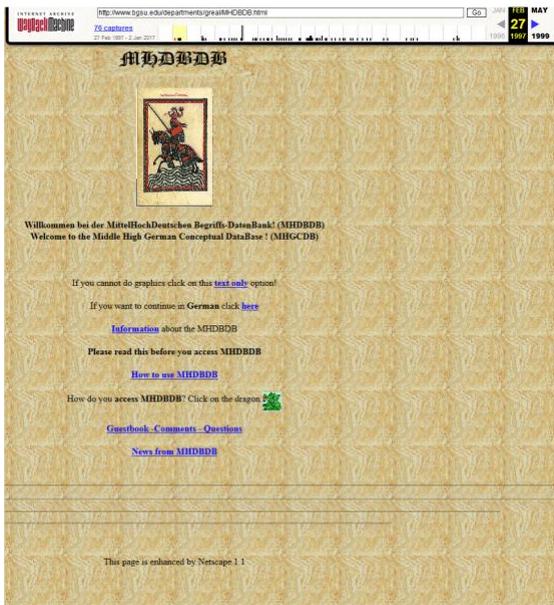


Abb. 2 + Abb. 3: Archivbild und Detail der MHDBDB-Website vom 27.02.1997, <https://web.archive.org/web/19970227152339/http://www.bgsu.edu/departments/greal/MHDBDB.html> (Link nicht mehr funktionsfähig, zuletzt abgerufen und gesichert am 07.02.2018)

Im Herbst 1998 wurde dieses System auf eine relationale Datenbank von ORACLE übertragen und eine neue Benutzeroberfläche auf der Basis von Websites im heutigen Verständnis erstellt.

3. Die MHDBDB im 21. Jahrhundert: Umzug von Ann Arbor nach Salzburg

Zwischen 2002 und 2004 fanden Verhandlungen mit der Paris Lodron-Universität Salzburg statt. 2004 wurde die MHDBDB schließlich in Salzburg installiert, ein kleiner Mitarbeiterstab (1,5 Stellen) zugesichert und von dort betrieben. Die Optik der Website änderte sich noch einmal geringfügig:

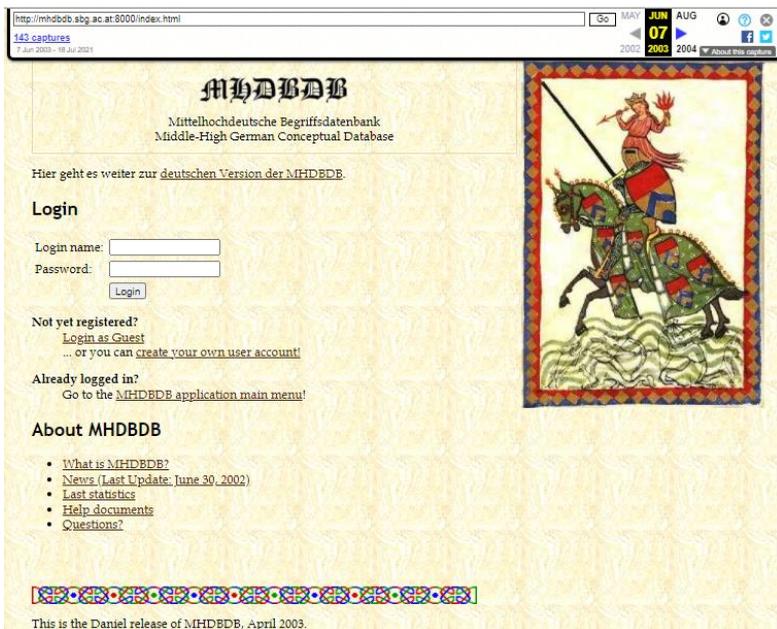


Abb. 4: Archivbild der MHDBDB-Website vom 07.06.2003,
<https://web.archive.org/web/20030607163507/http://mhdbdb.sbg.ac.at:8000/index.html> (zuletzt abgerufen und gesichert am 26.01.2022)

Das Rätsel um den »Daniel release« (unten am Bild) lässt sich über die alte Seitendoku aufklären, die für ein DH-Projekt aus dieser Zeit überraschend sorgfältig durchgeführt wurde. Die Releases waren die Versionennamen der Produktion, wobei »Daniel« noch vergleichsweise unspektakulär war, es gab etwa auch die Versionen »Craun«, »Erec«, »Wankelhart« oder »Fortimar«.

TEXT	LINE	POS	STRING	TAG	LINEC...	GRAMMAR	MEANING	WORD	LEMMA	NOTE	SOURCE
1 GUN	136100010	0	ein	1	1	0	-1	-1	-1 (null)	TS	
2 GUN	136100010	1	gart	0	1	0	-1	-1	-1 (null)	TS	
3 GUN	136100010	2	chom	0	1	32768	5361	45971	3400 (null)	TW	
4 GUN	136100010	3	seinem	0	1	1296	15291	32209	9131 (null)	TW	
5 GUN	136100010	4	litgeben	262144	1	1	19654	119410	12452 (null)	TW	
6 GUN	136100010	5	.	0	1	0	(null)	-1	(null) (null)	TW	
7 GUN	136100020	0	<	0	2	0	(null)	-8	(null) (null)	TW	
8 GUN	136100020	1	ich								
9 GUN	136100020	2	wil								
10 GUN	136100020	3	hie								
11 GUN	136100020	4	mic	0	2	256	6544	14531	4172 (null)	TW	
12 GUN	136100020	5	gesache	0	2	0	-1	-1	-1 (null)	TS	
13 GUN	136100020	6	leben	262144	2	0	(null)	(null)	(null) (null)	TS	
14 GUN	136100020	7	.	0	2	1	19654	23307310	(null) (null)	TW	
15 GUN	136100020	8	>	0	2	2	19654	21112400	(null) (null)	TW	

ID	STRING	STRING_LOWER	STRING_LANG	MEANINGLEMMA	TRIERID
12452	litgebe	litgebe	4B3C6E322819280001070101010100		12452 1101842

ME...	CATEGORY
1	19654 23307310 (null) (null) TW
2	19654 21112400 (null) (null) TW
3	19654 21111307 (null) (null) TW
4	19654 231112800 (null) (null) TW

ID	NAME	PARENT
23307310	Gastgewerbe	23307300

Abb. 5: Tabellenstruktur der relationalen Datenbank

Die folgende Zeit kann immerhin als direkter Augenzeugenbericht verfasst werden, denn meine erste Tätigkeit für die MHDBDB unter der ehemaligen Koordinatorin, Margarete Springeth, und dem sich bis heute als Berater und Mitarbeiter engagierenden Direktor, Klaus M. Schmidt, datiert ins Jahr 2010. 2012 bemühte ich mich zunächst – unter den gegebenen Bedingungen und mit de facto nicht vorhandenen finanziellen Mitteln – um ein kleineres optisches Update. Die Helmzier Ulrichs wurde von seinem Kopf gelöst und in ein Logo verwandelt, der Pixelhintergrund gegen ein *seamless* Pattern getauscht, die knallbunten, mit der Maus gezeichneten *divider* und andere visuelle Spuren der 1990er wurden peu à peu ausgetauscht. Auch inhaltlich kam es in den 2010ern neben der steten Erweiterung des Korpus zu zahlreichen Neuerungen wie beispielsweise in der Usability (Tooltip-Funktion bei der Textsuche, neue Filter, geschlechterspezifische Annotationen, breite Wortfeld-Annotationen, Neukonzeption des Begriffssystems, Ausgabemöglichkeit des text-/autorspezifischen Wortschatzes – Hapax-

legomena aka »[Unique-Words-Tool](#)«) sowie einem verstärkten Bemühen, die breiten Funktionalitäten der MHDBDB in der Fachcommunity zu verankern, etwa über implementierte Mini-Tutorials, themenspezifische Newsletter, Support über Social Media etc.

Darüber hinaus gab es zahlreiche Datenvernetzungen mit größeren Partnerprojekten: dem Trierer »[Wörterbuchnetz](#)«, dem Grazer »[Portal der Pflanzen des Mittelalters](#)«, dem italienisch-europäischen Poesie-Lexikon »[Lirica Europea di TrobVers e MHDBDB](#)« (vgl. Distilo 2013), der Kremser Bilddatenbank »[REALonline](#)«, der Oswald von Wolkenstein-Gesellschaft (Digitalisierung der Oswald-Ausgabe von K. K. Klein, siehe Literaturverzeichnis) sowie den diversen Editionsprojekten von Vlastimil Brom (Masaryk Universität Brunn), assoziiertem MHDBDB-Partner und Experte für deutschsprachige Texte in slawischer Umgebung (vgl. Brom 2019).¹

Vom Institut für Literaturwissenschaft und dem Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart wurde auf Basis der damals 10 Millionen MHDBDB-Tokens eine Trainingsdatei für einen Tree-Tagger des Mittelhochdeutschen erstellt (vgl. Ketschik [geb. Echelmeyer] [u. a.] 2017). Dabei handelt es sich um ein Werkzeug zur Annotation von Texten mit Part-of-Speech (PoS) und Lemma-Informationen, also der automatischen Zuordnung von Wörtern zu Wortarten.



Abb. 6: Logo der MHDBDB 2012–2022

4. MHDBDB 3.0: Die Reise geht weiter

Auf manche unserer Nutzer*innen mag es aufgrund des aktuellen Status quo der Website so wirken, als wäre die Entwicklung der MHDBDB seitdem zum Erliegen gekommen. Auf der dem Band zugrunde liegenden DH-Tagung prägte Joachim Hamm das treffende Bonmot der »Digitalen Editionsruine« (Gott sei Dank nicht auf die MHDBDB bezogen), das sich auf den ersten Blick auch für die Datenbank anbieten könnte. In Wahrheit ist jedoch das Gegenteil der Fall.

Die Verwaltung sämtlicher Daten (E-Editionen, Begriffssystem, Annotationen, Userdaten) findet seit 1998 in der relationalen ORACLE-Umgebung² statt; die Funktionen sind mit der objektorientierten Programmiersprache Java realisiert. War diese Technik 1998 noch auf dem neuesten Stand, so zeichnen sich inzwischen deutliche Mängel ab: Eine Vernetzung mit anderen Ressourcen, etwa in Form von Linked Open Data über das Semantic Web, ist, ebenso wie der unkomplizierte Import von Metadaten, kaum möglich. Metadaten sind im Grunde ›Daten über Daten‹ und beinhalten deskriptive, strukturelle, administrative oder technische Informationen zu einem Datensatz. Vor allem deskriptive Metadaten sind deshalb von großer Bedeutung, weil sie Daten auffindbar und nachnutzbar machen. Darüber hinaus können unter Metadaten auch Annotationen verstanden werden, die sich aber meist auf Teile eines Datensatzes beziehen, in Texten etwa auf einzelne Absätze oder Wörter, und deren Eigenschaften näher beschreiben (vgl. Schöch 2017, S. 223–233; Steiner 2021).

Aus den genannten Gründen des technischen Handicaps werden hinter den Kulissen bereits seit 2017 Vorkehrungen für einen Relaunch getroffen, für den bis heute weder die dafür benötigten finanziellen noch personellen Ressourcen zur Verfügung gestellt werden. Temporäre Hoffnung gab die Entwicklung eines Repositoriums der Universität Salzburg, ein Vorhaben, das aus Gründen leider wieder zum Erliegen gekommen ist und die avisierten Pilotprojekte ohne finales Ergebnis zurücklässt. Diese Umstände haben

dafür gesorgt, dass es leider nicht besonders wahrscheinlich ist, dass zum 50-jährigen Jubiläum der MHDBDB mit Ende des Jahres 2022 tatsächlich schon eine Beta-Version der komplett neu strukturierten, migrierten (Graph-)Datenbank³, an der seit Jahren gearbeitet wird, erscheinen kann – doch die Chancen stehen gut für 2023. Im Folgenden soll ein Ausblick über die neue Datenlage gegeben werden, die bereits heute vorliegt und nur darauf wartet, in ein funktionales Frontend eingebettet zu werden.



Abb. 7: Neues Logo der MHDBDB

Die wichtigsten Neuerungen sind die Nutzung von Standards, genormte Datenformate (insbesondere XML-TEI, RDF, SKOS, OWL), eine Anbindung ans Semantic Web⁴ und die Vernetzung mit Normdaten. Normdaten (engl. *authority files*), z. B. Wikidata und die GND, sind kontrollierte Vokabulare für bestimmte Domänen. In einer Normdatei hat eine Entität einen eindeutig referenzierbaren Eintrag, der mit weiterer spezifizierender Information ausgestattet sein kann. Solche Normdaten werden zukünftig auch von der MHDBDB nachgenutzt, damit User noch bessere Zugriffsmöglichkeiten auf das vorhandene Datenmaterial haben. Die Vernetzung mit Metadatenrepositorien ermöglicht das wechselseitige Anreichern der Daten (vgl. Steiner/Fritze 2021).

Beispiellos wird die neue Open-Access-Policy auf der Basis der FAIR-Prinzipien (*findable, accessible, interoperable* und *re-usable*)⁵ sein: Das MHDBDB-Volltextkorpus kann in Zukunft nicht nur vollständig (bisher nur in Auszügen in der Textsuche) gelesen, sondern zu weiten Teilen auch

als reiner Lesetext (PDF) oder Arbeitsversion (in diversen Datenformaten) heruntergeladen werden. Ebenso werden sämtliche Annotationen wie Begriffssystem (Semantik, Onomasiologie), Namenssystem (Onomastik), PoS/Grammatik, Wort- und Satzgrenzen, Phrasen- und Satzstrukturen, Wortfelder, Metadaten, Verknüpfungen zu Normdaten u. v. m. unter einer Creative Commons-Lizenz (voraussichtlich CC BY-NC-SA 3.0 AT) zur Verfügung gestellt (vgl. Hinkelmanns/Zeppezauer-Wachauer 2020; Hinkelmanns/Zeppezauer-Wachauer 2021).

Die neue MHDBDB setzt auf ein Datenmodell basierend auf unterschiedlichen Semantic-Web-Technologien wie RDF (*Resource Description Framework*)-Vokabulare und Ontologien⁶ sowie auf TEI (*Text Encoding Initiative*) zur Kodierung der Texte des Korpus (vgl. Pollin 2021b; Eibinger 2021). Sämtliche 666 E-Texte wurden bereits zur Gänze in das XML-TEI-Format konvertiert, das heute einen de facto-Standard in den Geisteswissenschaften darstellt. Die TEI-Texte wurden im Stand-off-Verfahren mit Linked Open Data verknüpft. Alle weiteren Forschungsdaten wie Annotationen sowie die bibliographischen Metadaten und die deskriptiven Metadaten zu Personen, Zeit, Orten und Ereignissen (basierend auf dem *Conceptual Reference Model* CIDOC-CRM für Kulturerbe-Daten, vgl. Pollin 2021a) liegen inzwischen als RDF-Daten vor. Sie werden direkt auf die Tokens der E-Texte bezogen. Die Vernetzung zwischen RDF- und TEI-Daten erfolgt mittels *Web Annotation Vocabulary* nach der Empfehlung des W3-Konsortiums.⁷

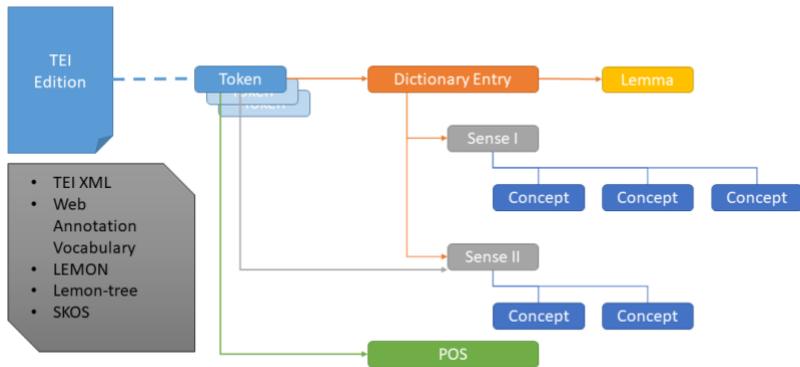


Abb. 8: Vereinfachtes Datenmodell der neuen MHDBDB

Die neue MHDBDB setzt durch die Verwendung kontrollierter Vokabularen und Ontologien zukünftig in sämtlichen Bereichen auf Nachhaltigkeit, um stabile Referenzen für Ressourcen sowie Interoperabilität zu gewährleisten. Kontrollierte Vokabulare und Normdaten helfen bei der Schematisierung von Inhalten geisteswissenschaftlicher Daten.

Die Wortartikel (Lemmata) wurden bereits nach den Vorgaben des *OntoLex-Lemon-Lexicography-Modules* kodiert. Weitere schon genutzte Ontologien bzw. Vokabulare sind *BibFrame 2.0*, die *GND Ontology* und eine im Rahmen unseres von der Österreichischen Akademie der Wissenschaften geförderten »Go! digital«-Projektes erstellte Semantic-Web-Ontologie für die Analyse narratologischer Muster in der Literatur und in Bildern des Mittelalters: *ONAMA (Ontology of Narratives of the Middle Ages)*. Diese Ontologie bietet ein Modell für die medienübergreifende Beschreibung von Handlungen, Akteur*innen, Schauplätzen und zeitlichen Strukturen. Neben den konstituierenden Grundelementen transmedialer Erzählkerne erfassen diese Beschreibungen die jeweiligen textlichen und bildlichen Umsetzungen im Detail und gehen damit über die reine Handlung eines Erzähltexts oder Bilderzyklus hinaus. ONAMA erlaubt es, Muster und Besonderheiten in den Strukturen von Erzählungen zu identi-

fizieren, die dann auf ihre zugrundeliegenden Ursachen und Funktionen hin untersucht werden können. Mit ONAMA können Bild- und Textquellen so annotiert werden, dass die generierten Daten Aufschluss über die Genese und Tradierung von Erzählkernen, Figurenkonstellationen, Handlungsmustern etc. im jeweiligen Medium und in medienübergreifenden Zusammenhängen geben (vgl. Nicka [u. a.] 2020; Zeppezauer-Wachauer [u. a.] 2021; Hinkelmans [u. a.] 2022; <http://onama.sbg.ac.at/ontology/>).

Das namensgebende Begriffssystem der MHDBDB war durch klassische, nach Sachgruppen strukturierte Thesauri inspiriert. Die Konzeption erfolgte in der Nachfolge von Rogets Thesaurus, als Weiterentwicklung der Dornseiff'schen Kategorisierung sowie vor allem in Anlehnung und Erweiterung des von Hallig und Wartburg entworfenen Begriffssystems (vgl. Roget 1852; Dornseiff 1970; Hallig/Wartburg 1963; Schmidt 1988, S. 40; Schmidt 1993, S. VII–X; Springeth 2009, S. 194).⁸ Die korpusorientierte Anforderung an das System hat über die Jahre dazu geführt, dass es um zahlreiche Kategorien erweitert wurde und neben Autosemantika auch weitere Funktionalitäten integriert wurden, etwa Synsemantika bzw. Funktionswörter, in den mhd. Texten vorkommende Fremdsprachen (Lateinisch, Französisch, Arabisch, Hebräisch, Griechisch, Slawisch, Tschechisch, Persisch, Italienisch, Spanisch, Ungarisch, Rätoromanisch, Englisch, Phönizisch-Punisch, Aramäisch, Jiddisch, Rotwelsch) oder eine Art wegbereitende Frühform von *Named-entity recognition* (NER). Das Design war konventionell hierarchisch und bei aller Nützlichkeit in seiner technischen Funktionalität sowie auch der Anwendungsfreundlichkeit limitiert. Usern, die das Begriffssystem zu schätzen und nutzen wussten, musste auch eine gewisse ›Leidensfähigkeit‹ attestiert werden, da die Arbeit mit Ziffern und Zahlencodes nicht besonders intuitiv war und eine längere Einarbeitung benötigte.

Aus diesem Grund wurde das Begriffssystem mittlerweile in einen polyhierarchischen *SKOS-Thesaurus*⁹ überführt, der es erlaubt, einzelnen semantischen Begriffen mehrere *parents* und *children* gleichzeitig zuzu-

ordnen, um standardisierte Notizen zu ergänzen und zudem verschiedene Labels (Synonyme) für die Begriffe zu vergeben. Die Anwendung erfolgt zukünftig über eine visuelle Suchmaske und via SPARQL (*SPARQL Protocol And RDF Query Language*, einer Abfragesprache für Graphdatenbanken), man benötigt also keine Zifferncodes mehr. Der weitaus größte Teil des Begriffssystems blieb bei diesen Arbeiten erhalten und wurde um vielfache standardisierte Relationen untereinander (*skos:broader*; *skos:narrower*; *skos:broadMatch*; *skos:narrowMatch*; *skos:related Match*; *skos:closeMatch*; *skos:exactMatch*) bzw. weitere Informationen (die Sprachen/Language-Tags Deutsch, Englisch, Französisch und Latein; *skos:altLabel*; *skos:editorialNote*; *skos:definition*) angereichert. Konzepte, die keine semantische Annotationsebene i. e. S. darstellen (z. B. Fremdsprachen, neuhochdeutsche Kommentare, Abkürzungen) wurden zusätzlich mit standardisierten Beschreibungen ausgezeichnet, etwa in TEI oder als Verknüpfung mit einschlägigen kontrollierten Vokabularien.

Besonders viel Arbeit verursachte die Neukonzeption der Namen: Die bisherige Namenssystematik, die das Vermächtnis des 1992 integrierten Namensprojektes von Pütz darstellt und überaus umfassend ist, wurde aus dem onomasiologischen Begriffssystem herausgelöst und zu einem komplexen Namensverzeichnis (Onomastikon) auf SKOS-Basis, mit bidirektional gerichteten Relationen ins Begriffssystem, umgestaltet. Es wird somit sowohl optional integrativ über das Begriffssystem oder aber separat über ein reines Namenslexikon durchsuchbar sein. Eine weitere Baustelle wird es sein, die frühe Form von NER, die bisher über das Begriffssystem realisiert wurde, zu professionalisieren und in heutige technische Standards zu überführen. Bislang wurde beispielsweise jedes Token ›Walther‹ über den Wortindex mit dem Begriffskonzept ›Literatur/Namen‹ verknüpft, um Walther von der Vogelweide als Personenentität in Texten zu identifizieren. Dies war in den 1990ern revolutionär; heutzutage gibt es mit Semantic Web und Normdaten aber andere Möglichkeiten.

Vergleichbares kann über Metadaten wie Dichterbiografien, die in der alten MHDBDB aufwändig als *plain text*-Beschreibungstexte angelegt sind, über bibliografische Metadaten zu den zugrundeliegenden Editionen und Handschriften oder über Gattungen/Genres gesagt werden. Beispielsweise die bisherigen Texttypen werden als kollaboratives Joint Venture verschiedener Netzwerke und Arbeitsgruppen in eine vollkommen neue Gattungstypologie umgebaut, die ebenfalls in SKOS beschrieben wird. Die Arbeiten an einer ersten Beta-Version werden in den nächsten Monaten abgeschlossen, und eine Publikation der Typologie auf GitHub, einem netzbasierten Dienst für das Teilen von Software-Entwicklungsprojekten, wird angestrebt. Für diese neue MHDBDB-Gattungstypologie wird sowohl auf Knowhow vor Ort zurückgegriffen¹⁰ als auch auf jenes internationaler Wissensverbände und Expert*innen, beispielsweise der »Brevitas – Gesellschaft zur Erforschung vormoderner Kleinepik«¹¹ (insbesondere der »Brevitas Systematik Kleinepik«); des »Netzwerk Offenes Mittelalter« zu Linked Open Data in der deutschsprachigen Mediävistik (vgl. hierzu auch Borek/Zeppezauer-Wachauer [u. a.] 2022) sowie zuvorderst von Marco Heiles (RWTH Aachen), dessen Mitarbeit an der Typologie im Bereich der Kategorisierung von Wissens- und Gebrauchsliteratur zentral ist. Das »Netzwerk Historische Wissens- und Gebrauchsliteratur« (HWGL), in dem sowohl Heiles als auch ich Mitglieder sind, liefert auf gemeinsamen Tagungen und per Mailinglist ebenfalls wertvolle Impulse zu Anwenderfragen.

Die Forschungsdaten wurden und werden also auf den neuesten technischen (und damit einhergehend auch inhaltlichen) Stand gebracht. Inzwischen sind alle Texte voll lemmatisiert, die Lyrik und der größte Teil der Kurzprosa zusammen mit einigen Epen sogar disambiguiert. Diese Daten werden besonders hilfreich bei der Entwicklung eines lernfähigen Disambiguierungsmoduls sein. Freilich sind solche Arbeiten niemals zur Gänze abgeschlossen, jedoch darf der aktuelle Datenstand mittlerweile wieder als gut kuratiert, dokumentiert und nachhaltig konzipiert bezeichnet werden.

Anders gestaltet sich die Situation bei der Präsentation dieser Daten in einer Weboberfläche. Das völlig neue Webdesign wurde bereits mit eigenen Mitteln entworfen und weitgehend implementiert. Die ursprünglich vereinbarte Umsetzung des Frontend durch das IT-Center der Universität Salzburg ist aufgrund der Einstellung des universitären Repositoriums jedoch zum Erliegen gekommen und liegt nun – unerwartet – wieder in den Händen des MHDBDB-Teams, da keine personelle oder finanzielle Unterstützung vorgesehen ist. Aktuell nehmen sich dieser mühevollen Kleinarbeit die (ehem.) MHDBDB-Mitarbeiter Peter Hinkelmanns, Daniel Schlager und Peter Färberböck auf Basis von TypeScript/ Angular an. Die Aussichten sind vielversprechend, wenn auch aus den genannten Gründen noch etwas Geduld erforderlich sein muss: Nach Fertigstellung wird die rundumerneuerte Datenbank nicht nur optisch ansprechend und anwendungsfreundlicher, sondern auch responsiv sein. Sie kann dann auch auf mobilen Geräten wie Tablets oder Smartphones genutzt werden. Die graphenbasierte Abfragesprache SPARQL wird komplexe Recherchen ermöglichen; eine visuelle Suchmaske dient dazu, dass auch User, die keine SPARQL-Erfahrung haben, das mögliche Maximum für ihre Forschung herausholen können. Eine Nutzer*innen-spezifische Oberfläche wird nicht nur den lang gehegten User-Wunsch einer Speichermöglichkeit über die letzten Suchen bieten, sondern beispielsweise auch das Anlegen eigener Text-Sammlungen.

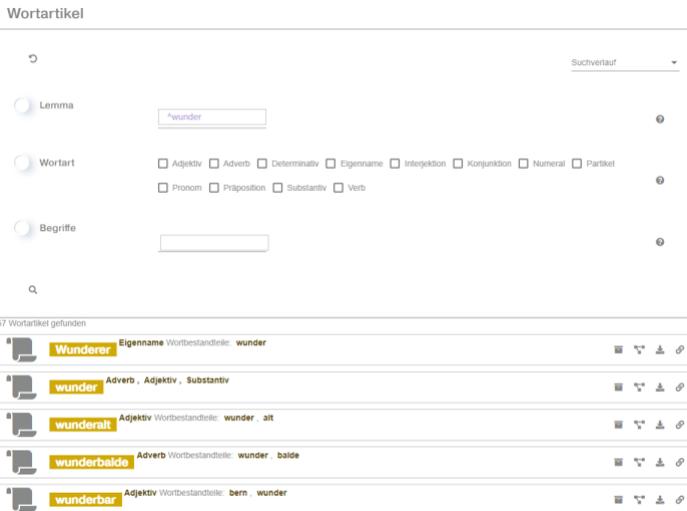


Abb. 9: Sneak peek Nr. 1 – Neues MHDBDB-Frontend (Wortartikel, die mit ›wunder-‹ beginnen)

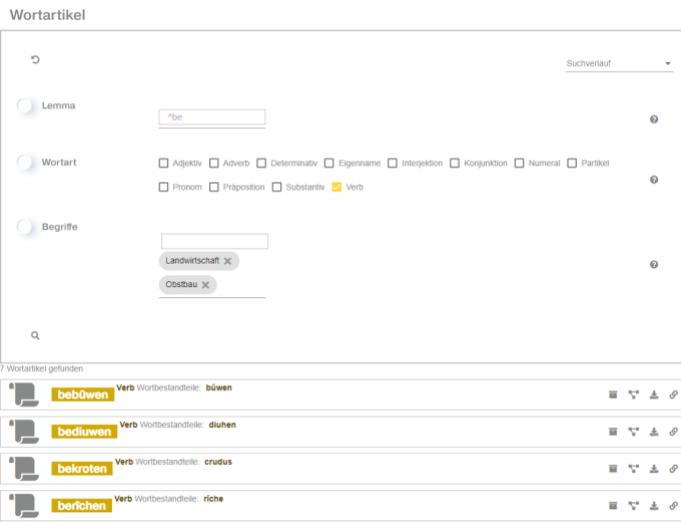


Abb. 10: Sneak peek Nr. 2 – Neues MHDBDB-Frontend (Wortsuche: Verba, die mit ›be-‹ beginnen und mit den Begriffen ›Landschaft‹ sowie ›Obstbau‹ assoziiert sind)

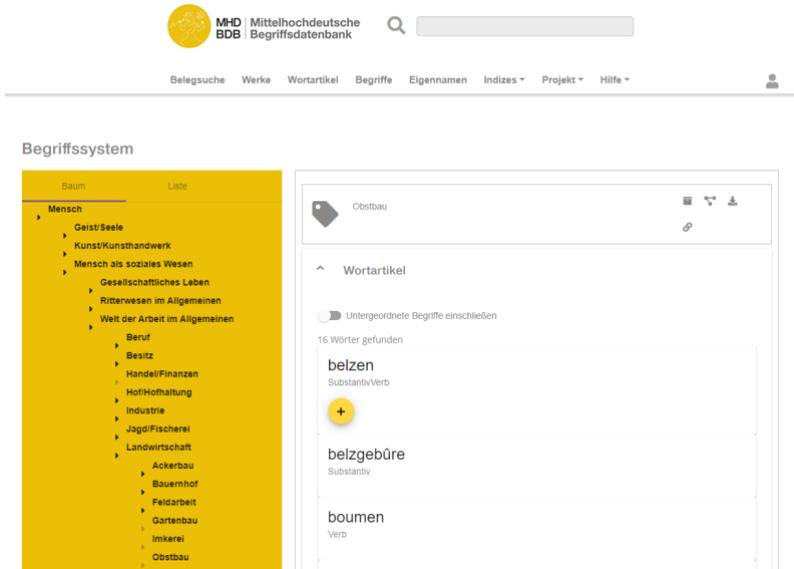
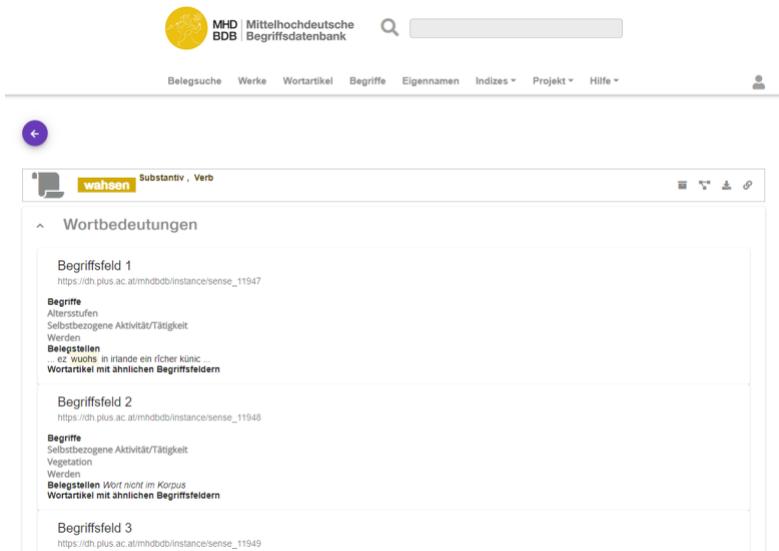


Abb. 11: Sneak peek Nr. 3 – Neues MHDBDB-Frontend (Begriffssystem, Beispiel ›Obstbau‹)



The image shows two screenshots of the MHD BDB web interface. The top screenshot displays the entry for 'wachsen' (Substantiv, Verb) with sections for 'Wortbedeutungen', 'Wortbildung', and 'Keyword in context'. The 'Wortbildung' section lists various derivatives like 'Wahsmuot', 'gewahsen', 'durchwachsen', etc. The 'Keyword in context' section provides example sentences. The bottom screenshot shows the same entry but with the 'Graphische Varianz' section expanded, displaying a table of word forms and their frequency in the corpus.

MHD BDB Mittelhochdeutsche Begriffsdatenbank

Belegsuche Werke Wortartikel Begriffe Eigennamen Indizes Projekt Hilfe

wachsen Substantiv, Verb

Wortbedeutungen

Wortbildung

Wahsmuot (Eigenname), **bewachsen** (Verb), **durchwachsen** (Adjektiv), **entwachsen** (Verb), **enwachsen** (Adverb, Negation, Verb), **erwachsen** (Verb), **gewahs** (Substantiv), **gewahsen** (Adjektiv), **gewahst** (Substantiv), **höchgewachsen** (Adjektiv), **langgewachsen** (Adjektiv), **metwachsen** (Adjektiv), **missgewahs** (Negation, Substantiv, Verb), **mittelwehshc** (Adjektiv), **selpawachsen** (Adjektiv), **underwahsenez** (Adjektiv, Präposition), **verwachsen** (Verb), **volwachsen** (Verb), **volwachseniu** (Adjektiv), **wahsender** (Adjektiv), **wahseter** (Substantiv), **wahsunge** (Substantiv), **wolgewachsen** (Adjektiv), **wirwahs** (Substantiv), **zuogewahsenez** (Adjektiv, Präposition), **zuowahsunge** (Substantiv, Präposition), **öfwachsen** (Präposition, Verb), **özwachsen** (Adverb, Präposition, Verb)

Keyword in context

... sin lip was wol **gewachsen** , schoene unde balt, ...
 ... sin Körme in zweinc jären **gewachsen** ze einem manne, beginnet ...
 ... erzoegen in solbers ja **wuochs** er bi ir vil harte ...
 ... erzooen in tenelande, si **wuochs** auch in der mätze ...
 ... ez **wuochs** in irlande ein richer künic ...
 ... sich niht betragen, nu **wuochs** diu maget junge, schoene ...
 ... ougenweide, do was ez **gewachsen** ze sibem järe lagen ...
 ... sagen unde singen er **wuochs** in einer waeste, der ...
 ... baz mohte geniezen, er **wuochs** unz an die stunde, ...

MHD BDB Mittelhochdeutsche Begriffsdatenbank

Belegsuche Werke Wortartikel Begriffe Eigennamen Indizes Projekt Hilfe

wachsen Substantiv, Verb

Wortbedeutungen

Wortbildung

Wahsmuot (Eigenname), **bewachsen** (Verb), **durchwachsen** (Adjektiv), **entwachsen** (Verb), **enwachsen** (Adverb, Negation, Verb), **erwachsen** (Verb), **gewahs** (Substantiv), **gewahsen** (Adjektiv), **gewahst** (Substantiv), **höchgewachsen** (Adjektiv), **langgewachsen** (Adjektiv), **metwachsen** (Adjektiv), **missgewahs** (Negation, Substantiv, Verb), **mittelwehshc** (Adjektiv), **selpawachsen** (Adjektiv), **underwahsenez** (Adjektiv, Präposition), **verwachsen** (Verb), **volwachsen** (Verb), **volwachseniu** (Adjektiv), **wahsender** (Adjektiv), **wahseter** (Substantiv), **wahsunge** (Substantiv), **wolgewachsen** (Adjektiv), **wirwahs** (Substantiv), **zuogewahsenez** (Adjektiv, Präposition), **zuowahsunge** (Substantiv, Präposition), **öfwachsen** (Präposition, Verb), **özwachsen** (Adverb, Präposition, Verb)

Keyword in context

Graphische Varianz

Wortform	Vorkommen im Korpus
wuochs	7
gewachsen	3
wachsen	1

Abb. 12, Abb. 13 + Abb. 14: Sneak peek Nr. 4 – Neues MHDBDB-Frontend (Wortartikel ›wachsen‹)

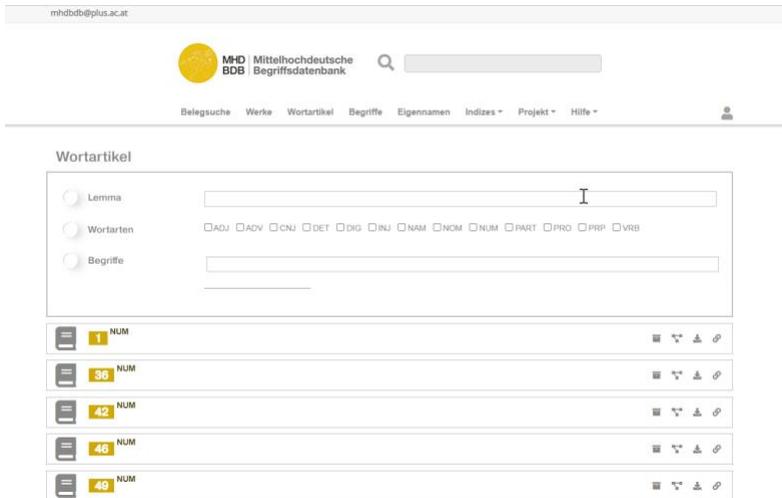


Abb. 15: *Sneak peek* Nr. 5 – Neues MHDBDB-Frontend (Wortsuche)

Weitere aktuelle Neuerungen finden sich auf der [MHDBDB-GitHub-Seite](#), die beständig erweitert wird, so beispielsweise um eine [Stoppwort-Liste](#) für die quantitative Untersuchung mittelhochdeutscher Texte (vgl. Hinkelmanns 2021a), ein [PoS-Tagger](#) für das Mittelhochdeutsche für [spaCy v3 \(GitHub\)](#) (vgl. Hinkelmanns 2021b) und eine [XQuery Module API](#) für Transkribus PageXML (vgl. Hinkelmanns 2021d). Das XQuery-Modul ermöglicht den Zugriff auf [Transkribus-PageXML-Dateien](#) über XQuery-Funktionen.

Die kommende Zeit steht weiterhin im Zeichen von ›Nachhaltigkeit‹: Erste Priorität hat die Verwendung von maschinenlesbaren, standardisierten, offenen, dokumentierten und kuratierten Datenformaten und kontrollierten Vokabularien. Die Umsetzung von Back- und Frontend wird das kleine MHDBDB-Team noch eine Zeit lang gut beschäftigen. Die nächste große Herausforderung wird darin bestehen, endlich die technischen und auch organisatorischen Voraussetzungen zu schaffen, um die verfügbaren

Forschungsdaten dauerhaft in einem zertifizierten Repository zu speichern und verfügbar zu halten.

5. Epilog

Mit dem vorliegenden Jubiläumsbeitrag wollte ich eine umfassende Rückschau über 50 Jahre MHDBDB liefern, die zukünftigen Generationen wissenschaftstheoretische, methodisch-konzeptuelle, aber auch ganz persönliche Einblicke in die Entwicklung dieses Urgesteins der Digitalen Mediävistik geben sollte.

Wir Mediävist*innen sprechen gern in Jahrhunderten. 1972–2022: ein halbes Jahrhundert Mittelhochdeutsche Begriffsdatenbank. Eine glückliche Zeit der Entwicklung, des Fortschritts, Lernens, Erfolges und der Neugierde auf all die neuen Möglichkeiten, die das breite Digital-Humanities-Feld bietet. Es waren aber auch Jahre, die viel Flexibilität, organisatorische Kreativität, Resilienz, Geduld und Beharrlichkeit von uns allen forderten. Ohne eine so treue Fachgemeinschaft wäre vieles davon nicht möglich gewesen. Dieser Beitrag sei daher Ihnen und Euch allen gewidmet, die die MHDBDB regelmäßig nutzen, sie hervorheben und zitieren, sie konzeptuell befruchten und mit immer neuen Forschungsideen und -wünschen aufwarten.

Wir blicken den nächsten 50 aufregenden Jahren hoffnungsvoll entgegen.

Anmerkungen

- 1 Vgl. beispielsweise Brom 2019. Dort weiterführende Literatur von Vlastimil Brom zur Arbeit mit und an der MHDBDB. E-Texte in der MHDBDB ediert und ingestiert von Brom: ›Di tutsch kronik von Behem lant‹ und ›Pulkava Chronik‹.
- 2 Ein klassisches relationales Datenbankmodell basiert auf der Speicherung von Informationen in verschiedenen Tabellen, die untereinander über Relationen (Beziehungen) verknüpft sind, siehe Abb. 5. Das Entwicklungsunternehmen

ORACLE gehört gemeinsam mit Microsoft SQL und IBM DB zu den Marktführern bei Datenbankmanagementsystem-Software.

- 3 Eine Graphdatenbank nutzt statt tabellarischen Relationen Graphen, um vernetzte Informationen darzustellen und abzuspeichern. Die Beziehungen werden dabei mithilfe von so genannten Knoten (Objekte) und Kanten (Verbindungen) ausgedrückt.
- 4 »Mit dem *Semantic Web* wird das reguläre World Wide Web um semantische Informationen ergänzt, die einen globalen Informationsgraphen ergeben.« Zitat und genauere Erläuterung s. Hinkelmanns 2021c.
- 5 »Die FAIR *Guiding Principles for scientific data management and stewardship* wurden 2016 veröffentlicht und besagen, dass Forschungsdaten auffindbar (*Findable*), zugänglich (*Accessible*), interoperabel (*Interoperable*) und wiederverwendbar (*Re-usable*) sein sollen. Sie bilden die Grundlage für eine disziplinen- und länderübergreifende Nachnutzung von Forschungsdaten.« (Stigler 2021).
- 6 »In der Informationswissenschaft und den Digital Humanities handelt es sich bei Ontologien um eine Form der Wissensrepräsentation. Man versucht Wissen aus der Welt oder thematisch kleiner gefassten Bereichen zu modellieren, abzubilden und zu formalisieren. Durch die Formalisierung wird es möglich, Daten und Wissen weiterzuverwenden, auszutauschen und sogar logische Schlussfolgerungen zu ziehen. [...] Ontologien stehen auch sehr stark in Zusammenhang mit dem *Semantic Web*, da sie dabei helfen, Informationen explizit und weiterverarbeitbar zu machen.« (Galka 2021).
- 7 Das *Web Annotation Vocabulary* spezifiziert das Set von RDF-Klassen, Prädikaten und Entitäten, die vom *Web Annotation Data Model*, einem strukturierten Modell für die Nutzung und Wiederverwendung von Annotationen, verwendet werden.
- 8 Onomasiologie wird seit Jahrzehnten innerhalb der deutschsprachigen historischen Lexikografie diskutiert. Ralf Plate schrieb beispielsweise schon 1992 in »Onomasiologische Umkehrlexikographie auf dem Prüfstand«: »Obgleich die Umkehrlexikographie [...] keine Erfindung des Computer-Zeitalters ist, hat ihr doch die Möglichkeit, die außerordentlich aufwendigen Umsortierungsprozeduren der Maschine zu überlassen, gesteigerte Aufmerksamkeit zuteil werden lassen. Dabei rückte eine Eigenschaft von Umkehrwörterbüchern ins Blickfeld, die zuvor kaum beachtet worden war: diejenigen Lemmata, die im Bedeutungswörterbuch an verschiedenen Stellen mit demselben Interpretament erläutert werden, erscheinen im Umkehrwörterbuch zusammen hinter diesem

jetzt als Stichwort auftretenden Interpretament. Das Umkehrwörterbuch läßt also Synonymreihen, Wortfelder der Objektsprache sichtbar werden und erfüllt so teilweise die Aufgabe eines onomasiologischen Begriffswörterbuchs. Onomasiologische Wörterbücher aber, unentbehrliche Hilfsmittel für Untersuchungen z.B. zur Stilistik, zu systematischen inhaltlichen Wortschatzbeziehungen oder zum sprachgeschichtlichen Bezeichnungswandel, sind in der historischen Lexikographie des Deutschen eines der größten Desiderate: ein ›Dornseiff‹ oder ›Wehrle-Eggers‹ für das Mittelhochdeutsche etwa, der die großen semasiologischen Wörterbücher von G. F. Benecke/W. Müller/F. Zarncke (BMZ) und M. Lexer ergänzen müßte, ist nicht vorhanden und nicht in Sicht.« Plate 1992, S. 313.

- 9 Das »*Simple Knowledge Organization System* (SKOS) ist ein W3C-Metadatenstandard für die digitale Organisation von Wissen. [...] SKOS versucht [...] Strukturen abzubilden und für den Austausch und die Verlinkung im Semantischen Web aufzubereiten.« (Bleier 2021).
- 10 Besonderer Dank ergeht an Univ.-Prof. Dr. Manfred Kern sowie Mag. Manuel Schwembacher, beide Universität Salzburg.
- 11 Brevitas – Gesellschaft zur Erforschung vormoderner Kleinelpeik e. V., vertreten durch: Silvan Wagner (1. Vorsitzender), Patrizia Barton (2. Vorsitzende), Friedrich Michael Dimpel (Schatzmeister): <http://brevitas.org/>.

Literaturverzeichnis

Handschriften

Codex Manesse Heidelberg, Universitätsbibliothek, cpg 848 ([online](#)).

Primärliteratur

Die Lieder Oswalds von Wolkenstein. Digitalisierung der Ausgabe von K.K.Klein im Auftrag der Mittelhochdeutschen Begriffsdatenbank MHDBDB und der Oswald von Wolkenstein-Gesellschaft von Bettina Hatheyer ([online](#)).

Sekundärliteratur

Bleier, Roman: Metadaten-Schemata für LZA: SKOS, in: Klug (2021) ([online](#)).
Borek, Luise/Zeppezauer-Wachauer, Katharina/Ketschik, Nora: Eindeutig Uneindeutig. Zur Modellierung von Unschärfe in der Mediävistik, in: Mittelalter.

- Interdisziplinäre Forschung und Rezeptionsgeschichte (16. Februar 2022) ([online](#)).
- Brom, Vlastimil: Die Mittelhochdeutsche Begriffsdatenbank als ein vielseitiges Arbeitsinstrument zur Analyse älterer deutschsprachiger Texte. Middle High German Conceptual Database: a flexible tool for analysing older German texts, in: Brünner Beiträge zur Germanistik und Nordistik 33, iss. Supplementum (2019), S. 173–184 ([online](#)).
- Debus, Friedhelm/Pütz, Horst P. (Hrsg.): Namen in deutschen literarischen Texten des Mittelalters. Vorträge. Symposium Kiel, 9.–12. 9. 1987, Neumünster 1989 (Kieler Beiträge zur deutschen Sprachgeschichte 12).
- Distilo, Rocco: Per un portale del lessico poetico europeo (TroVers/MHDBDB), in: *Forme e la storia. Rivista di filologia moderna*. VI, 1 (2013), S. 209–214 ([online](#)).
- Dornseiff, Franz: Der deutsche Wortschatz nach Sachgruppen. 7. Aufl., Berlin 1970.
- Eibinger, Julia: TEI (Text Encoding Initiative), in: Klug (2021) ([online](#)).
- Galka, Selina: Ontologie, in: Klug (2021) ([online](#)).
- Hallig, Rudolf/Wartburg, Walter von: Begriffssystem als Grundlage für die Lexikographie. Versuch eines Ordnungsschemas, Berlin 1963.
- Hinkelmanns, Peter (Hrsg.): List of Middle High German stopwords. 2021a ([online](#)).
- Hinkelmanns, Peter (Hrsg.): Middle High German pos tagger model for spacy. 2021b ([online](#)).
- Hinkelmanns, Peter (2021c): Semantic Web, in: Klug (2021) ([online](#)).
- Hinkelmanns, Peter (Hrsg.): XQuery Module API for Transkribus PageXML. 2021d ([online](#)).
- Hinkelmanns, Peter/Landkammer, Miriam/Nicka, Isabella/Schwembacher, Manuel/Zeppezauer-Wachauer, Katharina: Beyond the Plot. Der Vergleich mittelalterlicher Narrative im Semantic Web mit ONAMA, in: Vienna Doctoral Academy – Medieval Academy (Hrsg.): *Narrare – producere – ordinare. New approaches to the middle ages*, Wien 2022, S. 143–158.
- Hinkelmanns, Peter/Zeppezauer-Wachauer, Katharina: *ez ist ein wârheit, niht ein spel, daz netze was sinewel*. Die MHDBDB im Semantic Web, in: Fischer, Martin (Hrsg.): *Digitale Methoden und Objekte in Forschung und Vermittlung der mediävistischen Disziplinen*. Akten der Tagung Bamberg, 08.–10. November 2018, Bamberg 2020, S. 73–81 ([online](#)).
- Hinkelmanns, Peter/Zeppezauer-Wachauer, Katharina: Mittelhochdeutsche Begriffsdatenbank (MHDBDB), in: Klug (2021) ([online](#)).
- Ketschik [geb. Echelmeyer], Nora/Reiter, Nils/Schulz, Sarah: Ein PoS-Tagger für ›das‹ Mittelhochdeutsche, in: *Dhd 2017 Konferenzabstracts*. 2017, S. 141–147 ([online](#)).

- Klug, Helmut W. (Hrsg.): KONDE Weißbuch. Unter Mitarbeit von Selina Galka/Elisabeth Steiner im HRSM Projekt »Kompetenznetzwerk Digitale Edition«, Graz 2021 ([online](#)).
- Nicka, Isabella/Hinkelmanns, Peter/Landkammer, Miriam/Schwembacher, Manuel/Zeppezauer-Wachauer, Katharina: Erzählerische Spielräume. Medienübergreifende Erforschung von Narrativen im Mittelalter mit ONAMA, in: Schöch, Christof (Hrsg.): DHd 2020 Spielräume. Digital Humanities zwischen Modellierung und Interpretation, Paderborn 2020, S. 131–135 ([online](#)).
- Plate, Ralf: Onomasiologische Umkehrlexikographie auf dem Prüfstand, in: Zeitschrift für Dialektologie und Linguistik 59 (1992), S. 312–329.
- Pollin, Christopher (2021a): CIDOC-Conceptual Reference Model (CRM), in: Klug (2021) ([online](#)).
- Pollin, Christopher (2021b): RDF, RDFS, OWL, in: Klug (2021) ([online](#)).
- Pütz, Horst P.: Rechnergestützte Bearbeitung großer Datenmengen am Beispiel des entstehenden Lexikons, in: Debus, Friedhelm/Pütz, Horst P. (Hrsg.): Namen in deutschen literarischen Texten des Mittelalters. Vorträge. Symposion Kiel, 9.–12. 9. 1987, Neumünster 1989 (Kieler Beiträge zur deutschen Sprachgeschichte 12), S. 287–299.
- Pütz, Horst P./Schmidt, Klaus M.: Die Mittelhochdeutsche Begriffsdatenbank, in: Mittelalter-Philologie im Internet. ZfdA 130 (2001), S. 493–495 ([online](#)).
- Roget, Peter Mark: Thesaurus of English Words and Phrases. Classified and arranged so as to facilitate the Expression of Ideas and assist in Literary Composition [1. Aufl.], London 1852.
- Schmidt, Klaus M.: Tendenzen zum Realismus in der ritterlichen Epik der nachklassischen Periode. Untersuchungen zu Ulrichs von Liechtenstein Frauen dienst. Dissertation, University of Michigan, Ann Arbor MI 1972.
- Schmidt, Klaus M.: Wege zu Begriffsglossaren und einem Begriffswörterbuch mittelhochdeutscher Epik, in: Lenders, Winfried/Moser, Hugo (Hrsg.): Maschinelle Verarbeitung Altdeutscher Texte. Beiträge zum Symposion Mannheim 11./12. Juni 1971, Berlin 1978, S. 127–146.
- Schmidt, Klaus M.: Begriffsglossare und Indizes zu Ulrich von Liechtenstein, München 1980 (Indices zur deutschen Literatur 14/15).
- Schmidt, Klaus M.: Der Beitrag der begriffsorientierten Lexikographie zur systematischen Erfassung von Sprachwandel und das Begriffswörterbuch zur mittelhochdeutschen Epik, in: Bachofer, Wolfgang (Hrsg.): Mittelhochdeutsches Wörterbuch in der Diskussion (Germanistische Linguistik 84), Tübingen 1988, S. 35–49.
- Schmidt, Klaus M.: Begriffsglossar zu Ulrichs von Zatzikhoven ›Lanzelet‹, Tübingen 1993 (Indices zur deutschen Literatur 25).

- Schmidt, Klaus M.: Begriffsglossar und Index zur ›Kudrun‹, Tübingen 1994 (Indices zur deutschen Literatur 29).
- Schöch, Christof: Aufbau von Datensammlungen, in: Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hrsg.): Digital Humanities. Eine Einführung, Stuttgart 2017, S. 223–233 ([online](#)).
- Springeth, Margarete: Der analytische Weg ist das Ziel: Die Mittelhochdeutsche Begriffsdatenbank als Online-Textarchiv, in: Hofmeister, Wernfried/Hofmeister-Winter, Andrea (Hrsg.): Wege zum Text. Über die Verfügbarkeit mediävistischer Editionen im 21. Jahrhundert. Grazer Kolloquium 17.–19. September 2008, Tübingen 2009 (Beihefte zu editio, Bd. 30), S. 185–202.
- Steiner, Christian/Fritze, Christiane: Normdaten, in: Klug (2021) ([online](#)).
- Steiner, Elisabeth: Metadaten (allgemein), in: Klug (2021) ([online](#)).
- Stigler, Johannes: FAIR-Prinzipien, in: Klug (2021) ([online](#)).
- Woesner, Katrin: Begriffsglossar und Index zu Albrechts ›Jüngerem Titulek‹, Tübingen 2008.
- Zeppezauer-Wachauer, Katharina/Schwembacher, Manuel/Nicka, Isabella/Landkammer, Miriam/Hinkelmanns, Peter: Needful Things. Die Relationen der Dinge in einer Ontologie mittelalterlicher Narrative, in: MEMO. Medieval and Early Modern Material Culture Online 8 (2021) ([online](#)).

Online-Ressourcen

- BibFrame 2.0 (Bibliographic Framework Initiative):
<http://www.loc.gov/bibframe/>.
- Brevitas – Gesellschaft zur Erforschung vormoderner Kleinepik (e. V.):
<http://brevitas.org/>.
- GND (Gemeinsame Normdatei) Ontology: <https://d-nb.info/standards/elementset/gnd>.
- List of Middle High German stopwords: <https://github.com/Middle-High-German-Conceptual-Database/stopwords>
- Middle High German pos tagger model for spacy: <https://github.com/Middle-High-German-Conceptual-Database/Spacy-Model-for-Middle-High-German>.
- MHDBDB (Mittelhochdeutsche Begriffsdatenbank). 2004–2022 (laufend):
<http://mhdbdb.sbg.ac.at/>;
- GitHub: <https://github.com/Middle-High-German-Conceptual-Database>;
UniqueWords: <http://mhdbdb.sbg.ac.at/mhdbdb/App?action=UniqueWords>.
- HWGL (e. V.) (Blog „Netzwerk Historische Wissens- und Gebrauchsliteratur“):
<https://hwgl.hypotheses.org/>.
- Netzwerk Offenes Mittelalter. DFG-gefördertes Projekt zu Linked Open Data in der deutschsprachigen Mediävistik: <https://offenesmittelalter.org/>.

ONAMA (Ontology of Narratives of the Middle Ages): <http://onama.sbg.ac.at/>;
GitHub: <https://github.com/Middle-High-German-Conceptual-Database/onama>;
Zenodo: [10.5281/zenodo.4285986](https://zenodo.org/record/4285986).
OWL (Web Ontology Language): <https://www.w3.org/TR/owl-features/>.
MPS-Blog (Portal der Pflanzen des Mittelalters/Medieval Plant Survey) 2009–2022 (laufend): <http://medieval-plants.org/>.
REALonline. Bilddatenbank des Instituts für Realienkunde des Mittelalters und der frühen Neuzeit (IMAREAL) der Universität Salzburg. 2001–2022 (laufend):
<http://realonline.imareal.sbg.ac.at/>.
SKOS (Simple Knowledge Organization System Reference):
<http://www.w3.org/TR/skos-reference>.
spaCy - Industrial-Strength Natural Language Processing in Python:
<https://spacy.io/>;
GitHub: <https://github.com/explosion/spaCy>.
Transkribus: <https://readcoop.eu/transkribus/>.
Web Annotation Vocabulary: <https://www.w3.org/TR/annotation-vocab/>.
Wörterbuchnetz: <https://woerterbuchnetz.de/>.
XQuery Module API for Transkribus PageXML: <https://github.com/Middle-High-German-Conceptual-Database/xquery-pagexml-transkribus-module>.

Anschrift der Autorin:

Dr. Katharina Zeppezauer-Wachauer
Paris-Lodron-Universität Salzburg
Mittelhochdeutsche Begriffsdatenbank MHDBDB
Erzabt-Klotz-Straße 1
A-5020 Salzburg
E-Mail: katharina.wachauer@plus.ac.at